

**Genomweite CNV-Analyse und
zielgerichtetes Next-Generation-Sequencing
zur Identifizierung neuer Kandidatengene
bei Patienten mit HNPCC**

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Katrin Kayser

aus

Leverkusen

Bonn, März 2017

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Stefan Aretz

2. Gutachter: Prof. Dr. Jörg Höfeld

Tag der Promotion: 16. November 2017

Erscheinungsjahr: 2018

Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abkürzungsverzeichnis.....	1
1 Einleitung	1
2 Theoretische Grundlagen.....	3
2.1 Variation des menschlichen Genoms.....	3
2.2 Kopienzahlvarianten.....	3
2.2.1 Entstehungsmechanismen von CNVs	4
2.2.2 CNV-Detektion	5
2.2.3 Interpretation von CNVs	7
2.2.4 Pathomechanismus von CNVs	8
2.3 Erblicher Darmkrebs.....	9
2.3.1 Adenom-Karzinom-Sequenz	9
2.3.2 Zwei-Treffer-Hypothese nach Knudson	10
2.3.3 Familiäre adenomatöse Polyposis	11
2.3.4 MUTYH-assoziierte Polyposis.....	11
2.3.5 Polymerase Proofreading-assoziierte Polyposis.....	12
2.3.6 HNPCC und Lynch-Syndrom	12
2.4 Ziel der Arbeit	20
3 Material und Methoden.....	22
3.1 Patienten	22
3.2 Kontrollen.....	23
3.3 Materialien	23
3.3.1 Geräte und technische Ausrüstungen	23
3.3.2 Reagenzien und Kits	24

Inhaltsverzeichnis

3.3.3	Chemikalien	25
3.3.4	Lösungen	25
3.3.5	Software und Datenbanken	26
3.4	Primer Design	27
3.5	DNA Isolierung	28
3.6	Konzentrations- und Reinheitsmessung.....	28
3.7	Sanger-Sequenzierung.....	28
3.7.1	Polymerase-Kettenreaktion	29
3.7.2	<i>Cycle Sequencing</i> : Kettenabbruchmethode nach Sanger	30
3.7.3	Laserfluoreszenzdetektion	31
3.7.4	Aufreinigungsschritte	31
3.8	MLPA.....	32
3.9	Identifizierung und Validierung von Kopienzahlveränderungen	33
3.9.1	Genomweite SNP-Array-Analyse.....	33
3.9.2	In-silico Analyse der Array-Daten.....	36
3.9.3	Filterkriterien der CNV-Analyse	38
3.9.4	Quantitative PCR	42
3.9.5	Segregationsanalyse.....	46
3.9.6	Pathway-Analyse	46
3.9.7	Netzwerk-Analyse	46
3.10	Next-Generation-Sequencing	46
3.10.1	Nextera Rapid Capture Enrichment	47
3.10.2	Bioinformatische Datenanalyse	49
3.10.3	Varianten-Filterung	49
4	Ergebnisse	53
4.1	Keimbahnmutationsanalyse	53

4.1.1	Voruntersuchungen und Ausschluss von Patienten mit ursächlichen Mutationen	53
4.1.2	Untersuchung seltener Ursachen eines MSH2-Ausfalls	54
4.1.3	Kandidatengensequenzierung des Proteasom-Signalwegs	55
4.2	Differentialdiagnosen	55
4.2.1	Somatische Mutationen	56
4.3	CNV-Analyse	57
4.3.1	Suche nach Kandidatengen	57
4.3.2	CNVs in bekannten CRC-Genen und Interaktionspartnern von MSH2	58
4.3.3	Filterschritte auf Genebene	59
4.3.4	Priorisierung der Gene	61
4.3.5	Suche nach <i>Long Range</i> Regulatoren von <i>MSH2</i>	65
4.3.6	Netzwerk-Analyse	66
4.3.7	Pathway-Analyse	68
4.3.8	Segregationsanalyse	69
4.4	Next-Generation-Sequencing	73
4.4.1	Bekannte CRC-Gene der Literatur	73
4.4.2	Gene des Proteasom-Signalwegs	75
4.4.3	CNV-Gene	75
4.5	Zusammenfassung der Ergebnisse aus CNV- und NGS-Analyse	78
5	Diskussion	81
5.1	Ursächlich aufgeklärte Patienten des Kollektivs	81
5.2	CNV-Analyse	84
5.3	Sequenzierung der CRC-Gene der Literatur	95
5.4	Sequenzierung von Tumor-DNA	97
5.5	Limitationen	98
6	Zusammenfassung	100

Inhaltsverzeichnis

7	Ausblick	103
8	Literaturverzeichnis.....	104
9	Eigene Publikationen.....	120
	Anhang	I

Danksagung

Eidesstattliche Erklärung

Abkürzungsverzeichnis

®	rechtlich geschützt
µl	Mikroliter
∞	unendlich
A	Adenin
AFAP	attenuierte familiäre adenomatöse Polyposis
BAC	Bacterial Artificial Chromosome
BAF	B Allele Frequency
bp	basepairs
C	Cytosin
CGH	Comperative Genomic Hybridization
Chr	Chromosom
Cl	Chlorid
CN	Copy Number (Kopienzahl)
CNP	Copy Number Polymorphism
CNV	Copy Number Variation
CRC	Colorectal Cancer
CT	Cycle Threashold
ddNTPs	Didesoxyribonukleotidtriphosphate
Del	Deletion
del	deletiert
DEPC	Diethyldicarbonat
DGV	Database of Genomic Variants
DNA	Desoxyribonukleinsäure
dNTPs	Desoxyribonukleotidtriphosphat
Dup	Duplikation
dup	dupliziert
EC	Endometriumkarzinom
EDTA	Ethylendiamintetraessigsäure
EVS	Exome Variant Server
ExAC	Exome Aggregation Consortium
F	forward
FA	Familienanamnese
FAP	familiäre adenomatöse Polyposis
FISH	Fluorescence-In-Situ-Hybridization
FFPE	Formalin-Fixed, Paraffin-Embedded
FoSTeS	Fork Stalling and Template Switching
G	Guanin
H ₂ O	Wasser
HI	Haploinsuffizienz
HKG	House-Keeping-Genes
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HNR	Heinz-Nixdorff-Recall
HR	homologe Rekombination

Abkürzungsverzeichnis

IGV	Integrative Genome Viewer
IHC	Immunohistochemie
k	tausend
kb	kilobasen
KHCO ₃	Kaliumhydrogencarbonat
LCR	Low Copy Repeats
LBF	Log Bayes Faktor
LoF	Loss of Function
Log R	Logarithmus von R
LOH	Loss of Heterozygosity
LRR	Log R ratio
LS	Lynch-Syndrom
m	männlich
MAP	MUTYH-assoziierte Polyposis
mB	megabasen
MLPA	Multiplex Ligation-dependent Probe Amplification
mM	Millimol
MSI	Mikrosatelliteninstabilität
MSI-H	hohe Mikrosatelliteninstabilität
MSI-L	niedrige Mikrosatelliteninstabilität
MSS	Mikrosatellitenstabilität
n	Teilchenzahl
NAHR	nicht-alleleliche homologe Rekombination
n. u.	nicht untersucht
NaCl	Natriumchlorid
NaOH	Natriumhydroxid
NCBI	National Center for Biotechnology Information
ng	Nanogramm
NGS	Next-Generation-Sequencing
NH ₄ Cl	Ammoniumchlorid
NHEJ	Non-Homologous End-Joining
nM	Nanomol
p	kurzer Arm eines Chromosoms
PCR	Polymerase-Kettenreaktion
PLIGU	Patienten- und Laborinformationssystem für genetische Untersuchungen
pM; pmol	picomol
PPAP	Polymerase Proofreading-assoziierte Polyposis
q	langer Arm eines Chromosoms
Q	Qualität
qPCR	quantitative Polymerase-Kettenreaktion
R	reverse
RefSeq	NCBI Reference Sequence
RNA	Ribonukleinsäure
RTC	reversible Terminator Chemie
RVIS	Residual Value of Intolerance Score
SBE	Single Base Extension

SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
T	Thymin
t	Zeit
TCGA	The Cancer Genome Atlas
Tris	Tris(hydroxymethyl)-aminomethan
TSG	Tumorsuppressorgen
UPD	Uniparentale Disomie
USCS	University of California Santa Cruz
v	Version
VUS	Variance of Unkown Significance
w	weiblich
WGS	Whole Genome Sequencing
WHO	World Health Organisation
Z	Zielsequenz
Δ	Delta
ϵ	epsilon
λ	delta

1 Einleitung

Das kolorektale Karzinom (*Colorectal Cancer*, CRC) stellt mit etwa 62.400 Neuerkrankungen im Jahr die zweithäufigste Krebserkrankung in Deutschland dar (RKI Stand 2013, <http://www.rki.de/>). Eine genetische Prädisposition wird für etwa 20-30% aller CRC angenommen, wobei nur etwa 5% der Fälle auf eine monogene Form mit hochpenetranter Keimbahnmutation zurückzuführen sind (Hegde et al., 2014). Die monogenen Formen des erblichen Darmkrebses werden in polypöse und nicht-polypöse Erkrankungen unterteilt. Das Lynch-Syndrom (LS), im deutschen Sprachraum auch als hereditärer nicht-polypöser Darmkrebs (*Hereditary Non-Polyposis Colorectal Cancer*, HNPCC) bezeichnet, gilt als die häufigste monogene Form. Anders als beim sporadischen Darmkrebs, bei dem das durchschnittliche Erkrankungsalter bei 70 Jahren liegt, tritt das LS meistens vor dem 50. Lebensjahr auf. Zudem lässt sich in den betroffenen Familien eine Häufung syndrom-spezifischer extrakolonischer Tumoren finden, allen voran das Endometriumkarzinom bei Frauen. Für das LS sind Mutationen in Mismatch-Repair (MMR)-Genen und Deletionen im *EPCAM*-Gen verantwortlich (Lynch et al., 2009; Kuiper et al., 2010). Bei vorhandener Mutation liegt das Lebenszeitrisiko für Darmkrebs zwischen 30-70% und für das Endometriumkarzinom zwischen 40-50% (Steinke et al., 2014).

Bei klinischem Verdacht auf ein Tumorsyndrom ist die Identifizierung der genetischen Ursache aufgrund des hohen Lebenszeitrisikos und dem hohen Erkrankungsrisiko bei Familienangehörigen bedeutend für die Versorgung der Patienten und ihrer Angehöriger. Wird ein Tumorsyndrom rechtzeitig erkannt, ist durch die angebotenen Vorsorge- bzw. Früherkennungsmaßnahmen oft eine effektive Krebsprävention möglich. Zu den Vorsorge- und Früherkennungsprogrammen beim LS gehören eine jährliche Koloskopie sowie für Frauen zusätzlich ein transvaginaler Ultraschall mit Endometriumbiopsie. Wurde die familiäre Keimbahnmutation bei einer erkrankten Person identifiziert, ist es mittels prädiktiver genetischer Testung von Familienmitgliedern möglich, alle Anlageträger in einer Familie präsymptomatisch zu identifizieren und die Vorsorgemaßnahmen auf die tatsächlichen Anlageträger einer Familie zu beschränken. Nicht-betroffene Familienmitglieder werden psychisch entlastet und können aus dem strengen Vorsorgeprogramm entlassen werden.

Bei bis zu 50% der Patienten, welche die klinischen Auffälligkeiten für ein LS aufweisen, lässt sich trotz umfassender diagnostischer Screening-Methoden keine ursächliche Keimbahnmutation nachweisen (Bonis et al., 2007, p. 20; McPhillips et al., 2005; Obermair et al., 2010). Die familiäre Häufung von Tumoren wie auch das Auftreten von Tumoren in jungen Jahren (< 50) sprechen allerdings für eine zugrunde liegende genetische Prädisposition. Neben der Möglichkeit, dass Mutationen in den bekannten, LS-assoziierten Genen mit routinediagnostischer Methoden übersehen werden, können auch weitere, bisher unbekannte Gene ursächlich für das Auftreten der Erkrankung sein. Ziel dieser Arbeit war es daher, bei mutationsnegativen Patienten mit klinischem Verdacht auf Vorliegen eines LS mittels neuer molekulargenetischer Methoden (CNV-Analyse, Hochdurchsatzsequenzierung) sowohl Mutationen in bereits etablierten Genen zu finden, wie auch weitere ursächliche Gene oder Risikovarianten zu identifizieren und weiter zu charakterisieren.

In den letzten Jahren konnten seltene Kopienzahlvarianten (*Copy Number Variations*, CNVs), insbesondere größere heterozygote Deletionen, als ein quantitativ wichtiger Mutationstyp bei einer Reihe erblicher Tumorsyndrome nachgewiesen werden (Kuiper et al. 2010; Krepischi et al., 2012). Es ist deshalb davon auszugehen, dass größere Kopienzahlvarianten auch zum Mutationsspektrum noch unbekannter ursächlicher Gene des LS gehören. Für die Suche nach neuen ursächlichen Genen wendeten wir deshalb eine genomweite CNV-Analyse zur Identifizierung seltener Keimbahn-Deletionen und -Duplikationen und eine Hochdurchsatz-Sequenzierung von Kandidatengenen an. Hierfür wurde ein großes, gut charakterisiertes Kollektiv nicht-verwandter Patienten mit starkem Verdacht auf Vorliegen eines LS (erfüllte klinische Kriterien und isolierter Ausfall des MSH2-Proteins im Tumor) untersucht.

2 Theoretische Grundlagen

2.1 Variation des menschlichen Genoms

Lange Zeit wurden nur die beiden genetischen Größenextreme untersucht: große, mikroskopisch sichtbare chromosomale Veränderungen (Zerres et al., 1992) oder Einzelbasen-Polymorphismen (*Single Nucleotide Polymorphisms*, SNPs) bzw. pathogene Punktmutationen (International HapMap Consortium, 2005; Lander et al., 2001). Im Größenbereich dazwischen liegen die Strukturvarianten, mit denen man sich erst seit den letzten zehn Jahren eingehend befasst (Iafrate et al., 2004; Korbelt et al., 2007; McCarroll et al., 2008; Redon et al., 2006; Sebat, 2004).

Strukturvarianten sind kleinere oder größere Umbauten des Genoms. Inversionen oder Translokationen mit gleichbleibender Kopienzahl werden als balancierte Strukturvarianten bezeichnet. Strukturvarianten mit veränderter Kopienzahl, sogenannte Kopienzahlvarianten (*Copy Number Variations*, CNVs), zählen zu den unbalancierten Strukturvarianten. Sie können als Deletionen (Kopienzahl = 0 oder 1) oder Duplikationen / Multiplikationen (Kopienzahl ≥ 3) vorliegen (Feuk et al., 2006).

Werden CNVs mit berücksichtigt, kommt man zu der Einschätzung, dass zwei Menschen bis zu 99,5% genetisch identisch sind (Levy et al., 2007). Diese ~0,5% inter-individuellen Unterschiede bestimmen unter anderem, wie sensibel wir auf Umwelteinflüsse reagieren, wie und ob wir auf bestimmte Arzneimittel ansprechen und für welche Erkrankungen wir ein erhöhtes Risiko tragen (Feuk et al., 2006; Hastings et al., 2009). Diese Varianten sind daher Mittelpunkt der genetischen Forschung.

2.2 Kopienzahlvarianten

Hinsichtlich der Anzahl an betroffenen Basenpaaren nehmen CNVs den größten Anteil aller Varianten im menschlichen Genom ein (Conrad et al., 2010; Levy et al., 2007; Pang et al., 2014; Redon et al., 2006). Die Größendefinition eines CNVs ist durch die verbesserten Auflösungsmöglichkeiten von ursprünglich 1 kb (Feuk et al., 2006) auf heutzutage ~50 bp (MacDonald et al., 2014) herabgesetzt worden. Noch kleinere Insertionen oder Deletionen werden zusammengefasst als „Indels“ bezeichnet.

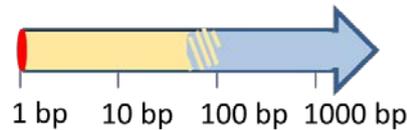


Abb. 1 SNVs, Indels und CNVs

Eigene Abbildung. Rot: SNVs, gelb: Indels, blau: CNVs

2.2.1 Entstehungsmechanismen von CNVs

CNVs werden in „rekurrente CNVs“ mit identischer Länge und gemeinsamen Bruchpunkten und in „nicht-rekurrente CNVs“ mit unterschiedlicher Länge und Bruchpunkten unterteilt. Die zugrundeliegenden Entstehungsmechanismen sind die nicht-allelele homologe Rekombination (NAHR) für rekurrente CNVs und das *Non-Homologous End-Joining (NHEJ)* sowie das *Fork Stalling and Template Switching (FoSTeS)* für nicht-rekurrente CNVs (Abb. 2) (Stankiewicz und Lupski, 2002, 2010).

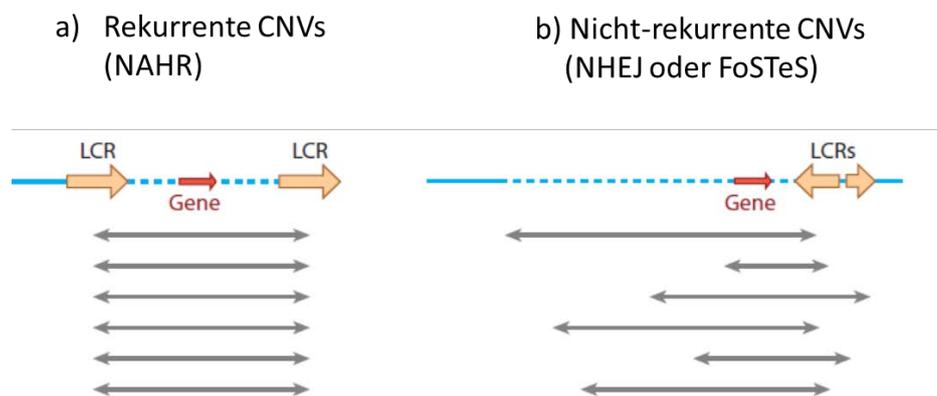


Abb. 2 Rekurrente und nicht-rekurrente CNVs

a) Rekurrente CNVs in Regionen mit flankierenden Bereichen von *Low Copy Repeats (LCRs)*, mit gleicher Länge und homologen Bruchpunkten. Entstehungsmechanismus: nicht-allelele homologe Rekombination.

b) Nicht-rekurrente CNVs mit unterschiedlich großen CNVs und versetzten Bruchpunkten. Entstehungsmechanismus: *Non-Homologous End-Joining* und *Fork Stalling and Template Switching*. Befinden sich LCRs in der Umgebung, gruppieren sich die Bruchpunkte dort.

(Stankiewicz und Lupski, 2010).

Nicht-allelele homologe rekombination

Rekurrente CNVs entstehen in Regionen mit *Low Copy Repeats (LCR)*. Dies sind kurze, repetitive DNA-Segmente von 10-400 kb mit hoher Sequenzidentität (Bailey et al., 2001). In der S-Phase der Mitose oder Meiose kann es durch die hohe Homologie dieser Regionen zu falschem

Aneinanderlagern der Schwesterchromatide kommen. Durch ungleiches Crossing-over entstehen so in der Tochterzelle Kopienzahlveränderungen mit identischer Länge und Bruchpunkten (Stankiewicz und Lupski, 2002). Spielt sich dieser Vorgang während der Meiose ab, besteht die Möglichkeit, dass es zu der Entstehung eines CNVs kommt, der für das Auftreten einer genetischen Erkrankung des Kindes verantwortlich ist. In somatischen, sich teilenden Zellen können diese gebildeten Kopienzahlveränderungen zur Krebsentstehung führen (Gu et al., 2008).

Non-Homologous End-Joining

NHEJ ist ein Reparaturmechanismus für Doppelstrangbrüche. Er findet in jeder Phase des Zellzyklus statt und ist unabhängig von LCR oder anderen Mediatoren. Beim NHEJ werden die beiden gebrochenen DNA-Enden enzymatisch modifiziert und abschließend ligiert. Während der Modifizierung kann ein größeres DNA-Stück entfernt oder doppelt synthetisiert werden, sodass an den Verbindungsstellen im Reparaturprodukt Deletionen oder Duplikationen entstehen (Gu et al., 2008; Weterings and van Gent, 2004).

Fork stalling and template switching

FoSTeS ist ein Modell, welches CNVs mit ungleichen Bruchpunkten und komplexen Strukturen erklärt. Es wird auf DNA-Replikationsfehler zurückgeführt. Während der Neusynthese bricht die Replikationsgabel ab, der Folgestrang löst sich und bindet aufgrund von Mikrohomologien an eine andere, offene Replikationsgabel in örtlicher Nähe. Der transferierte Strang wird weiter prozessiert. Bindet der Folgestrang in eine Replikationsgabel in Vorwärtsrichtung, führt dies zu einer Deletion, springt er nach hinten, kommt es zur Duplikation. Dieser Vorgang kann sich mehrere Male wiederholen (Gu et al., 2008; Lee et al., 2007). Auch Inversionen können dadurch zustande kommen.

2.2.2 CNV-Detektion

Durch die 1997 eingeführte Array-Technologie war die Identifizierung unbalancierter submikroskopischer Kopienzahlveränderungen möglich (Iafrate et al., 2004; Sebat, 2004). Es wird technisch zwischen der Array-CGH (*Comparative Genomic Hybridisation*) und den SNP-Arrays unterschieden. Bei der Array-CGH werden Proben- und Kontroll-DNA kompetitiv an eine Ziel-DNA (*Bacterial Artificial Chromosome*, BAC oder Oligonukleotide) gebunden. Die Proben- und Ziel-DNA sind dabei zur Unterscheidung unterschiedlich farbig fluoreszenzmarkiert. Anhand

Theoretische Grundlagen

von Unterschieden in der Fluoreszenzintensität zwischen der Proben- und der Kontroll-DNA können Kopienzahlveränderungen ermittelt werden. Die SNP-Arrays bestehen aus einer Oberfläche, auf der unzählige über das Genom verteilte SNP-Marker gebunden sind. Die SNPs beruhen auf HapMap-Genotypisierungsdaten. Die neuesten SNP-Arrays beinhalten > 1 Million solcher SNP-Marker mit einer Allelfrequenz von 1% (Altshuler et al., 2010; Belmont et al., 2005). Im Vergleich zur Array-CHG haben SNP-Arrays den Vorteil, dass neben den Kopienzahlveränderungen auch Informationen über den Genotyp erfasst werden, sodass auch Regionen mit kopienzahlneutralen aber homozygoten Genotypen (z.B. bei Isodisomien) erfasst werden können. Sie kommen auf eine Auflösung von bis zu 5 kb (Conrad et al., 2006; McCarroll et al., 2008; Redon et al., 2006).

Weitere Techniken zur CNV-Detektion sind: *Die Fluorescence-In-Situ-Hybridization (FISH)*, die *Multiplex Ligation-dependent Probe Amplification (MLPA)*, die quantitative PCR (qPCR) oder *Southern Blots*. Diese Techniken eignen sich jedoch nicht für genomweite hypothesenfreie Untersuchungen. Sie benötigen spezifische Sonden oder Primer (Emanuel und Saitta, 2007) und kommen daher nur zur gezielten Detektion begrenzter Genombereiche oder zur Validierung zum Einsatz. Abb. 3 zeigt die Auflösungen der unterschiedlich verfügbaren Detektionstechniken an.

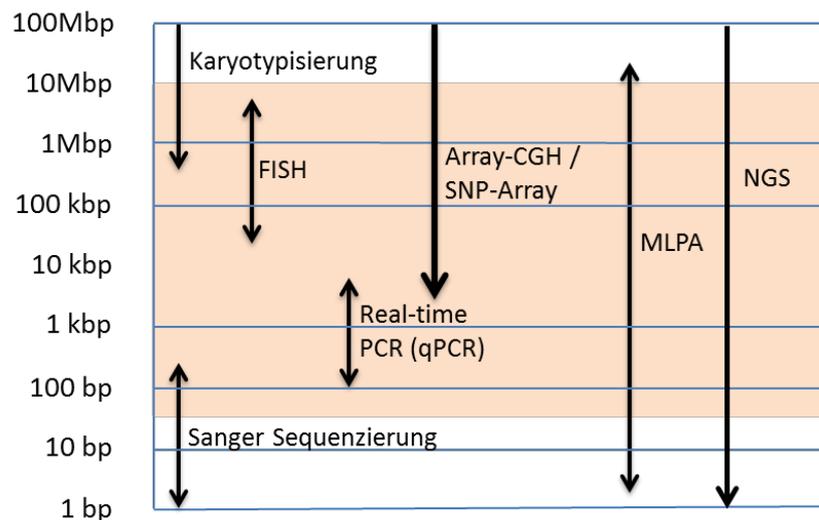


Abb. 3 Auflösung verschiedener Detektionsmethoden von genomischen Veränderungen

In Anlehnung an <http://www.fisherbiotec.com.au>

Orange: CNV-Bereich

Durch die sinkenden Preise der Hochdurchsatz-Sequenzierung in Verbindung mit verbesserten Algorithmen und besserer Abdeckung (*Coverage*) wird zukünftig das *Whole Genome Sequencing* (WGS) die Methode der Wahl werden, um alle Arten von Varianten - SNV, Indels, Strukturvarianten - auf einmal zu erfassen und somit ein komplettes Bild der genomischen Varianz zu erhalten.

Große internationale Kollaborationsprojekte wie das 1000-Genome-Projekt (TGP) veröffentlichten *Whole-Genome-Sequenzierdaten* von über tausend Individuen und machten diese in Datenbanken wie dem UCSC Genome Browser (<https://genome.ucsc.edu/>) (1000 Genomes Project Consortium et al., 2012) zugänglich. Dafür analysierten sie eine große Anzahl von Proben einer breiten Population, waren aber in ihrer Auflösung limitiert. Dadurch ist es nur möglich, häufige genetische Variationen, die mit einer Allelfrequenz von 1% in der Population auftreten, zu analysieren.

2.2.3 Interpretation von CNVs

Die phänotypischen Auswirkungen von CNVs sind sehr variabel und reichen von fehlenden bis embryonal letalen Effekten (Beckmann et al., 2007; Hurles et al., 2008). Die Übergänge dazwischen sind fließend und die Ursächlichkeit der CNVs nicht immer sicher nachzuvollziehen (Sebat, 2004). Hochpenetrante pathogene Mutationen von benignen Varianten zu unterscheiden ist daher eine komplexe Aufgabe. Neben Segregationsanalysen und funktionellen Untersuchungen bieten Frequenzeinschätzungen von CNVs eine Möglichkeit, die Pathogenität zu bewerten. Ähnlich wie SNPs handelt es sich bei häufig auftretenden CNVs nicht um hoch penetrante Mutationen, die monogenen Erkrankungen wie den erblichen Tumorsyndromen zugrunde liegen können (Kuiper et al., 2010). Seltene SNVs oder CNVs hingegen können hochpenetrant sein und innerhalb der betroffenen Familie mit dem Phänotyp segregieren (Shlien und Malkin, 2010). Die *Database of Genomic Variants* (DGV) ist eine CNV-Datenbank gesunder Individuen und enthält Einträge aus 61 Studien mit insgesamt 6.266.574 CNVs (<http://dgv.tcag.ca/dgv/app/home>, Stand: Juli 2015). Hierdurch ist eine Frequenzeinschätzung von detektierten CNVs möglich.

2.2.4 Pathomechanismus von CNVs

CNVs erzielen ihre pathogene Wirkung über verschiedene Mechanismen (vgl. Abb. 4):

- A) Komplette Gen-Deletionen in dosissensitiven Genen führen in der Regel zu einem Funktionsverlust des betreffenden Proteins, während komplette Gen-Duplikationen eine Überexpression des Proteins bewirken können.
- B) CNVs, die Gene nur partiell betreffen, können je nach Lage des Bruchpunktes zu einer Inaktivierung des Gens führen und einen Funktionsverlust des betreffenden Proteins zur Folge haben.
- C) Durch CNVs, die zwei benachbarte Gene zum Teil betreffen, können Fusionsproteine mit veränderter Funktion oder komplettem Funktionsverlust entstehen.
- D) Als Positionseffekt bezeichnet man die veränderte Expression eines Gens, häufig durch Translokation der regulatorischen Einheiten (Kleinjan und van Heyningen, 2005). Bei CNVs spricht man von einem Positionseffekt, wenn ein CNV den regulatorischen, nicht kodierenden Gen-Bereich betrifft und dieses dadurch über- oder unterexprimiert wird (Lupski und Stankiewicz, 2005). Dieser Effekt ist auch noch nachweisbar, wenn die regulatorische Einheit bis zu 1,5 Mb zum assoziierten Ziel-Gen entfernt liegt (Klopocki et al., 2008).

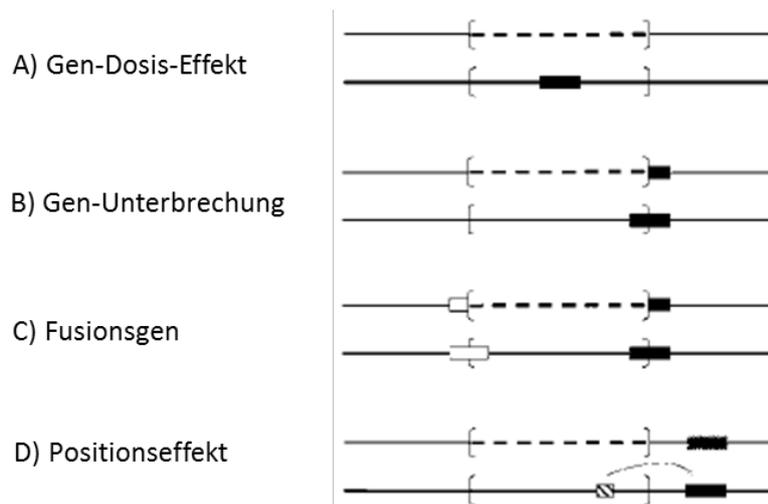


Abb. 4 Modell von CNV-Effekten

Horizontale Linien zeigen homologe chromosomale Bereiche an. Die vom CNV betroffenen Bereiche sind in Klammern gesetzt. Gestrichelte Linien zeigen CNVs (Deletion oder Duplikation) an. Gene sind durch ausgefüllte schwarze Rechtecke wiedergegeben (Lupski und Stankiewicz, 2005).

2.3 Erbllicher Darmkrebs

Nach dem Brustkrebs bei der Frau und dem Prostata- sowie Lungenkrebs beim Mann stellt das kolorektale Karzinom (CRC) in Europa die zweithäufigste Krebstodesursache beider Geschlechter dar (WHO 2015; <http://www.euro.who.int/>; RKI <http://www.rki.de/>). Etwa 70% davon sind sporadisch, d. h. sie treten meist als Einzelfälle in der Familie in höherem Lebensalter auf und haben keine monogene Ursache. Bei 20 - 30% der Erkrankten sieht man eine familiäre Häufung von Darmkrebserkrankungen (oligogen oder multifaktoriell), aber nur bei etwa 5% handelt es sich um eine monogen erbliche Form, die durch Keimbahnmutationen mit hoher Penetranz hervorgerufen werden (Abb. 4) (Hegde et al., 2014). Unter den Patienten mit Verdacht auf ein monogenes kolorektales Tumorsyndrom bleiben 30-50% ohne aufgeklärte Ursache. Die monogenen Formen werden in die verschiedenen Polyposis-Erkrankungen und den erblichen Darmkrebs ohne Polyposis (Lynch-Syndrom, LS oder *Hereditary Non-Polyposis Colorectal Cancer*, HNPCC) unterteilt. Bestimmte Polyposisformen weisen überschneidende Merkmale zum LS auf. Konnte ein LS molekulargenetisch nicht bestätigt werden, sollte dies differentialdiagnostisch berücksichtigt werden.

2.3.1 Adenom-Karzinom-Sequenz

Bei der Entstehung von Darmkrebs (sporadisch wie hereditär) kommt es zur schrittweisen Umwandlung von gesunder Darmschleimhaut über adenomatöse Polypen zum Karzinom. Dieser Vorgang wird als Adenom-Karzinom-Sequenz bezeichnet (Vogelstein et al., 1988).

Adenome sind gutartige Geschwulste aus Drüsenepithelzellen und weisen eine dysplastische Morphologie und veränderte Differenzierung der betroffenen Epithelzellen auf (Fearon, 2011). Damit aus einem Adenom ein Karzinom wird, ist eine Abfolge von somatischen Mutationen und epigenetischen Veränderungen einer Zelllinie der Darmschleimhaut notwendig. Da bei den autosomal-dominant erblichen Darmkrebsarten die initiale Mutation bereits in allen Zellen vorliegt, ist die Wahrscheinlichkeit für die Entstehung von Karzinomen erhöht.

Beim sporadischen Darmkrebs (Abb. 5a) können sowohl Mutationen in Onkogenen als auch in Tumorsuppressorgenen (TSG) und DNA-Reparaturgenen für die Ausbildung des Adenoms ursächlich sein. Bei Mutationen in Tumorsuppressorgenen bedarf es, entsprechend der „Zwei-Treffer Hypothese“, eines Verlustes des zweiten Allels, bis es zur Tumorbildung kommt.

Theoretische Grundlagen

Bei der FAP (Kap. 2.3.3) entwickeln sich aufgrund der heterozygoten Keimbahnmutation im Tumorsuppressorgen APC viele Adenome schon in jungem Alter. Nichtsdestotrotz sind noch weitere individuelle somatische Mutationen oder epigenetische Veränderungen wie Methylierungen in spezifischen Onkogenen, Reparaturgenen und TSG erforderlich, bis ein Karzinom auftritt (Abb. 5b) (Chung, 2000; Vogelstein et al., 2013). Beim HNPCC / LS (Kap. 2.3.6) kommt es wie beim sporadischen CRC durch unabhängige Mutationen zur Entstehung vereinzelter Adenome im mittleren Lebensalter Abb. 5c) (Toribara und Sleisenger, 1995). Der sich später ereignende *second hit* durch somatische Mutation des Wildtyp-Allels des entsprechenden DNA-Reparatur-Gens verursacht den Mismatch-Repair (MMR)-Funktionsverlust und hat eine Anhäufung weiterer Mutationen in den betroffenen Zellen zur Folge (Rustgi, 2007; Fishel 1995; Vilar und Gruber, 2010). Vom Adenom zum Karzinom dauert es dann meist weniger als drei Jahre (Pino et al., 2009).

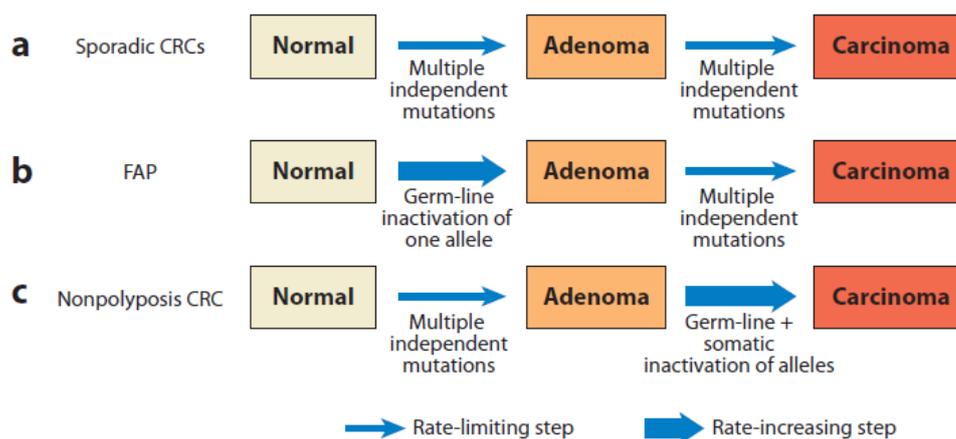


Abb. 5 Adenom-Karzinom Tumor Entwicklung

Übernommen aus (Fearon, 2011)

2.3.2 Zwei-Treffer-Hypothese nach Knudson

Die Zwei-Treffer-Hypothese nach Knudson besagt, dass bei Vorliegen einer Keimbahnmutation in einem Tumorsuppressor- oder DNA-Reparaturgen erst die somatische Inaktivierung des zweiten Allels die Krebsentstehung initiiert. Bei dem zweiten Treffer handelt es sich häufig um eine größere Deletion des entsprechenden Chromosomenabschnitts, die z.B. mittels eines *Loss of Heterozygosity* (LOH) nachgewiesen werden kann. Doch auch Punktmutationen oder epigenetische Veränderungen wie eine Promotor-Hypermethylierung des entsprechenden Gens

(z. B. für das DNA-Reparaturgen *MLH1* beschrieben) (Nagasaka et al., 2010) können den *second hit* darstellen und initiieren ggf. eine maligne Tumortransformation (Knudson, 2001).

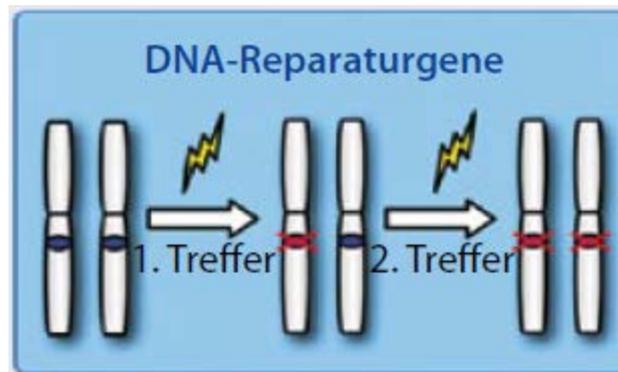


Abb. 6 Zwei-Treffer-Hypothese nach Knudson

Aus „Klinik und Genetik des familiären Darmkrebses“ (Steinke et al., 2010)

2.3.3 Familiäre adenomatöse Polyposis

Die familiäre adenomatöse Polyposis (FAP) ist ein autosomal-dominant vererbtes Tumorsyndrom und repräsentiert < 1% aller kolorektalen Karzinomerkrankungen. Sie betrifft etwa eine von 10.000 Personen in der Bevölkerung und ist damit das häufigste gastrointestinale Polyposis-Syndrom (Galiatsatos und Foulkes, 2006). Bei der klassischen FAP finden sich hunderte bis tausende adenomatöse Polypen im gesamten Dickdarm (Lynch und de la Chapelle, 2003). Die ersten Polypen können bereits im Kindesalter auftreten und entwickeln sich, sofern unbehandelt, in fast 100% der Fälle bis zum Alter von 35-40 Jahren zu Karzinomen weiter (Bisgaard et al., 1994).

Die genetische Ursache sind Mutationen im *APC*-Gen, einem Tumorsuppressorgen, welches als Antagonist im Wnt-Signalweg und für die Zellzykluskontrolle verantwortlich ist. Das mutierte *APC*-Gen aktiviert über die Akkumulation von β -Catenin mehrere Transkriptionsfaktoren und leitet so die Tumorgenese ein (Galiatsatos und Foulkes, 2006). Die Erkrankung besitzt eine nahezu hundertprozentige Penetranz (Galiatsatos und Foulkes, 2006). Mildere Verlaufsformen mit einem höheren Erkrankungsalter und einem geringeren Polypenaufreten werden als attenuierte FAP (AFAP) bezeichnet.

2.3.4 MUTYH-assoziierte Polyposis

Im Gegensatz zu der FAP / AFAP wird die MAP autosomal-rezessiv durch biallelische *MUTYH*-Mutationen vererbt (Sampson und Jones, 2009). Das Erkrankungsbild der MAP, verursacht

durch bialele *MUTYH*-Mutationen, ist sehr heterogen und weist phänotypische Überschneidungen mit dem LS auf (Aceto et al., 2005; Aretz et al., 2006; Jones et al., 2002; Morak et al., 2010; Sieber et al., 2003). So kann bei einer MAP in seltenen Fällen auch eine hohe Mikrosatelliteninstabilität (MSI-H) vorliegen (Colebatch et al., 2006; Morak et al., 2014). Die MSI-H ist hierbei jedoch ursächlich somatischen Mutationen in MMR-Genen geschuldet. Durch den hier gestörten *base-excision-repair* ist die Mutationsrate im Tumor stark erhöht. Ein HNPCC/LS mit Ausfall eines MMR-Gens kann somit vorgetäuscht werden.

2.3.5 Polymerase Proofreading-assozierte Polyposis

Spezifische Keimbahnmutationen in den Exonuklease-Domänen der Polymerase-Gene *POLE* (c.1270C>G; p.Leu424Val; NM_006231) und *POLD1* (c.1433G>A; p.Ser478Asn; NM_002691) sind seit kurzem als sehr seltene, hochpenetrante Mutationen bei Patienten mit frühmanifesten multiplen kolorektalen und duodenalen Adenomen und Karzinomen beschrieben worden (Palles et al., 2013, Spier et al 2015). Für die Erkrankung wurde der Begriff "Polymerase Proofreading-assozierte Polyposis (PPAP)" eingeführt (Briggs und Tomlinson, 2013). Zudem wurde für Mutationen in *POLD1* ein erhöhtes Risiko für Endometriumkarzinome beschrieben (Briggs und Tomlinson, 2013; Palles et al., 2013). PPAP-Patienten weisen eine große inter- und intrafamiliäre Variabilität in Bezug auf die Anzahl der Adenome und Karzinome sowie dem Erkrankungsbeginn auf. Die Tumoren wurden als mikrosatellitenstabil (MSS) und chromosomeninstabil charakterisiert (Palles et al., 2013).

Eine kürzlich erschienene Studie konnte zeigen, dass sich die bekannte Exonuklease-Domain-Keimbahnmutation in *POLE* auch bei Patienten mit MSI-H und MMR-Verlust im Tumor finden lässt. Die Mutation wurde bei 3/1188 (0,25%) Indexpatienten aus Familien mit familiärem Darmkrebs nachgewiesen (Elsayed et al., 2014). Diese Frequenz entspricht der früherer Studien (Palles et al., 2012; Valle et al., 2014). Bei HNPCC-Familien ohne genetische Ursache in den MMR-Genen sollte deshalb auch eine *POLE*-Mutation als sehr seltene Ursache in Betracht gezogen werden.

2.3.6 HNPCC und Lynch-Syndrom

Das HNPCC (Hereditary Non-Polyposis Colorectal Cancer) bezeichnet ein klinisches Syndrom, das durch folgende Kriterien definiert wird: (i) Erfüllte Amsterdam- oder (revidierte) Bethesda-Kriterien (siehe Box 1 und 2); (ii) Vorliegen des Verlusts mindestens eines Mismatch-Repair-

Proteins im Tumor oder Nachweis einer Mikrosatelliteninstabilität (MSI) im Tumor; (iii) Kein Nachweis einer ursächlichen Mutation in einem der MMR-Gene (*MLH1*, *MSH2*, *MSH6* und *PMS2*) oder im EPCAM-Gen. Einige Autoren benutzen dafür auch die Bezeichnung „Lynch Like Syndrome“ (Mensenkamp et al., 2014; Rodríguez–Soler et al., 2013). Von einem Lynch-Syndrom (LS) im engeren Sinne spricht man, wenn die ursächliche Keimbahnmutation identifiziert werden konnte.

Als weitere Unterform soll noch auf den „Familial CRC type X“ hingewiesen werden. Bei diesem mutmaßlich erblichen Syndrom (erfüllte Amsterdam-Kriterien), liegt weder ein Verlust eines MMR-Proteins im Tumor vor, noch weist der Tumor eine Mikrosatelliteninstabilität auf (Lindor et al., 2005). Die genetische Ursache wird daher abseits eines Mismatch-Repair-Defizits vermutet (Valle et al., 2007).

Im klinischen Alltag werden die Begriffe LS und HNPCC häufig synonym verwendet. In dieser vorliegenden Arbeit wird die Bezeichnung LS verwendet, wenn die ursächliche Mutation bekannt ist und HNPCC, wenn die genetische Ursache (noch) unbekannt ist.

Klinik und Klassifikation

Das LS (OMIM #120435) stellt die häufigste monogene Prädisposition für Kolonkarzinome dar und ist verantwortlich für etwa 3% aller Kolonkarzinome (Hampel et al., 2008). Man schätzt die Inzidenz in der Bevölkerung auf etwa 0,2% (1 auf 500) (Steinke et al., 2013). Frauen entwickeln in 39-50% ein Endometriumkarzinom. Das vollständige Tumorspektrum mit dazugehörigem Lebenszeitrisiko für das LS sind in Tab. 1 wiedergegeben.

Tab. 1 Tumorspektrum und Lebenszeitrisiko für LS Patienten (für alle MMR-Gene)

Tumor	Risiko für Männer	Risiko für Frauen
Kolonkarzinom	34-73%	32-59%
Endometriumkarzinom	-	39-50%
Ovarialkarzinom		7-8%
Magenkarzinom	1-6%	
Urothelgänge	2-8%	
Gallengänge	1-4%	
Dünndarm	1-4%	
Gehirn/CNS	~ 2%	
Pankreas	~ 4%	
Haut (Talgdrüse)	abhängig vom betroffenen Gen	

Daten aus dem HNPCC-Konsortium aus Steinke et al., 2013

Mutationsträger entwickeln häufig bereits in jungem Alter (vor dem 50. Lebensjahr) einen für das LS typischen Tumor. Nicht selten entstehen mehrere synchrone oder metachrone Tumoren im Laufe des Lebens.

Das mittlere Erkrankungsalter von Mutationsträgern liegt mit 45 Jahren deutlich vor dem mittleren Erkrankungsalter sporadischer Kolonkarzinome, welche meist erst mit über 60 Jahren auftreten (Toribara und Sleisenger, 1995). Insgesamt ist der Krankheitsverlauf allerdings sehr variabel und dann zum Teil klinisch nur schwer von sporadischen Fällen zu unterscheiden (Hampel et al., 2005; Watson et al., 2008). Andererseits können sporadische CRC in einzelnen Fällen auch sehr früh auftreten (Mensenkamp et al., 2014); die Unterscheidung zwischen erblichen und nicht-erblichen Fällen ist daher klinisch oft schwierig.

International anerkannte klinische Kriterien sollen bei der Identifikation von Patienten mit einem HNPCC bzw. LS helfen und einen Hinweis geben, wann eine Untersuchung auf HNPCC-typische Veränderungen im Tumorgewebe sinnvoll ist. Bei zusätzlich auffälligem Resultat der Tumoruntersuchung wird die Keimbahnmutationssuche eingeleitet. Des Weiteren werden den Patienten und ihren Angehörigen engmaschige Vorsorgemaßnahmen empfohlen. Die klinischen Kriterien (Amsterdam- und Bethesda-Kriterien) sind in Box 1 und 2 dargestellt. Sie beziehen im Wesentlichen die folgenden drei Punkte mit ein:

1. Frühes Erkrankungsalter (< 50. Lebensjahr)
2. Synchrone und / oder metachrone Tumoren aus dem HNPCC-Spektrum
3. Positive Familienanamnese

Amsterdam-II-Kriterien

Alle Kriterien müssen zutreffen:

1. Mindestens drei Familienangehörige mit histologisch gesichertem kolorektalem Karzinom oder einem Karzinom des Endometriums, Dünndarms, Ureters oder Nierenbeckens, davon einer mit den beiden anderen erstgradig verwandt; FAP muss ausgeschlossen sein
2. Wenigstens zwei aufeinander folgende Generationen betroffen
3. Diagnosestellung bei mindestens einem Patienten vor dem Alter von 50 Jahren

Box 1 Amsterdam-II-Kriterien (Vasen et al., 1999)

Revidierte Bethesda-Kriterien

Mindestens eines der genannten Kriterien muss erfüllt sein:

- a) Patienten mit kolorektalem Karzinom vor dem 50. Lebensjahr
- b) Patienten mit synchronen oder metachronen kolorektalen Karzinomen oder anderen HNPCC-assoziierten Tumoren*, unabhängig vom Alter
- c) Patienten mit kolorektalem Karzinom mit MSI-H-Histologie** vor dem 60. Lebensjahr
- d) Patient mit kolorektalem Karzinom (unabhängig vom Alter), der einen Verwandten 1. Grades mit einem kolorektalem Karzinom oder einem HNPCC-assoziierten Tumor vor dem 50. Lebensjahr hat
- e) Patient mit kolorektalem Karzinom (unabhängig vom Alter), der mindestens zwei Verwandte 1. oder 2. Grades hat, bei denen ein kolorektales Karzinom oder ein HNPCC-assoziiertes Tumor (unabhängig vom Alter) diagnostiziert wurde

* Zu den HNPCC-assoziierten Tumoren gehören Tumoren in: Kolorektum, Endometrium, Magen, Ovarien, Pankreas, Urothel, Gallengang, Dünndarm, Gehirn (meist Glioblastom), Talgdrüsenadenome und Keratoakanthome

**Hohe Mikrosatelliteninstabilität (MSI-H) bei Vorliegen von Tumor-infiltrierenden Lymphozyten, Crohn-ähnlicher lymphozytärer Reaktion, muzinöser / Siegelring-Differenzierung oder medullärem Wachstumsmuster

Box 2 Revidierte Bethesda-Kriterien (Umar et al., 2004)

Genetik des Lynch-Syndroms

Das LS wird wie die FAP autosomal-dominant vererbt (Wiederholungsrisiko für erstgradig Verwandte von 50%). Im Tumor lässt sich der Funktionsverlust eines MMR-Proteins nachweisen. Ursächlich hierfür sind meist heterozygote pathogene Keimbahnmutationen in einem der vier ursächlich bekannten MMR-Gene: *MSH2*, *MSH1*, *MSH6* oder *PMS2* (Lynch et al., 2015). Daneben führen auch Deletionen der letzten beiden Exons von *EPCAM*, welches oberhalb (*upstream*) von *MSH2* liegt, über eine epigenetische Inaktivierung von *MSH2* zu der Erkrankung. Dies geschieht nur in Zellen, in denen *EPCAM* exprimiert wird. Dadurch kommt es zu einem mosaikartigen *MSH2*-Inaktivierungsmuster (Ligtenberg et al., 2009).

Neben der Beteiligung bei der Apoptose und Zellzykluskontrolle besteht die Hauptaufgabe der MMR-Proteine darin, DNA-Replikationsfehler zu korrigieren (Bonis et al., 2007; Jiricny, 2006).

Die Proteine arbeiten als Heterodimere zusammen: MSH2 geht hierbei sowohl mit MSH6 (MutS- α), als auch mit MSH3 (MutS- β) einen Komplex ein (Acharya et al., 1996). Diese beiden MutS-Heterodimere sind für die Detektion der Fehlpaarungen zuständig, wobei MutS- α Einzelnukleotid-Substitutionen und kleinere Insertions-Deletions-Schleifen erkennt und bindet und MutS- β eine stärkere Affinität für größere Insertions-Deletions-Schleifen mit bis zu 10 Nukleotiden besitzt (Boland und Goel, 2010). Für die komplette Reparatur der fehlgepaarten Nukleotide lagert sich in einem darauffolgenden Schritt ein weiteres Heterodimer MutL- α (MLH1-PMS2) an. Mit Hilfe zusätzlicher Proteine wie EXO-1, PCNA und der DNA-Polymerase, wird der DNA-Tochterstrang bis zur Fehlpaarung herausgeschnitten und resynthetisiert. MLH1 ist auch in der Lage Heterodimere mit PMS1 und MSH3 einzugehen, deren Funktionsweisen jedoch nicht tiefergehend untersucht wurden (Boland und Goel, 2010).

Die beiden Proteine MSH2 und MLH1 übernehmen in ihren Komplexen jeweils die zentrale Aufgabe. Zellen, welche einen Expressionsverlust von MSH2 oder MLH1 aufweisen, zeigen keine MMR-Aktivität mehr. Bei dem Verlust der MSH6- oder PMS2-Expression ist die MMR-Aktivität nur vermindert. Patienten mit inaktivierenden *MSH6*- oder *PMS2*-Mutationen zeigen daher meist einen milderen Phänotyp mit verminderter Penetranz (Boland et al., 2008). Die Funktion von *MSH3* und *PMS1* ist noch nicht vollständig geklärt (Boland und Goel, 2010). Eine Assoziation mit dem LS konnte für *MSH3* und *PMS1* bisher nicht nachgewiesen werden (Boland et al., 2008; Boland und Goel, 2010). Kürzlich wurden compound-heterozygote Keimbahnmutationen in *MSH3* mit einer rezessiv vererbten Polyposis-Erkrankung assoziiert (Adam et al., 2016).

Routine-Diagnostik des HNPCC

Liegt der Verdacht eines LS vor, wird der Tumor zunächst auf Mikrosatelliteninstabilität (MSI) sowie Expressionsverlust eines MMR-Proteins mittels Immunohistochemie (IHC) untersucht.

Mikrosatelliten sind kurze, über das Genom verteilte Nukleotidsequenzen von 1-5 bp, die sich tandemartig 8-25-mal wiederholen (Tautz und Schlötterer, 1994). Eine Mikrosatelliteninstabilität liegt vor, wenn im Vergleich zwischen normalen Zellen und Tumorzellen Längenunterschiede der Mikrosatelliten bestehen. Eine hohe Mikrosatelliteninstabilität (MSI-H) ist das Kennzeichen eines MMR-Defekts (Lynch und de la Chapelle, 2003). Dessen ungeachtet weisen etwa 5% aller bestätigten LS-Patienten keine MSI-H auf (Hendriks et al., 2004).

Kolorektale Karzinome von Patienten mit LS zeigen in 95% eine MSI-H (Aaltonen et al., 1993; Peltomäki et al., 1993). Der genaue Mechanismus der Entstehung von MSI bei MMR-Verlust ist in Abb. 7 wiedergegeben.

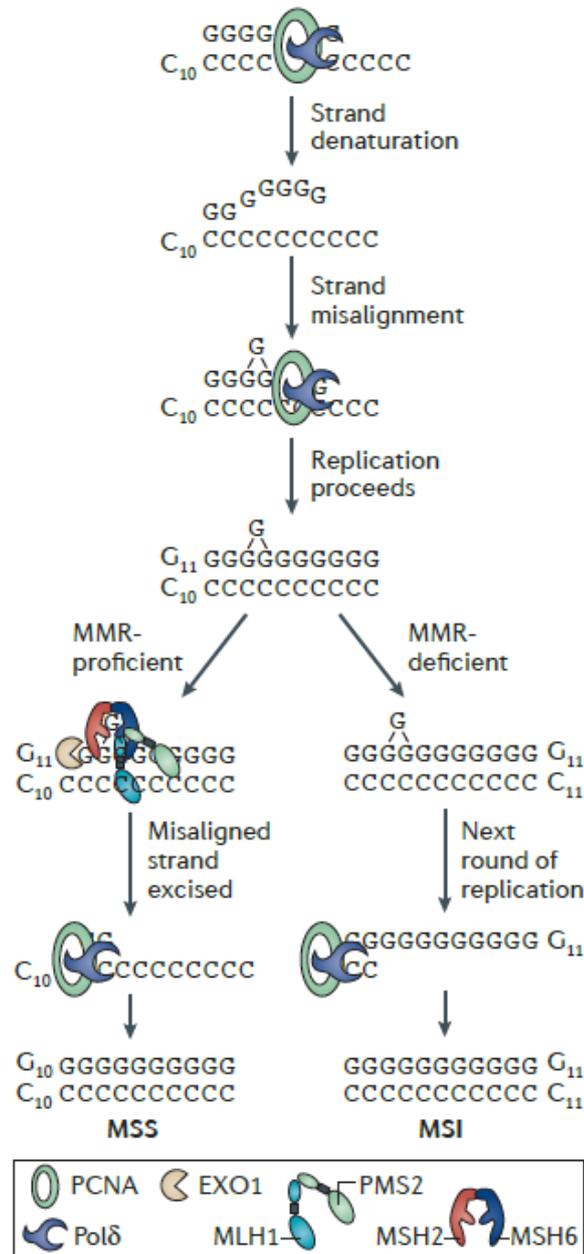


Abb. 7 Molekularer Mechanismus der MSI

Während der Replikation der Mikrosatelliten-Region (C10) kommt es häufiger zu Fehlern der Polymerase und es entstehen zu lange (oder zu kurze) Fragmente (G11). Bei funktionierender MMR-Aktivität wird das Extra-Nukleotid durch die MMR-Heterodimere MutS (*MSH2-MSH6*) und MutL (*MLH1-PMS2*) herausgeschnitten. In Zellen mit gestörter MMR-Funktion bleibt das Extra-Nukleotid jedoch bestehen und wird bei wiederholter DNA-Replikation fest in den neuen Strang eingebaut (C11). EXO1: Exonuklease 1, MSI: Mikrosatelliteninstabilität, MSS: Mikrosatellitenstabilität, PCNA: *Proliferating Cell Nuclear Antigen*, Pol δ: DNA Polymerase-δ (Lynch et al., 2015).

Für die immunohistochemische Untersuchung werden markierte Antikörper gegen die MMR-Proteine verwendet. Zellen mit zwei defekten Kopien eines MMR-Gens weisen keine Expression des zugehörigen Gens auf, entsprechend fällt die immunhistochemische Färbung negativ aus. Die immunohistochemische-Testung aller vier MMR-Proteine kann somit Hinweise liefern, welches der MMR-Gene am wahrscheinlichsten von einer pathogenen Mutation betroffen ist und schränkt damit die Suche nach pathogenen Keimbahnmutationen ein (Giardiello et al., 2014). Bei ursächlichen *MSH2*-Mutationen fällt immer auch MSH6 im Tumorgewebe mit aus, da MSH2 der einzige Bindungspartner von MSH6 ist. Bei Mutationen im *MSH6*-Gen hingegen bleibt die MSH2-Expression üblicherweise erhalten, da MSH2 noch weitere Bindungen mit MSH3 eingehen kann (Lynch et al., 2015).

Alternative Gründe für den MSH2-Verlust

In den allermeisten Fällen kommt es entsprechend der Zwei-Treffer-Hypothese nach Knudson (Kap. 2.3.2) zum MSH2-Verlust bei Vorliegen einer pathogenen Keimbahnmutation im *MSH2*-Gen und der anschließenden somatischen Inaktivierung des zweiten Allels. Nachfolgend werden weitere mögliche Ursachen für das Vorliegen des MSH2-Proteinverlustes im Tumorgewebe vorgestellt.

1. In der Routinediagnostik nicht identifizierte Mutationen

Zu den Mutationen, die durch die gängigen Methoden (Sanger, NGS, Deletions-Screening) in der Routinediagnostik bisher nicht erfasst werden, gehören tief-intronische Mutationen im Promotorbereich, somatische Mosaik und komplexe strukturelle Veränderungen wie Inversionen. Tief-intronische Mutationen können zu der Generierung einer neuen Spleißstelle führen. Die dabei gebildeten Pseudoexons sind durch eine Transkriptanalyse nachweisbar. Eine ursächliche, tief-intronische Mutation wurde bisher in einer australischen Familie nachgewiesen. Die Mutation (c.212-478T>G) befindet sich im Intron 1 des *MSH2*-Gens und verursacht die Aktivierung einer kryptischen Spleißstelle mit Einschluss eines Pseudoexons in die RNA (Clendenning et al., 2011). Eine 10 Mb große Inversion, welche die Exons 1-7 von *MSH2* einschließt, wurde von zwei unabhängigen Arbeitsgruppen beschrieben (Rhees et al., 2014; Wagner et al., 2002).

2. Epigenetische Veränderungen

Angeborene epigenetische Mutationen in MMR-Genen durch monoallelische Methylierung und posttranskriptionellem Expressionsverlust sind eine sehr seltene alternative Ursache für ein LS in MMR-Gen mutationsnegativen Fällen (Lynch et al., 2015). Die allermeisten dieser bei *MLH1* berichteten Epimutationen entstehen *de novo* (Hitchins et al., 2005). Ob die epigenetische Programmierung eines Menschen vererbbar ist, wurde lange Zeit kontrovers diskutiert (Chong und Whitelaw, 2004). In der Regel wird das epigenetische Muster in der frühen Embryogenese reprogrammiert, sodass eine direkte Vererbung ausgeschlossen ist (Lee et al., 2002). Die Transmission einer *MSH2*-Epimutation wurde bisher einmalig berichtet. Das Methylierungsmuster im Bereich des *MSH2*-Promotors wurde in drei nachfolgenden Generationen nachgewiesen. Zwei SNPs im Bereich des *MSH2*-Promotors konnten mit der hereditären Epimutation assoziiert werden (Chan et al., 2006).

3. Posttranskriptionelle Degradierung von MSH2

Ein interessanter Interaktionsmechanismus zwischen Proteinen des Proteasom-Pathways und *MSH2* ist bei humoralen Krebszellen beschrieben worden. Demnach führen somatische Deletionen von *PRKCZ* zu einer messbaren Degradierung von *MSH2* mit *MSH2*-Ausfall im Tumorgewebe (Diouf et al., 2011; Hernandez-Pigeon et al., 2005). Darüber hinaus konnte gezeigt werden, dass auch Deletionen weiterer Gene des Signalwegs (*MTOR*, *HERC1*, *PRKCZ* oder *PIK3C2B*) diesen Effekt auf *MSH2* ausüben. Ein gezielter Knockdown dieser Gene rekapituliert den *MSH2*-Degradierungsvorgang in Leukämiezellen (Diouf et al., 2011). Ob auch *Loss of Function* (LoF)-Mutationen der Keimbahn der o.g. Proteasom-Gene bei Patienten mit HNPCC einen *MSH2*-Verlust im Tumor bewirken können, ist eine bisher noch nicht überprüfte Hypothese. Abb. 8 zeigt die Interaktion der Proteasom-Pathway-Gene und ihre Verbindung zu *MSH2*.

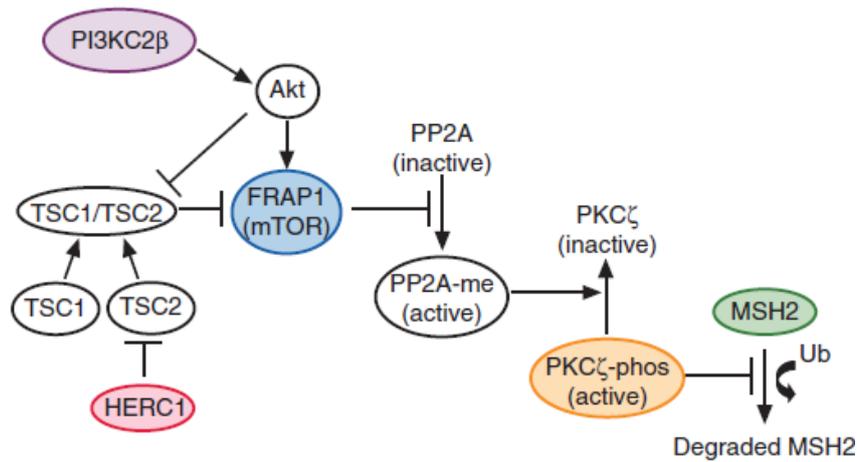


Abb. 8 Somatische MSH2-Degradierung über den Proteasom-Pathway

Übernommen aus (Diouf et al., 2011)

4. Biallele somatische Mutationen von *MSH2* (Phänokopien)

Treten zwei biallele somatische Ereignisse in *MSH2* auf, so kommt es zum *MSH2*-Verlust im Tumor mit MSI, der leicht mit dem HNPCC verwechselt werden kann. Man spricht daher auch von „Phänokopie“. Die betroffenen Patienten zeigen häufig einen milderen Verlauf der Erkrankungen und ein späteres Auftreten der Tumore. Drei Studien kamen zu dem Ergebnis das dieses Phänomen recht häufig auftritt und > 50% der Patienten mit Verdacht auf ein LS ausmacht (Geurts-Giele et al., 2014; Haraldsdottir et al., 2014; Mensenkamp et al., 2014).

5. Verlust beider Bindungspartner

Ein weiterer theoretisch denkbarer Mechanismus bei dem es zu einem *MSH2*-Verlust ohne Vorliegen einer ursächlichen Mutation kommen kann wäre, wenn *MSH2* beide Bindungspartner *MSH3* und *MSH6*, z.B. durch pathogene Mutationen in *MSH3* und *MSH6*, verliert.

2.4 Ziel der Arbeit

Bei bis zu 50% der Patienten welche die klinischen Kriterien für das LS erfüllen und einen MMR-Defekt im Tumor aufweisen, lässt sich keine ursächliche Keimbahnmutation in den bekannten MMR-Genen nachweisen. Das primäre Ziel dieser Arbeit war daher die Suche nach neuen ursächlichen Genen beim genetisch ungeklärten HNPCC mit *MSH2*-Ausfall. Hierfür wurde eine genomweite CNV-Analyse an einem Kollektiv von 95 HNPCC-Patienten durchgeführt. Umfangreiche Filterschritte wurden angewandt, um seltene und damit potentiell pathogene

CNVs in neuen Kandidatengenem aufzufinden. Die anschließende Sequenzierung der Kandidatengene diente der Identifizierung zusätzlicher Punktmutationen. Um die Relevanz der Gene weiterhin zu erforschen wurde eine Segregationsanalyse, eine Netzwerk- und Pathway-Analyse durchgeführt.

Ergänzende Untersuchungen betrafen in der Literatur beschriebene alternative Ursachen für den MSH2-Verlust wie eine Degradierung von MSH2 durch Mutationen bestimmter Gene des Proteasom-Biosynthesewegs oder eine zusätzliche Untersuchung auf seltene in der Literatur beschriebene *MSH2*-Mutationen (eine spezifische große Inversion und eine tief intronische Mutation).

Die Aufklärung der Pathophysiologie der Erkrankung hat eine wichtige Bedeutung für die Versorgung der Patienten. Erst wenn die ursächlichen Mechanismen bekannt sind, kann den Betroffenen eine sichere Diagnose gestellt werden, das exakte Wiederholungsrisiko für Verwandte berechnet, und eine prädiktive genetische Testung von Familienangehörigen ermöglicht werden. Außerdem wird durch das Aufdecken der genetischen Ursachen ein tiefgreifendes Verständnis der biologischen Grundlagen ermöglicht.

3 Material und Methoden

3.1 Patienten

Das Patientenkollektiv umfasst 95 nicht-verwandte Indexpatienten des Deutschen HNPCC-Konsortiums und einige Patienten eines niederländischen Kollektivs. Sie wurden in den Jahren 2002 - 2012 am Institut für Humangenetik Bonn oder in den anderen in Tab. 1 aufgelisteten Zentren rekrutiert. Die Studie wurde von den lokalen Ethikkommissionen aller teilnehmenden Zentren angenommen und in Übereinstimmung mit den Vorgaben der Deklaration von Helsinki durchgeführt.

Die Bedingungen für eine Teilnahme waren:

- (i) das Vorliegen einer schriftlichen Einverständniserklärung für die Teilnahme am Forschungsprojekt
- (ii) die Erfüllung der klinischen Verdachtskriterien für das HNPCC (revidierte Bethesda-Kriterien, vgl. Kap. 2.4.1)
- (iii) einen *MSH2/MSH6*-Verlust im Tumor, nachgewiesen mittels Immunohistochemie*
- (iv) keine nachgewiesene ursächliche Keimbahnmutation in *MSH2* und *EPCAM*

* Zwei Patienten des Kollektivs zeigten keinen *MSH2/MSH6*-Verlust im Tumor. Bei einem dieser Patienten konnte jedoch nur ein Adenom untersucht werden, weshalb die Untersuchung nur eingeschränkte Aussagekraft hat. Der zweite Patient weist eine hohe Mikrosatelliteninstabilität im Tumorgewebe auf, was für einen MMR-Defekt spricht. Beide Patienten erfüllen die Amsterdam-Kriterien und waren somit interessante Kandidaten für die genomweite CNV-Analyse

Eine hohe Mikrosatelliteninstabilität ist ein weiteres Merkmal des LS, es wurden jedoch auch Patienten mit niedriger Mikrosatelliteninstabilität (MSI-L) und mikrosatellitenstabile Fälle (MSS) mit aufgenommen, sofern sie die klinischen Kriterien erfüllten und einen Ausfall von *MSH2/MSH6* im Tumor aufwiesen.

Tab. 2 Zusammensetzung des Kollektivs

Zentrum	Anzahl Patienten	m	w	A	B	pos. FA*	IHC-MSH2	MSI-H	VUS MSH2
Bonn	66	41	25	12	54	36	64	54	7
Bochum	8	4	4	1	7	8	8	6	1
Dresden	7	5	2	1	6	5	7	7	1
Düsseldorf	1	1	-	-	1	-	1	-	-
Regensburg	2	-	2	1	1	1	2	2	-
München	7	3	4	1	6	2	7	7	-
Leiden, NL	4	3	1	-	4	4	4	2	-
Summe	95	57	38	16	79	56	93	78	9

m: männlich, w: weiblich, A: Amsterdam-Kriterien erfüllt; B: Bethesda-Kriterien erfüllt, IHC: Immunohistochemie (Ausfall von MSH2), FA: Familienanamnese; MSI-H: hohe Mikrosatelliteninstabilität; VUS: Varianten mit unklarer funktioneller Relevanz (*Variants of Unknown Significance*)

*pos. FA: erstgradig Verwandter mit Tumor des HNPCC-Spektrums

3.2 Kontrollen

Die verwendeten Heinz-Nixdorf recall (HNR)-Kontrollen stammen aus einer Längsschnittstudie des Universitätsklinikums Essen. Die Studie ermittelt den populationsbasierten Einfluss von Risikofaktoren für kardiovaskuläre Erkrankungen. Das Gesamtkollektiv umfasst über 4.500 gesunde Frauen und Männer im Alter zwischen 45 und 75 Jahren (Altersdurchschnitt: 60 Jahre). (Schermund et al., 2002). Aus diesem Gesamtkollektiv wurden 1320 Kontrollen zufällig ausgewählt. Sie sind alle deutscher Abstammung und wurden auf dem gleichen SNP-Array (HumanOmniExpress-12, Illumina, HG 19) wie unsere Patienten genotypisiert.

3.3 Materialien

3.3.1 Geräte und technische Ausrüstungen

Pipetten	<ul style="list-style-type: none"> – Multipette® Xstream, Eppendorf AG – Research Plus: 2,5-20 µl; 10-100 µl; 20-200 µl; 100-1000 µl – Research Pro: 1-Kanal und 8-Kanal: 0,5-10 µl, 5-100 µl
Pippetier Roboter	<ul style="list-style-type: none"> – Biomek NXP Laboratory Automation Workstation – Beckman Coulter GmbH
96 Well PCR Platten	<ul style="list-style-type: none"> – Corning Costar Thermowell, Fisher Scientific – Framestar, 4titude – Hard-Shell Plate (HSP), BioRad – MicroAmp, Applied Biosystems / Life Technologies – MIDI Platte, Thermo Scientific – Multiply-PCR Plate, Sarstedt

Material und Methoden

	<ul style="list-style-type: none">– MultiScreen, Merck Millipore
384 well qPCR-Platten	Microtiterplatte, Applied Biosystems, Life Technologies
Platten Adhesivfolien	<ul style="list-style-type: none">– Microseal 'A', BioRad– Microseal 'B', BioRad– Adhesive Plate Seal, 4titude– PCR seal, 4titude
Reagenzien Reservoir	25 µl, Thermo Scientific
Genotypisierungssystem	BeadStation 500GX System, Illumina
Genotypisierungs - Scanner	i-scan, Illumina
Vortexer	<ul style="list-style-type: none">– VELP® scientificate– Vortex Genie 2, Scientific Industries
Zentrifugen	<ul style="list-style-type: none">– Centrifuge, 5430, 5430R, 5810R Eppendorf AG– Heraeus™ Megafuge 1.0, Thermo Scientific– Heraeus™ Biofuge fresco™ Microzentrifuge, ThermoScientific
DNA Aufkonzentrations- Filtersystem	Amicon Ultrazentrifuge 0,5 ml, Merck Millipore
Waagen	TE3102S/ TE 3135-DS, Sartorius AG
Thermocycler	PTC-225, MJ Research Inc., Peltier Thermal Cycler
Nukleinsäure-Messgeräte	<ul style="list-style-type: none">– 2100 Bioanalyzer, Agilent Technologies– NanoDrop™ 2000 Thermo Scientific Inc. PeqLab Biotechnologie GmbH– Qubit® Fluorometric Quantitation 2.0, Life Technologies
Sequenzierer	<ul style="list-style-type: none">– Genetic Analyzer 3500XL– Genetic Analyzer 3130XL– Applied Biosystems, Life Technologies– MiSeq 2000, Illumina
qPCR – Gerät	ViiA™ 7 Real-Time PCR System, Applied Biosystems™
Schüttel-, Mix-, Magnet- und Heizgeräte	<ul style="list-style-type: none">– High Speed Micro Plate Shaker, VWR– MS 3 digital Kleinschüttler, IKA®– Magnetic stand-96, Ambion– Thermomixer comfort, Eppendorf AG– Tru Temp Microheating System, Illumina

3.3.2 Reagenzien und Kits

Agilent DNA 1000 Kit	Agilent Technologies
BigDye® Terminator v1.1	Applied Biosystems, Life Technologies
BigDye® Removal: DyeEx 2.0 Spin Kit (250)	Qiagen
Hi-Di™ Formamide	Applied Biosystems, Life Technologies
HotStarTaq DNA Polymerase (Master Mix)	Qiagen GmbH
Infinium II Whole-Genome Genotyping Kit	Illumina

Lichrosolv® destilliertes Wasser zu Chromatographiezwecken	Merck Millipore
MS-102-2002 MiSeq Reagent Kit v2 (300 cycles)	Illumina
Nextera® Rapid Capture Custom Enrichment Kit; 48 samples	Illumina
Power SYBR®-Green I Master Mix	Applied Biosystems, Life Technologies
Qubit® 2.0 NGS Starter Kit	Life Technologies
SALSA® MLPA® PCR Kit, Reagent kit SALSA® MLPA® probemix	MRC-Holland
– P003 MLH1/MSH2	
– P072 EPCAM	
Sephadex™ G-50 Superfine	GE-Healthcare
TE-Puffer	Invitrogen

3.3.3 Chemikalien

- Agencourt, AMPureXP, Beckman Coulter
- Diethyldicarbonat (DEPC), PanReac, AppliChem
- Destilliertes Wasser für Chromatographiezwecke
- Ethanol, PanReac, AppliChem
- Etylendiamintetraessigsäure reinst (EDTA), Merck kGaA
- KHCO₃, Merck kGaA
- NaCl, Merck kGaA
- NaOH, Merck kGaA
- NH₄Cl, Chemsolute, Th. Geyer GmbH
- TE-Puffer, Life Technologies
- Tris-Cl, Fluka Biochemika
- Tween 20, Roche Diagnostics GmbH

3.3.4 Lösungen

- EDTA [1 mM], 0,1% DEPC: 500 µl 0,5 M EDTA, 250 µl DEPC, ad 250 ml Wasser
- Ethanol 70%: 700 mL reines Ethanol auf 1 l destilliertes Wasser
- Ethanol 80%: 800 mL reines Ethanol auf 1 l destilliertes Wasser
- Frischblutlysispuffer: 155 mM NH₄Cl, 10 mM KHCO₃, 0,1 mM EDTA
- Kernlysispuffer (NaCl-EDTA-Puffer): 75 mM NaCl, 24 mM EDTA pH 8,0
- NaOH 0,2 N: 200 µl NaOH 1,0 N, 800 µl Wasser für Chromatographiezwecke

Material und Methoden

- TE-Puffer (Tris-EDTA TE-4): 10 mM Tris-Cl, 0,1 mM Na-EDTA, pH 8,0
- Tris-Cl 10 mM, ph 8,5: 20 µl Tris-Cl 1,0 M, 1980 µl destilliertes Wasser
- Tween 20 0,5%: 5 µl Tween 20, 995 µl destilliertes Wasser

3.3.5 Software und Datenbanken

Software

3500 Data Collection Software	Applied Biosystems, Life Technologies
BaseSpace	Illumina
Biomek® Software	Beckman Coulter
Bioanalyzer-2100 Expert Software vB.02.08	Agilent Technologies
Coffalyser	MRC-Holland
DesignStudio	Illumina
Excel 2010	Microsoft Corporation
GenomeStudio V2011.1 (Genotyping v. 1.9.4)	Illumina
IGV (Integrative Genomics Viewer) v. 2.3.52 (67)	Broad Institute
Illumina Experiment Manager	Illumina
MiSeq Control Software	Illumina
NanoDrop 2000/2000c Software v.1.5	Thermo Fisher Scientific, Peqlab Biotechnologie
PLIGU (Patienten- und Laborinformationssystem für genetische Untersuchungen), Microsoft Access 2003	Microsoft Corporation
Power Point 2010	Microsoft Corporation
qPCR Analysis VIIA™ 7 Software v. 1.2	Applied Biosystems / Life Technologies
QuantiSNP v2	University of Oxford
SeqMan Pro v. 9.0 (2011)	DNAStar® Inc. Lasergene
Sequencing Analysis Software v. 5.4	Applied Biosystems, Life Technologies
Sequence Pilot Software v. 4.1.2	JSI medical systems
VariantStudio v. 2.2	Illumina
Word 2010	Microsoft Corporation

Datenbanken

Database of Genomic Variants (DGV)	http://dgv.tcag.ca/dgv/app/home
Ensembl, Welcome trust SANGER Institute	http://www.ensembl.org/index.html
Exome Aggregation Consortium (ExAC) v0.3	http://exac.broadinstitute.org/
Exome Variant Server (EVS)	http://evs.gs.washington.edu/EVS/

Genecards, Weizmann Institute of Science	http://www.genecards.org/
Human Splicing Finder v3.0	http://www.umd.be/HSF3/
LOVD, Leiden Open Variation Database v.3.0	http://www.lovd.nl/3.0/home
MutationTaster	http://www.mutationtaster.org/
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov
PolyPhen-2 v2.2.2	http://genetics.bwh.harvard.edu/pph2/
Primer3 v. 0.4.0	http://bioinfo.ut.ee/primer3/
PROVEAN Craig Venter Institute	http://provean.jcvi.org/
SIFT, Craig Venter Institute	http://sift.jcvi.org/
STRING v.9	http://www.string-db.org
The Cancer Genome Atlas (TCGA)	https://cancergenome.nih.gov/
USCS Genome Browser, University Santa Cruz	http://genome.ucsc.edu/
1000 Genomes-Projekt	http://browser.1000genomes.org

3.4 Primer Design

Für die Auswahl der Oligonukleotid-Primer wurde das öffentlich zugängliche Onlineprogramm Primer3 verwendet (<http://bioinfo.ut.ee/primer3-0.4.0/>) (Rozen und Skaletsky, 1999). Tab. 3 gibt die verwendeten Einstellungen zur Auswahl der Primer durch Primer3 wieder. Alle Primer wurden von der Firma Metabion bezogen.

Tab. 3 Allgemeine Vorgaben zur Auswahl der Primer

Produktgröße	400-1000 bp (PCR); 110-150 bp (qPCR)
Primer Größe	min 18, opt. 20, max. 24 Basen
Primer Temperatur	min 57 °C, opt. 60 °C, max. 62 °C
Primer GC-Anteil	min. 20%, opt. 46%, max. 60%

Die erhaltenen Oligonukleotide wurden auf Homologien in anderen Regionen getestet. Dazu wurden die Programme „In-Silico PCR“ und „BLAST“ des UCSC Genome Browsers verwendet (Kent et al., 2002) (<http://genome.ucsc.edu/>). Es wurden nur Primer verwendet, die an eine spezifische Sequenz im Genom binden. Alle in dieser Arbeit verwendeten Primer sind im Anhang in Tab. A.1 bis Tab. G.1 aufgelistet.

3.5 DNA Isolierung

Die DNA-Extraktion aus Patientenblutproben erfolgte im Rahmen der allgemeinen Diagnostik mit der nach Miller entwickelten Aussalzmethode (Miller et al., 1988). Hierfür wurden 10 mL venöses Vollblut verwendet und im ersten Schritt einer Zellyse unterzogen. Ein anschließender Proteinverdau mit Proteinase K setzt die Nukleinsäure frei, die im nachfolgenden Schritt durch Isopropanol gefällt werden kann.

3.6 Konzentrations- und Reinheitsmessung

Zur Konzentrations- und Reinheitsmessung von Nukleotiden (DNA oder einzelsträngige Oligonukleotide), wurde das NanoDrop™ 2000 Spectralphotometer verwendet. Etwa 1,5 µl Probenlösung werden auf den unteren Messfühler des Geräts aufgetragen. Mittels des Lambert-Beerschen Gesetzes (Tab. 4) wird die Konzentration anhand der Extinktion (optische Dichte) berechnet. Sie beschreibt das Verhältnis von Lichtintensität der Referenzprobe (Puffer) zur Probenlösung (DNA in Puffer).

Tab. 4 Lambert-Beersches Gesetz zur Messung der optischen Dichte

$c = (A\lambda \times \epsilon\lambda) / b$
<hr/>
c = Nukleinsäurekonzentration [ng/µl]
Aλ = Extinktion [AU]
ελ = Extinktionskoeffizient in Abhängigkeit zur Wellenlänge [ng x cm-1 / µl]
b = die Länge zwischen oberem und unterem Messfühler in cm

Neben der Konzentration kann auch die Reinheit der Nukleinsäure bestimmt werden. Dazu wird die Extinktion bei 260 nm (Absorptionsmaximum von Nukleinsäuren) und 280 nm (Absorptionsmaximum von Phenolen, Proteinen und anderen Verunreinigungen) gemessen und ins Verhältnis gesetzt. Ein Verhältnis 260 nm / 280 nm ≥ 1,8 wird als rein für DNA, ein Wert von mindestens 2,0 als rein für RNA, angesehen. Je kleiner der Wert, desto größer die Verunreinigung.

3.7 Sanger-Sequenzierung

Die Sanger-Sequenzierung gliedert sich in drei Arbeitsschritte. Zunächst wird die Ziel-DNA-Sequenz durch die Polymerase-Kettenreaktion (*Polymerase Chain Reaction*, PCR) (3.7.1) amplifiziert. Im daran anschließenden *Cycle Sequencing* (3.7.2), dem entscheidenden Schritt der

Sanger-Methode, werden fluoreszenzmarkierte Didesoxynukleotidtriphosphate in die Sequenz eingebaut und die Sequenz hierdurch zum Abbruch gebracht. Im letzten Schritt werden die Didesoxynukleotide per Laserfluoreszenz detektiert (3.7.3). Zwischen jedem Vorgang ist ein Waschschrift (0) zur Entfernung der Enzyme, Puffer sowie überschüssiger Nukleotide durchzuführen.

3.7.1 Polymerase-Kettenreaktion

Die PCR stellt eine *in vitro*-Methode zur zielgerichteten Vervielfältigung der DNA dar (Mullis und Faloona, 1987). Die Reaktion erfolgt in drei Schritten und wird zyklisch bis zu 30-mal wiederholt. Bei jedem Schritt wird der zu amplifizierende Bereich verdoppelt, das PCR-Produkt wächst somit exponentiell.

Im ersten Schritt wird der DNA-Doppelstrang bei 95 °C denaturiert (1. Denaturierung). Die beiden DNA-Einzelstränge dienen im nächsten Schritt als Matrize für jeweils einen spezifischen komplementären Vorwärts (*forward*, F)- bzw. Rückwärts (*reverse*, R)-Primer. Diese lagern sich nun durch eine Temperaturniedrigung knapp unterhalb der Schmelztemperatur der Primer (~58 °C) an (2. Anlagerung). Im letzten Schritt wird die Temperatur auf 72 °C eingestellt. Dabei entfaltet die thermostabile Taq-Polymerase ihre volle Aktivität und bindet die in der Reaktion beigefügten freien Desoxynukleotidtriphosphate (dNTPs) an die freien 3'OH-Enden der Primer (3. Elongation).

Tab. 5 zeigt den verwendeten Reaktionsansatz. Das gewählte PCR-Programm (Tab. 6) basiert auf zwei Stufen mit unterschiedlichen Anlagerungstemperaturen. Nach 15 Zyklen sorgt eine um 2-3 °C erniedrigte Primer-Anlagerungstemperatur dafür, dass vermehrt spezifische Anlagerungsprodukte gebildet werden. Die initiale Denaturierungszeit von 15 Minuten wird benötigt, um die zu Beginn eingesetzte genomische DNA vollständig zu denaturieren.

Tab. 5 PCR Reaktionsansatz

	Volumen
DNA [20ng/μl]	5 μl
Taq PCR MasterMix*	12,5 μl
F-Primer [5pmol/ μl]	2 μl
R-Primer [5pmol/ μl]	2 μl
H ₂ O RNase-frei (bidest)	3,5 μl

*1000 Einheiten Taq DNA Polymerase

Tab. 6 PCR-Programm

	Temperatur in °C	Zeit in Minuten	Wiederholung
Initiale Denaturierung	94	15:00	
Denaturierung	94	0:30	15-mal
Primer-Anlagerung	58	0:20	
Elongation	72	0:45	
Denaturierung	94	0:30	15-mal
Primer-Anlagerung	55	0:20	
Elongation	72	0:45	
Finale Elongation	72	7:00	
Endkühlung	8-15	∞	

3.7.2 Cycle Sequencing: Kettenabbruchmethode nach Sanger

Die enzymatische Kettenabbruch (Didesoxy)-Methode nach Sanger (Sanger et al., 1992) beruht auf der Anlagerung fluoreszenzmarkierter Didesoxynukleotidtriphosphate, welche die wachsende Kette zum Abbruch bringen. Die Reaktion beginnt mit der Denaturierung des DNA-Doppelstrangs bei 94 °C. Beim Abkühlen des Ansatzes auf 50 °C lagert sich der zugegebene Vorwärts- oder Rückwärtsprimer an. Bei 60 °C beginnt die Polymerase, die in dem Reaktionsmix (BigDye®, Beckman-Coulter) im Überschuss enthaltenen Desoxyribonukleotidtriphosphate und Didesoxyribonukleotidtriphosphate (dNTPs und ddNTPs) einzubauen. Die dNTPs und ddNTPs, denen am 3' Ende eine Hydroxylgruppe (-OH) zur Anlagerung weiterer Nukleotide fehlt, konkurrieren dabei mit den dNTPs. Jedes Mal, wenn ein ddNTP eingebaut wird, bricht die Kette ab. Dies geschieht zufällig und kann jede Stelle des Basenstranges betreffen, weshalb eine Mischung von Fragmenten jeder Länge entsteht. Basenspezifische Fluoreszenzmarker, die an die ddNTPs gekoppelt sind, ermöglichen eine anschließende Detektion.

3.7.3 Laserfluoreszenzdetektion

Bei der Laserfluoreszenzdetektion werden die DNA-Fragmente kapillarelektrophoretisch aufgetrennt und anschließend detektiert. Dabei wird das Fluoreszenzsignal der ddNTP's mittels Laser in ein optisches Signal umgesetzt. Nachfolgend kann abgelesen werden, welche Base (A, T, G oder C) sich an dieser letzten Position der Kette befindet. Die Sequenz beginnt mit dem kürzesten Fragment und endet mit dem längsten. Tab. 7 gibt den Reaktionsansatz, Tab. 8 das verwendete Programm für das *Cycle Sequencing* wieder.

Tab. 7 *Cycle Sequencing* Reaktionsansatz

	Volumen
PCR Produkt ad [60ng/ µl]	x µl
BigDye®	1,6 µl
TE-Puffer	3,4 µl
F- oder R-Primer [5ng/ µl]	1 µl
Wasser	ad 20 µl

Tab. 8 *Cycle Sequencing* Programm

	Temperatur in °C	Zeit in Minuten	Wiederholung
Initiale Denaturierung	94	5:00	
Denaturierung	94	0:15	25-mal
Primer-Anlagerung	50	0:10	
Elongation	60	4:00	
Finale Elongation	60	10:00	
Endkühlung	8-15	∞	

3.7.4 Aufreinigungsschritte

1. Agencourt® *Solid-Phase Reversible Immobilization*

Die *Solid-Phase Reversible Immobilization (SPRI®) beads*-Technologie eignet sich zur Aufreinigung niedrig konzentrierter DNA (DeAngelis et al., 1995). Sie beruht auf paramagnetischen SPRI®-beads (Partikel), welche durch ihre spezielle Beschichtung eine reversible Bindung mit der DNA eingehen können. Wird von außen ein Magnetfeld aufgebaut, binden sich die *beads* mit der gebundenen DNA an dieses Feld und die verbleibenden Kontaminanten (Primer, dNTPs, nicht verbrauchte Polymerase) können herausgewaschen

werden. Danach wird die DNA von den *beads* gelöst und in ein separates Gefäß überführt (<http://www.beckmancoulter.de>).

2. Sephadex® Gelfiltration (Größen-Ausschluss-Chromatographie)

Sephadex® ist der Handelsname für ein quervernetztes Dextran-Gel (GE-Healthcare Bio-Sciences AB, Pharmacia), welches für die Gelfiltration verwendet wird. Dextrane sind hochmolekulare, verzweigte Glucose-Einheiten. Für diese Aufreinigung, die nach dem *Cycle Sequencing* erfolgt, wird das Sephadex®-Pulver auf eine Microfilterplatte verteilt und durch die Zugabe von je 0,3 mL Wasser pro *well* (Platten-Vertiefung) zum Quellen gebracht. Nach dem Abzentrifugieren von überschüssigem Wasser erhält man eine Gelsäule aus einer porösen Matrix, welche alle niedermolekularen Kontaminanten (Salze, Pufferreste) auffängt. Die hochmolekulare aufgereinigte DNA wird durch erneute Zentrifugation (5 Min. bei 2500 rpm) aufgefangen (<http://www.gelifesciences.com>).

3.8 MLPA

Die MLPA (*Multiplex Ligation-dependent Probe Amplification*) ist eine etablierte semiquantitative Methode zum Nachweis abweichender Kopienzahlen (Kap. 2.2.2 „CNV Detektion“) (Schouten et al., 2002). Mit dieser Methode können hetero- und homozygote Deletionen und Duplikationen im Größenbereich von einem Exon sowie im Promoterbereich eines Gens erfasst werden. Voraussetzung dafür ist, dass die Region, in der die Kopienzahlveränderung nachgewiesen werden soll, vorher bekannt ist.

Das Prinzip beruht auf spezifischen MLPA-Sonden. Diese beinhalten eine kurze Sequenz, die zur Zielregion komplementär ist. Als Erkennungssequenz tragen die Sonden zusätzliche Oligonukleotide mit variabler, sondenspezifischer Länge, welche nicht komplementär zur Zielregion sind und deshalb nicht gebunden werden. Eine Untersuchung von *MSH2* und *EPCAM* auf Deletionen und Duplikationen mittels MLPA wurde vor Beginn der wissenschaftlichen Arbeit im Rahmen der genetischen Diagnostik durchgeführt.

Die Proben-DNA (100 ng) wird thermisch denaturiert und die MLPA-Sonden (MRC-Holland, Niederlande), jeweils zwei aneinandergrenzende Sonden pro Zielregion, lagern sich an die Zielregion an. Durch eine Ligierungsreaktion werden die hybridisierten Fragmente miteinander verbunden. Nur an die Zielsequenz gebundene und damit ligierte Sonden werden in der

anschließenden PCR-Reaktion (Kap. 3.7.1) amplifiziert. Pro Reaktion können bis zu 60 verschiedene Sondenpaare parallel eingesetzt werden. Durch die unterschiedlich langen Erkennungssequenzen der MLPA-Sonden werden die Amplifikate im Sequenzer (Genetic Analyzer 3130XL, Life Technologies) elektrophoretisch aufgetrennt und können dadurch unterschieden werden.

Die Peakhöhe im Vergleich zu einer Referenz entscheidet über das Vorliegen einer Kopienzahlveränderung. Die Peakhöhe ist dabei proportional zur relativen DNA-Menge. Bei einer Deletion liegt ein kleinerer Peak, bei einer Duplikation ein größerer Peak im Vergleich zu den Kontrollen vor. Die Differenz zwischen Kontrollpeak und Probenpeak beträgt dabei > 25%. Die Ergebnisse werden mit dem Coffalyser (MRC-Holland) ausgewertet. Tab. 9 gibt das Programm der MLPA wieder.

Tab. 9 MLPA-Programm

Reaktion	Temperatur in °C	Zeit in Min.	Wiederholung
Denaturieren der Proben DNA	98	8:00	
	25°C	Pause	
Hybridisierungsreaktion	95	2:00	
	60	16 Stunden	
Ligasereaktion	54	Pause	
	54	15:00	
	98	5:00	
	4	Pause	
PCR-Reaktion	95	0:40	35-mal
	60	0:30	
	72	0:90	
	72	20:00	
	15	∞	

3.9 Identifizierung und Validierung von Kopienzahlveränderungen

3.9.1 Genomweite SNP-Array-Analyse

Für die CNV-Analyse wurde das SNP-Array „HumanOmniExpress-12“ der Firma Illumina verwendet, das über 700.000 über das Genom verteilte SNP-Marker enthält. Die genomische Test-DNA wird an vorgefertigte artifizielle Oligonukleotide gebunden, welche über kleine *beads* (Silicium Perlen) auf dem *Array* (der Matrix) fixiert sind. Die verwendete Technologie des

Material und Methoden

Illumina Infinium HD Assays basiert auf der Einzelbasen-Verlängerungs (*single base extension*, SBE)-Reaktion. Auf jedem *bead* sind viele hunderttausende Kopien von zielgerichteten 50-mer Oligonukleotidsonden gebunden. Jedes Oligonukleotid gibt einen spezifischen Locus im Genom wieder (Steemers et al., 2006). Die Eigenschaften des HumanOmniExpress-12 SNP-Arrays sind in Tab. 10 aufgeführt.

Tab. 10 Eigenschaften des HumanOmniExpress-12 SNP-Arrays

BeadChip	HumanOmniExpress-12
Anzahl SNP-Marker	733.202
Benötigte DNA Menge	200 ng
Assay	Infinium HD
Durchschnittliche Call Rate	99,84% (<99%)
Reproduzierbarkeit	99,99% (>99.9%)
Standard Log R	0,15 (<0,3)
Abstand (kb)	4,0 (mittlerer)/ 2,1 (Median)
Anzahl SNPs in 10kb RefSeq Genen	382.197
Nicht-synonyme SNPs	15.062

Als Ausgangskonzentration werden 200 ng genomische DNA benötigt. Diese wird zunächst denaturiert, neutralisiert und anschließend über Nacht isothermisch amplifiziert. Diese *whole-genome*-Amplifizierung erhöht die Menge der DNA um den Faktor 1000-3000. Bei 200 ng Ausgangskonzentration erhält man dementsprechend 0,2-0,6 mg DNA. Als nächstes wird die amplifizierte DNA in Fragmenten von 200-400 bp enzymatisch fraktioniert, mit Isopropanol präzipitiert und auf niedermolekulare Zielkonzentration resuspendiert. Diese so vorbereiteten genomischen Fragmente können nun komplementär mit den Oligonukleotidsonden auf dem BeadArray hybridisieren (<http://technology.illumina.com>). Die Oligonukleotidsonden enden jeweils exakt ein Basenpaar vor dem zu analysierenden SNP. Durch die nun stattfindende SBE-Reaktion baut die Polymerase ein einzelnes der vier hapten-gelabelten Didesoxynukleotide ein, welches komplementär zur Ziel-DNA ist (Abb. 9). Die folgende immunohistochemische Fluoreszenzmarkierung folgt einem 2-Farben Code, wobei die Nukleotide Adenin und Thymin rot, Guanin und Cytosin grün erscheinen (Steemers et al., 2006). Die Verwendung von zwei Farben schränkt die mögliche Auswahl an SNPs ein, da nur vier von sechs SNP-Gruppen verwendet werden können (A/G, T/C, A/C, und T/G) (Gunderson et al., 2006). Die verbleibenden Genotypen A/T und G/C können somit nicht mit erfasst werden. Sie machen

jedoch nur einen geringen Teil der genomweiten SNPs aus und können daher vernachlässigt werden (Steemers und Gunderson, 2007).

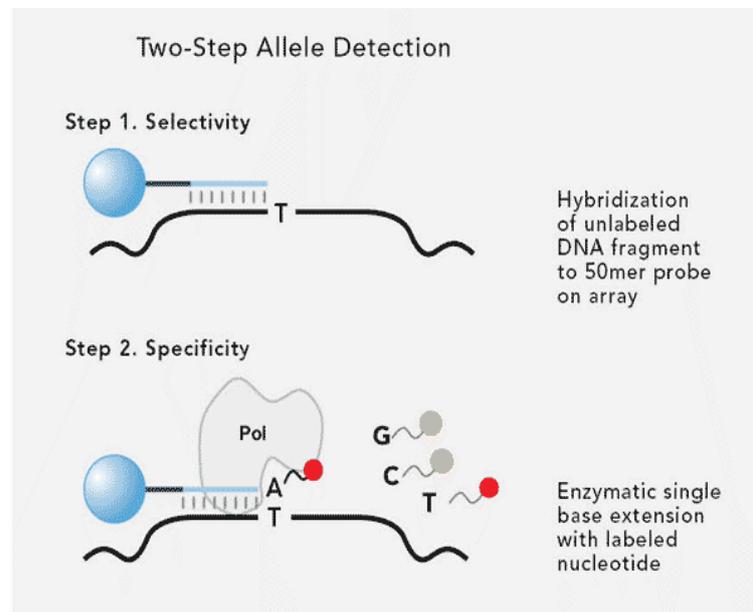


Abb. 9 Einzelbasenpaar-Verlängerungsreaktion (*single base extension*)

Schritt 1): Bindung der Ziel-DNA an die komplementäre, fixierte Oligonukleotidsonde, welche ein Basenpaar vor dem zu analysierenden SNP liegt.

Schritt 2): Abhängig von der Ziel-DNA baut die Polymerase ein spezifisches markiertes Nucleotid ein, welches anhand des 2-Farben Codes (siehe oben) detektiert wird (<http://technology.illumina.com/>).

Ein hochauflösender Laserscanner (i-scan, Illumina) detektiert das von den eingebauten Nucleotiden ausgehende Fluoreszenzsignal. Dieses wird in Abhängigkeit zur Gesamtmenge an gebundener DNA in Form von Signalintensität erfasst und vom Laserscanner in ein optisches Bildsignal umgewandelt. Für diploide Organismen werden bei biallelen Genorten drei Cluster (AA, BB und AB) (Abb. 10) erwartet.

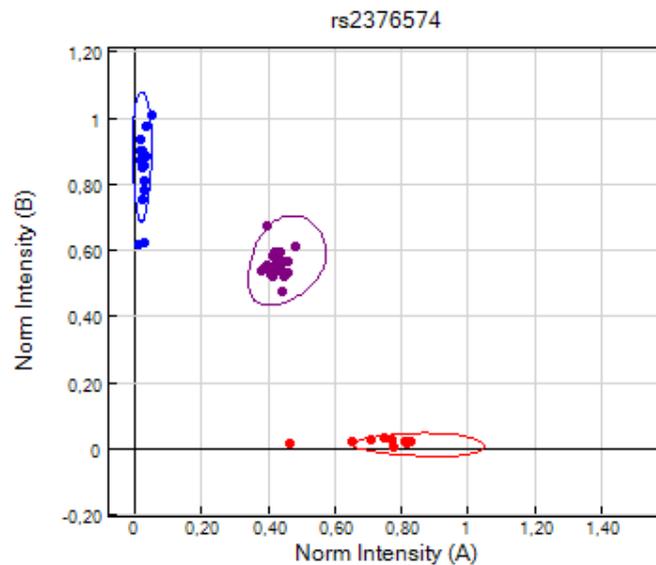


Abb. 10 Typischer diploider Genotypisierungsplot eines spezifischen Locus

Die drei verschiedenen Cluster repräsentieren die AA, BB, bzw. AB Genotypen (GenomeStudio, Illumina).

3.9.2 In-silico Analyse der Array-Daten

GenomeStudio

Aus den Rohdaten des SNP-Arrays berechnet die Software *GenomeStudio* (Illumina) verschiedene Werte, die für die Auswertung der Daten benötigt werden. Die *B Allele Frequency* (BAF) und die *Log R Ratio* (LRR) sind die zwei relevanten Werte für die CNV-Analyse.

Die BAF beschreibt den Genotyp eines SNP-Markers einer Person. Abhängig von dem Nukleotid, das während der *single base extension* eingebaut wurde, wird ein unterschiedliches Fluoreszenzsignal emittiert. Bei einer normalen Kopienzahl (CNV=2) kann ein homozygoter Genotyp (AA/BB) vorliegen. Es wird eine hauptsächlich rote bzw. grüne Fluoreszenz detektiert. Bei einem heterozygoten Genotyp (AB) ergibt sich eine Mischfarbe. Die normalisierten Rohdaten werden zusammen mit dem Referenzdatensatz der HNR-Proben auf ein Koordinatensystem aufgetragen. Proben mit gleichem Genotyp (AA, AB oder BB) werden zu Clustern zusammengefasst. Die BAF bezeichnet die allelische Zusammensetzung eines SNPs innerhalb des Genotyp-Clusters. Sie liegt zwischen 0 und 1. Dabei hat der Genotyp AA eine BAF von 0, der Genotyp AB eine BAF von 0,5 und BB eine BAF von 1.

Bei einer heterozygoten Deletion (CNV=1; Abb. 11) erwartet man durch den Verlust des zweiten Allels immer einen homozygoten Genotypen. Die BAF liegt daher bei 0 oder 1. Bei einer Duplikation (CNV=3; Abb. 12) ergeben sich mehrere mögliche Kombinationen. Die beiden homozygoten Genotypen (AAA oder BBB) führen weiterhin zu einer BAF von 0 oder 1. Bei den heterozygoten Genotypen (ABA/AAB bzw. BAB/BBA) liegt die BAF bei 0,33 bzw. 0,66.

Mit der BAF alleine kann noch keine Aussage über das Vorliegen eines CNVs getroffen werden. Dies ist erst in Kombination mit der Log R Ratio (LRR) möglich. Die LRR gibt Auskunft über die Intensität des Fluoreszenz-Signals eines Markers. Sie ergibt sich aus dem Verhältnis der gemessenen Intensität des Markers bei dem jeweiligen Patienten im Vergleich zu einer Kontrollgruppe. Entspricht die Intensität des Probenmarkers der zu erwartenden Intensität, liegt eine LRR von Null vor. Bei Patient und Kontrollgruppe liegt also die gleiche Kopienzahl vor. Ist die Intensität der Marker beim Patienten im Vergleich zur Kontrollgruppe erniedrigt, so liegt ein Wert kleiner Null vor (Deletion Abb. 11), ist sie wie bei einer Duplikation erhöht, so liegt eine LRR größer Null vor (Abb. 12).

Die Standardabweichung der LRR (Intensität Patient zu Referenz) ist ein Qualitätsmerkmal des Arrays. Proben mit einer hohen LRR weisen aufgrund des hohen Signalrauschens (*signal noise*) mehr falsch-positive CNVs auf, daher wurden Proben mit einer Standardabweichung LRR > 0,3 ausgeschlossen. Ein weiteres Qualitätsmerkmal ist die *Call Rate*. Sie gibt den prozentualen Anteil an Markern wieder, welcher während des Scanvorgangs erkannt wurde. Die *Call Rate* sollte nicht kleiner als 98% sein.

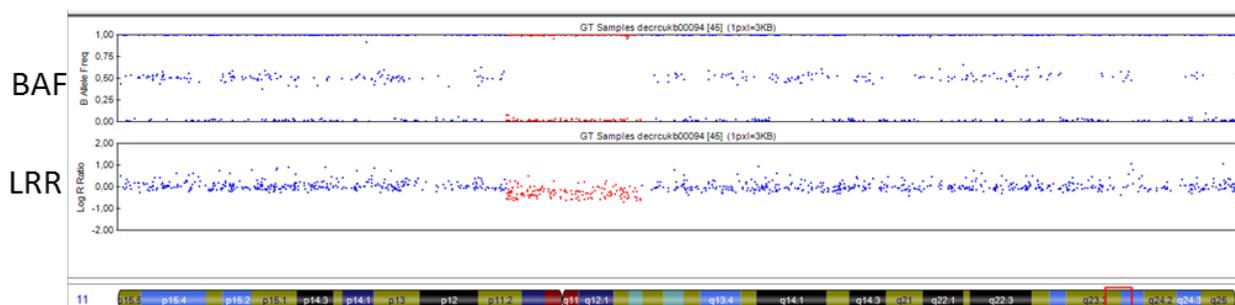


Abb. 11 Darstellung einer heterozygoten Deletion auf Chromosom 11 (GenomeStudio)

Jeder Punkt repräsentiert einen SNP. BAF: *B Allele Frequency*, LRR: Log R Ratio. Die roten Marker beschreiben den Bereich der Kopienzahlveränderung. Die BAF liegt für alle untersuchten SNPs im deletierten Bereich bei 0 oder 1. Die LRR ist sichtbar herabgesetzt (< 0).

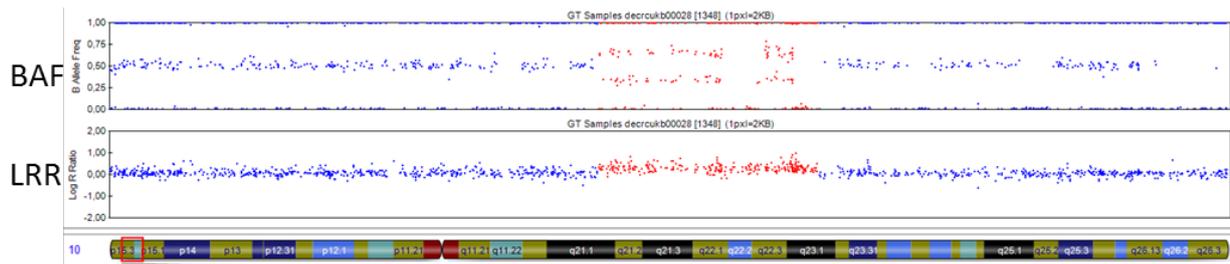


Abb. 12 Darstellung einer heterozygoten Duplikation auf Chromosom 10 (GenomeStudio)

Jeder Punkt repräsentiert einen SNP. BAF: *B Allele Frequency*, LRR: *Log R Ratio*. Die roten Marker beschreiben den Bereich der Kopienzahlveränderung. Die BAF zeigt die vier verschiedenen Genotypen: BBB, BBA, BAA, und AAA, an. Die LRR ist sichtbar erhöht (> 0).

QuantisNP

Zur Identifizierung der potentiellen CNVs wurde das Programm *QuantisNP* (v2) verwendet, welches in einer vergleichenden Studie mit als bestes Detektierprogramm abgeschnitten hat (Dellinger et al., 2010; Winchester et al., 2009). *QuantisNP* wurde von der Universität Oxford und dem *Wellcome Trust Centre for Human Genetics* entwickelt (<http://www.well.ox.ac.uk/QuantisNP>). Dem Programm liegt ein Algorithmus zugrunde, der als *Hidden Markov Model* bezeichnet wird (Colella et al., 2007). Hierbei werden die Wahrscheinlichkeiten, mit der ein Zustand ein bestimmtes Signal emittiert, abgebildet. Der eigentliche Zustand ist verborgen (*hidden state*). Dieser Algorithmus greift auf die von *GenomeStudio* wiedergegebene *B Allele Frequency*, die *Log R Ratio* und die chromosomale Position der entsprechenden SNPs zurück. *QuantisNP* errechnet für jeden detektierten CNV den *Log Bayes-Faktor* (LBF), einen Wahrscheinlichkeitswert für die Echtheit des CNVs. Je höher der Wert, desto wahrscheinlicher ist die detektierte Kopienzahlveränderung echt. Die Entwickler von *QuantisNP* empfehlen einen Schwellenwert von $LBF=30$ einzusetzen, da ab diesem Wert die Falsch-Positiv-Rate auf unter 1% sinkt (<http://sites.google.com/site/quantisnp/howto>). Das bedeutet, dass für alle CNVs mit einem $LBF \geq 30$ die detektierte Kopienzahlveränderung mit sehr hoher Wahrscheinlichkeit auch wirklich vorliegt. Da dadurch jedoch das Auflösungsvermögen verringert wird, wurde in dieser Arbeit der LBF auf bis zu ≥ 10 herabgesetzt. Alle CNVs wurden daraufhin validiert, um mögliche falsch-positiv detektierte CNVs zu entfernen.

3.9.3 Filterkriterien der CNV-Analyse

Aufgabe dieser Arbeit war es, seltene, hochpenetrante CNVs, welche als krankheitsverursachend für den Phänotyp einzustufen sind, zu ermitteln. Um diese aus der

Gesamtzahl der detektierten CNVs ausfindig zu machen, wurden stringente Filterkriterien angewandt. Die ersten vier Filterschritte sind technischer Natur. Die Filterschritte 5-8 dienen einer ersten Beurteilung der CNVs bezüglich ihrer ätiologischen Relevanz.

1. Qualitätskontrolle

Die Standardabweichung der *Log R Ratio* beschreibt die Abweichung der Intensität des Probensignals bei einem Patienten im Vergleich zur Referenz. Für sie wurde ein Schwellenwert von $\leq 0,3$ festgelegt. Alle Patienten mit einer Standardabweichung *Log R Ratio* größer 0,3 wurden aus der Analyse ausgeschlossen. Die *Call Rate* bezieht sich auf den Anteil aller detektierten Marker. Es wurden nur Patienten eingeschlossen, deren *Call Rate* größer oder gleich 98% betrug.

2. Länge der CNVs ≥ 10 kb

Es wurden nur CNVs mit einer Größe ab 10 kb in die Analyse eingeschlossen. Dies hat vor allem technische Gründe. Kleinere CNVs sind mit dem verwendeten Array aufgrund der vorhandenen Auflösung (mittlerer Abstand der SNP-Marker 4 kb) schwierig zu detektieren. Zudem ist die Rate an falsch-negativen Befunden aufgrund der geringen SNP-Abdeckung in diesem Bereich sehr hoch.

3. Anzahl konsekutiver SNP-Marker ≥ 5

Um die Anzahl an falsch-positiven CNVs weiter einzuschränken, wurden nur CNVs berücksichtigt, bei denen die Kopienzahlveränderung anhand von mindestens fünf konsekutiven SNP-Markern detektiert wurde.

4. Log Bayes Faktor ≥ 10 / ≥ 20

Der *Log Bayes Faktor* (LBF) ist das von *QuantiSNP* errechnete Wahrscheinlichkeitsmaß für die Echtheit eines detektierten CNVs (Kap. 3.9.2). Für diese Arbeit wurde ein LBF von ≥ 10 für Deletionen und ≥ 20 für Duplikationen ein LBF gewählt. Dieser Wert wurde im Laufe dieser Arbeit empirisch ermittelt. Begonnen wurde mit einem LBF von ≥ 30 . Anschließend wurde der Wert in Zehnerschritten herabgesetzt, um weitere, möglicherweise relevante CNVs mit einzubeziehen. Bei den Duplikationen mit einem LBF ≤ 20 zeigte sich jedoch, dass diese häufig nicht mehr verifiziert werden konnten, daher wurde der LBF 20 für Duplikationen auf \geq gesetzt, während bei Deletionen ein LBF von ≥ 10 möglich war. Alle CNVs unterhalb dieser Schwelle wurden nicht analysiert.

5. Ausschluss von häufigen putativen CNVs

Um Array-spezifische Artefakte als auch in der Population häufige CNVs auszuschließen, wurden die Patienten-CNVs mit den CNVs von 1320 *In-House* (HNR)-Kontrollen verglichen. Die Kontroll-DNA wurden hierzu auf dem gleichen SNP-Array (HumanOmniExpress-12, Illumina) wie die Patienten-DNA genotypisiert. Die CNVs der Patienten und Kontrollen wurden nach den oben beschriebenen Schritten 1 - 4 gefiltert. Deletionen und Duplikationen wurden getrennt voneinander verglichen. Es wurden alle CNVs, die mindestens drei Mal ($\geq 3x$) in den Kontrollen auftraten (Frequenz $\geq 0,2\%$), herausgefiltert. Diese Frequenz wurde entsprechend des Auftretens von HNPCC und LS in der Allgemeinbevölkerung, etwa 1/500 (Steinke et al., 2013), gewählt.

6. Ausschluss von CNVs in nicht-proteinkodierenden Genen

CNVs in proteinkodierenden Genbereichen wurden über die *RefSeq*-Gen-Annotation (Pruitt et al., 2007) des *UCSC Genome Browsers* ermittelt. CNVs, die lediglich olfaktorische Gene und Pseudogene enthielten, wurden von der Analyse ausgeschlossen. Ebenso CNVs in nicht-kodierenden Bereichen, mit Ausnahme von intronischen Gen-Regionen. Als weitere Ausnahme wurde zusätzlich nach potentiell cis-regulatorischen Elementen von *MSH2* gesucht. Weit entfernte *long-range* cis-regulatorische Elemente können bis zu 1,5 Mb entfernt vom Zielgen liegen und seine Expression regulatorisch beeinflussen (Epstein, 2009). Für die Suche nach CNVs in cis-regulatorischen Bereichen von *MSH2* wurde daher der Chromosomenbereich 1,5 Mb vor und nach *MSH2* (chr2:47.580.206-47.760.367) mit eingeschlossen.

7. Ausschluss von CNVs in Bereichen segmentaler Duplikationen

Segmentale Duplikationen sind definitionsgemäß genomische Regionen > 1 kb mit 90-98%iger Sequenzidentität. Durch die hohe Sequenzidentität ist es schwierig, eine ausreichend hohe spezifische Markerabdeckung zu erreichen, dort detektierte CNVs sind meist Artefakte. Da in diesen Regionen auch Primer nicht spezifisch binden, ist eine Validierung aufwändig und teils unmöglich. Des Weiteren sind die meisten Gene, welche in Bereichen segmentaler Duplikationen lokalisiert sind, selten kopienzahlsensitiv und somit nicht krankheitsverursachend. Aus diesen Gründen wurden CNVs, welche größtenteils von Bereichen segmentaler Duplikationen betroffen waren, ausgeschlossen.

8. Filtern gegen die *Database of Genomic Variants*

Die *Database of Genomic Variants* (DGV) (dgv.tcag.ca/) (MacDonald et al., 2014) umfasst eine Datenbank humaner genomischer Kopienzahlvarianten, die bei gesunden Kontrollpersonen identifiziert wurden. Sie ermöglicht damit einen umfassenden Vergleich der in dieser Studie ermittelten CNVs mit denen der Allgemeinbevölkerung hinsichtlich Ausdehnung und Frequenz. Es wurden alle CNVs aus unserer Analyse ausgeschlossen, welche in der DGV mit einer Frequenz von größer 0,2% und/oder bei mindestens 3 Personen beschrieben wurden. Es wurden hierfür in der DGV nur Studien berücksichtigt, welche die Kopienzahlveränderungen mittels Oligo-Array-CGH oder SNP-Array detektiert hatten. Der in dem CNV enthaltene Genbereich musste vollständig von dem CNV der Vergleichsstudie abgedeckt sein, um ausgeschlossen zu werden. Deletionen wurden getrennt von Duplikationen verglichen.

9. Validierung der CNVs

Die Software *GenomeStudio* bietet die Möglichkeit, alle CNVs zu visualisieren und zu analysieren. Durch diese optische Beurteilung kann eine erste Einschätzung der Validität der berechneten CNVs vorgenommen werden. Abb. 11 und Abb. 12 zeigen eine beispielhafte Darstellung einer Deletion und Duplikation im *GenomeStudio*. Alle CNVs mit Ausnahme vereinzelter, sehr großer CNVs mit einem LBF > 60, wurden zusätzlich per qPCR validiert. Alle für die qPCR verwendeten Primer sind in Tab. A.1 und Tab. A.2 aufgelistet.

10. Priorisierung der CNVs

Die gefilterten CNVs wurden in zwei parallelen Schritten weiter priorisiert. Das erste Kriterium berücksichtigte die Art der Mutation (Deletion/Duplikation, komplett/partiell/intronisch) zusammen mit der Genfunktion. Für die Beschreibung der Genfunktion wurden die PubMed- und die GeneCards-Datenbank herangezogen. Hierbei wurden Gene, welche hinsichtlich der bekannten Funktion in einem möglichen Zusammenhang zur Tumorentwicklung stehen (Onkogen/Tumorsuppressorgen) bzw. in relevante onkologische Signalwege involviert sind, ausgewählt. Die beschriebene Genfunktion musste mit dem zu erwartenden Effekt (*gain of Function/Loss of Function*) übereinstimmen.

Das zweite Kriterium berücksichtigte, welche der 76 Gene sensitiv auf den Verlust eines Allels reagieren, bzw. intolerant gegenüber trunkierenden Mutationen sind. Dafür wurden zwei Scores, der Haploinsuffizienz- (HI) und der Intoleranz (*Residual Value of Intolerance Score*,

RVIS)-Score, herangezogen. Niedrige Scores stehen hierbei für haploinsuffiziente bzw. variantenintolerante Gene. Die Schwelle wurde auf $\leq 35\%$ für den HI-Score und auf $\leq 25\%$ für den RVIS-Score gesetzt. Dies entsprach den 10% aller detektierten Gene mit den niedrigsten Scores. Die Scores lassen sich nur auf Deletionen oder partielle Duplikationen mit einem vermuteten LoF-Effekt anwenden. Komplette Gen-Duplikationen, bei denen man von einem *gain of function* ausgeht, können nicht bewertet werden.

Weitere Bewertungskriterien waren zum einen das Vorkommen von somatischen Mutationen in diesen Gene bei Adenokarzinomen des TCGA (*The Cancer Genome Atlas*). Interessant waren solche Gene, die somatische Mutationen $\geq 3\%$ der 295 nicht-hypermutierten Adenokarzinome aufwiesen. Die hypermutierten Tumordaten wurden ausgeschlossen, um die Daten gegen *Passenger*-Mutationen zu korrigieren. Zum anderen wurde untersucht, ob die CNV-Gene laut Daten des UniGene Est-Profiles (NCBI) in Darmschleimhaut exprimiert werden.

3.9.4 Quantitative PCR

Die quantitative PCR (qPCR) ist eine etablierte Methode der Validierung von CNVs. Von der Methodik her unterscheidet sich die qPCR nur unwesentlich von einer herkömmlichen PCR. Ein wichtiges Unterscheidungsmerkmal ist der Einsatz von Fluoreszenzfarbstoffen zum Reaktionsgemisch, deren Fluoreszenzsignal proportional zur Konzentration des PCR-Produkts ist. Während der PCR-Reaktion wird der Anstieg des Fluoreszenzsignals konstant gemessen. Der Eintritt in die Exponentialphase in Relation zur Anzahl der gelaufenen PCR-Zyklen ist bedeutend für die Ermittlung der Menge des Produkts (Klein 2002, Kubista et al., 2006).

Als Farbstoff wurde Power SYBR[®]-Green I (Applied Biosystems), ein asymmetrisches Cyanin, eingesetzt. Es interkaliert unspezifisch in doppelsträngige DNA, in geringerem Maße aber auch in einzelsträngige DNA und RNA. Die Verwendung von Power SYBR[®]-Green I bietet, im Vergleich zu fluoreszierenden TaqMan-Sonden, den Vorteil einer einfachen und kostengünstigen Anwendung. Ein Nachteil ist jedoch, dass auch Nebenprodukte und RNA-Verunreinigungen mitdetektiert werden. Um dies zu vermeiden, muss bei der qPCR auf besondere Reaktionsbedingungen (s. u.) sowie auf möglichst spezifische Primer geachtet werden. Die Schmelzkurvenanalyse gibt Aufschlüsse über vorhandene mitamplifizierte Nebenprodukte (Klein, 2002). Werden mehrere Peaks detektiert (Abb. 13A), muss die qPCR unter veränderten Bedingungen wiederholt werden.

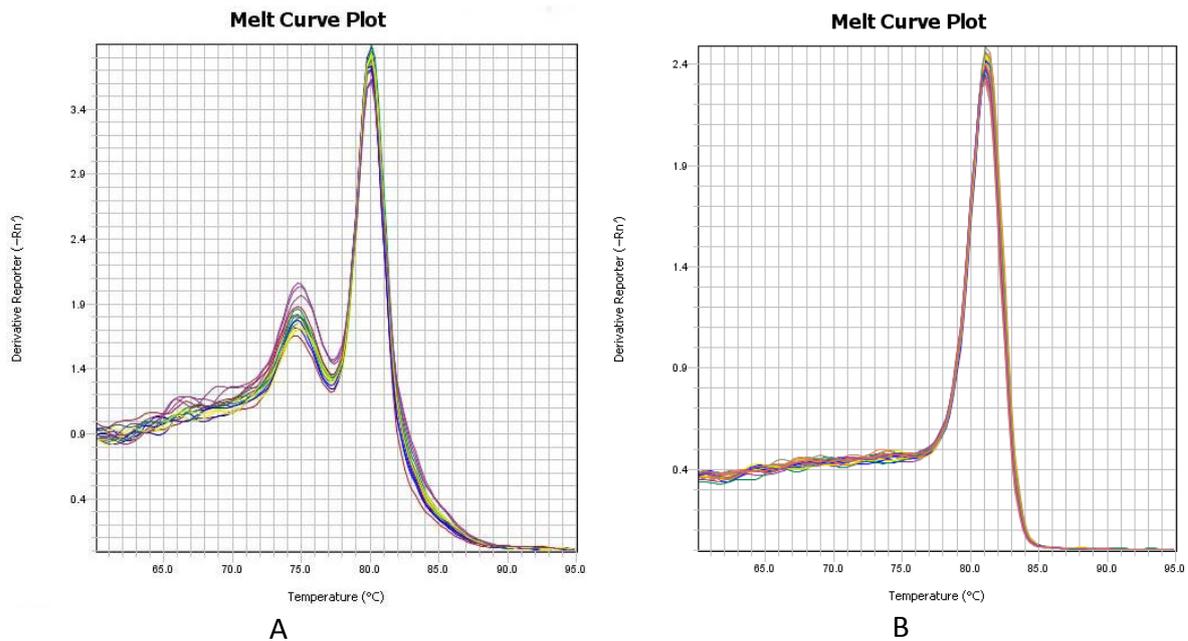


Abb. 13 Beispiel für Schmelzpunktkurven der PCR-Produkte.

A zeigt die Schmelzpunktkurve eines unspezifischen PCR-Produktes neben dem Hauptprodukt.

B zeigt den idealen Kurvenverlauf mit nur einem signifikanten Peak.

Reaktion

Jeder Reaktionsansatz enthält DNA-Triplikate von vier gesunden Kontrollen und die Patienten-DNA, jeweils in einer Konzentration von 10 ng/ μ l. Für jede Zielregion wurden, je nach Größe des Bereichs, zwei bis drei Primerpaare verwendet. Neben der Zielregion werden auch drei interne Kontrollbereiche in so genannten Housekeeping genes (HKG) mit amplifiziert. Diese Gene werden ubiquitär exprimiert und liegen üblicherweise mit normaler Kopienzahl (CNV = 2) vor. Hier wurden die drei Gene *BNC1* (basonuclin1), *CFTR* (cystic fibrosis transmembrane conductance regulator) und *RPP38* (ribonuclease P/MRP 38kDA subunit) verwendet. Das qPCR-Programm und der Reaktionsansatz sind in Tab. 11 und Tab. 12 dargestellt. Die PCR-Läufe wurden auf dem VIIATM 7 Real-Time PCR System, Applied Biosystems durchgeführt.

Tab. 11 qPCR Programm

	Temperatur in °C	Zeit in Minuten	Wiederholung
	50	2:00	
Initiale Denaturierung	95	10:00	
Denaturierung	95	0:15	
Primer-Anlagerung mit Sequenz-Elongation	60	1:00	40 Mal
Endkühlung	8-15	∞	

Tab. 12 qPCR Reaktionsansatz

	Volumen
DNA [10ng/μl]	2 μl
Power SYBR® Green I	5 μl
F-Primer [50pmol/ μl]	0,4 μl
R-Primer [50pmol/ μl]	0,4 μl
destillierte H ₂ O	2,2 μl
Gesamtvolumen	10 μl

Berechnung: $2^{-\Delta\Delta t}$ -Methode

Nach der PCR werden anhand des gemessenen Fluoreszenzsignals die CT-Werte (Cycle Threshold) für jede Position auf der Platte wiedergegeben. Der CT-Wert beschreibt die Anzahl an Zyklen, bei welcher das Signal einen festgesetzten Schwellenwert überschreitet und die Reaktion in die Exponentialphase übergeht (Abb. 14) (Kubista et al., 2006).

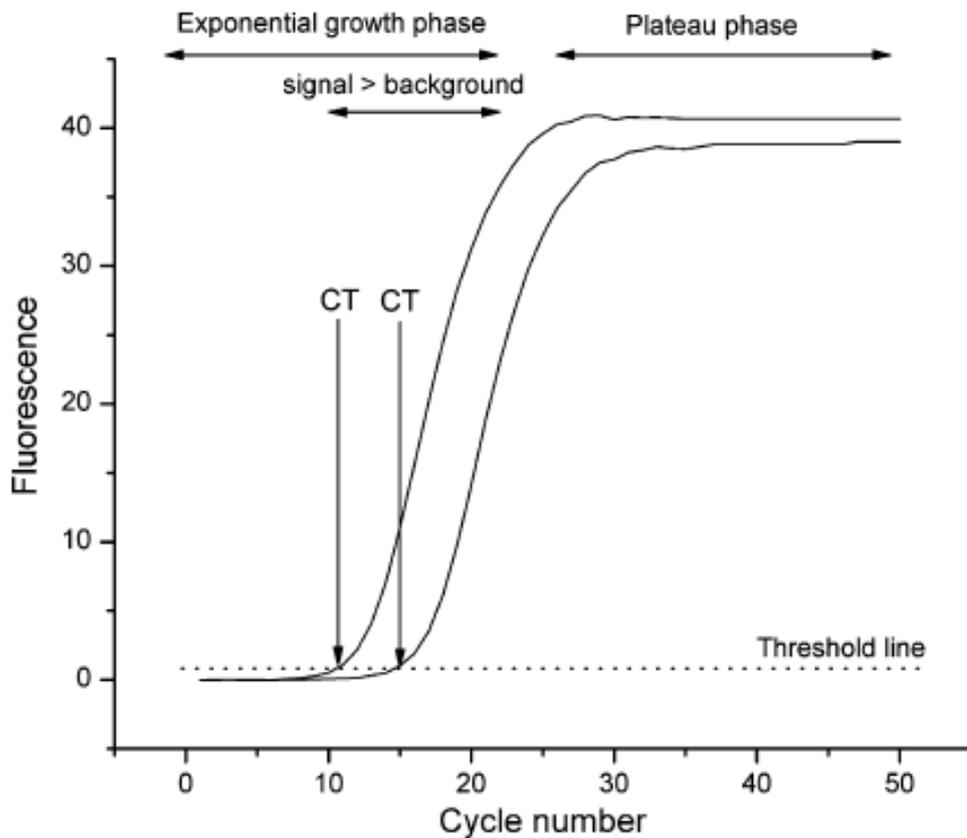


Abb. 14 qPCR Reaktionskurve

Dieser Wert gilt als Maß für die Konzentration der Ziel-DNA in der PCR-Reaktion. Der CT-Wert reagiert sehr sensibel auf äußere Einflüsse, wie eine veränderte Zielkonzentration (erwünscht), als auch Unterschiede im Reaktionsansatz, dem verwendeten Instrument und dem pH-Wert der Lösung (unerwünscht). Daher wird jeder Messwert gegen alle drei internen Kontrollen (HKG) normalisiert. Durch Subtraktion des CT-Werts der internen Kontrolle (HKG) von der Zielsequenz (Z) erhält man den ΔCT -Wert ($\Delta\text{CT} = \text{CT}_Z - \text{CT}_{\text{HKG}}$). Aus allen drei ΔCT -Werten wird der Mittelwert gebildet.

Im nächsten Schritt wird der $\Delta\Delta\text{CT}$ -Wert berechnet, indem man von dem Mittelwert der ΔCT -Differenzen des untersuchten Patienten (P) den ΔCT -Mittelwert einer gesunden Referenzperson (R) subtrahiert. Jeder Wert wird in Bezug zu allen vier Patienten gesetzt ($\Delta\Delta\text{CT} = \Delta\text{CT}_P - \Delta\text{CT}_R$).

$2^{-\Delta\Delta\text{CT}}$ gibt die Anzahl an DNA-Zielabschnitten in der qPCR nach Normalisierung mit einer endogenen Kontrolle (HKG) und in Relation zu jeder der vier Referenzpersonen (R) wieder. Die

Multiplikation $2 \times 2^{-\Delta\Delta CT}$ zeigt die relative Kopienzahl des gesuchten DNA-Abschnitts im Genom des Patienten an (Livak und Schmittgen, 2001).

3.9.5 Segregationsanalyse

Dort wo DNA von betroffenen Familienangehörigen verfügbar war, wurde überprüft, ob der ermittelte CNV co-segregiert. Hierfür wurde eine qPCR bei dem betroffenen Familienangehörigen durchgeführt. Der Index-Patient diente hierbei als Positiv-Kontrolle. Es wurden dieselben Primer wie zur Validierung des CNVs verwendet.

3.9.6 Pathway-Analyse

Eine Pathway-Analyse wurde mit einer webbasierten Software (Ingenuity Pathway Analyse) durchgeführt. Hierfür wurden alle CNV-Gene gemeinsam mit 38 relevanten CRC-Kandidatengenen (Tab. 1.1.1.1F) hochgeladen, um nach einer zufälligen Häufung der identifizierten Kandidatengene in bestimmten Signalwegen zu suchen oder Gene zu identifizieren, die mit bekannten onkologischen Signalwegen assoziiert sind und damit für die Pathogenese der Erkrankung eine mögliche Rolle spielen. Alle Pathways mit einem p-Wert $\leq 0,05$ wurden in Betracht gezogen.

3.9.7 Netzwerk-Analyse

Für die Suche nach funktioneller Assoziation zwischen MSH2 und den Proteinen der CNV-Analyse wurde eine „STRING“-Abfrage durchgeführt. Diese Datenbank verwendet bekannte und prognostizierte Assoziationen zwischen Proteinen und basiert auf *genomic context* Analysen, Hochdurchsatz-Untersuchungen, Co-Expressionsdaten und Algorithmus-basierten Analyseverfahren (Franceschini et al., 2013).

3.10 Next-Generation-Sequencing

Das *Next-Generation-Sequencing* (NGS) bietet die Möglichkeit, große DNA-Bereiche in kurzer Zeit zu sequenzieren und Punktmutationen sowie kleine Deletionen und Duplikationen (*Indels*) aufzuspüren. Für den vorliegenden Zweck wurde über die webbasierte Illumina-Software *DesignStudio* ein *Customized targeted NGS Panel* mit 104 ausgewählten Genen erstellt. Dieses individuelle *Custom-Panel* umfasst die 66 interessantesten Kandidatengene aus der CNV-Analyse sowie weitere 38 Gene aus der Literatur, die bereits im Zusammenhang mit einem erhöhten Darmkrebsrisiko beschrieben wurden.

Um die Sequenzierreaktion zu starten, bedarf es zunächst manueller Vorbereitung der DNA. Hierbei wurde das Nextera Rapid Capture Enrichment-Protokoll angewandt.

3.10.1 Nextera Rapid Capture Enrichment

Das dreitägige Nextera Rapid Capture Enrichment-Protokoll umfasst 14 Arbeitsschritte. In einem Lauf können die DNA-Proben von bis zu zwölf Patienten gleichzeitig angereichert werden. Die eingesetzte genomische DNA (50 ng) wird gleichmäßig fragmentiert und mit universellen Adaptoren (*nextera transposomes*) ligiert. Nach dem ersten Amplifizierungsschritt werden die Fragmente mit unterschiedlichen Indizes versehen. Diese wurden den Personen zuvor zugeteilt und ermöglichen später eine automatische Zuordnung der Sequenzen nach dem Lauf. Nach diesem Schritt können die zwölf verschiedenen DNA-Proben in einem Probengefäß weiterverarbeitet werden (*pooling*). Weitere Amplifizierungs- und (beheizte) Waschschriffe folgen, um eine *sample-library* mit hoher Spezifität und Reinheit zu erhalten. Nach der letzten Konzentrations- (Qubit®, Life Technologies) und Fragmentlängenmessung (Bioanalyzer, Agilent Technologies) wird die Probe auf eine finale Endkonzentration von 6-8 pM verdünnt, mit NaOH dehybridisiert und auf das Sequenziergerät (MiSeq 2000, Illumina) geladen.

Reaktion (bridge amplification und SBS-Reaktion)

Die vorpräparierten Zielbereiche werden maschinell auf die *flow cell* (Durchflusszelle) transferiert. Diese Mikroglassplatte mit 8-Kanälen ist der Ort, wo die eigentliche Sequenzierreaktion stattfindet. Auf ihrer Oberfläche befinden sich mehrere Millionen Oligonukleotide. Diese Oligonukleotide sind komplementär zu den ligierten Adaptersequenzen der Fragmente und hybridisieren mit diesen. Die gebundenen Fragmente werden durch eine Brückenbildung auf der *flow cell*-Oberfläche amplifiziert (*bridge Amplification*). Es wird ein dichtes Cluster von einer Million Kopien des Originalfragments gebildet. Diese Million von Clustern werden basierend auf dem Prinzip des *sequencing by synthesis* auf der *flow cell*-Oberfläche parallel sequenziert (Abb. 15).

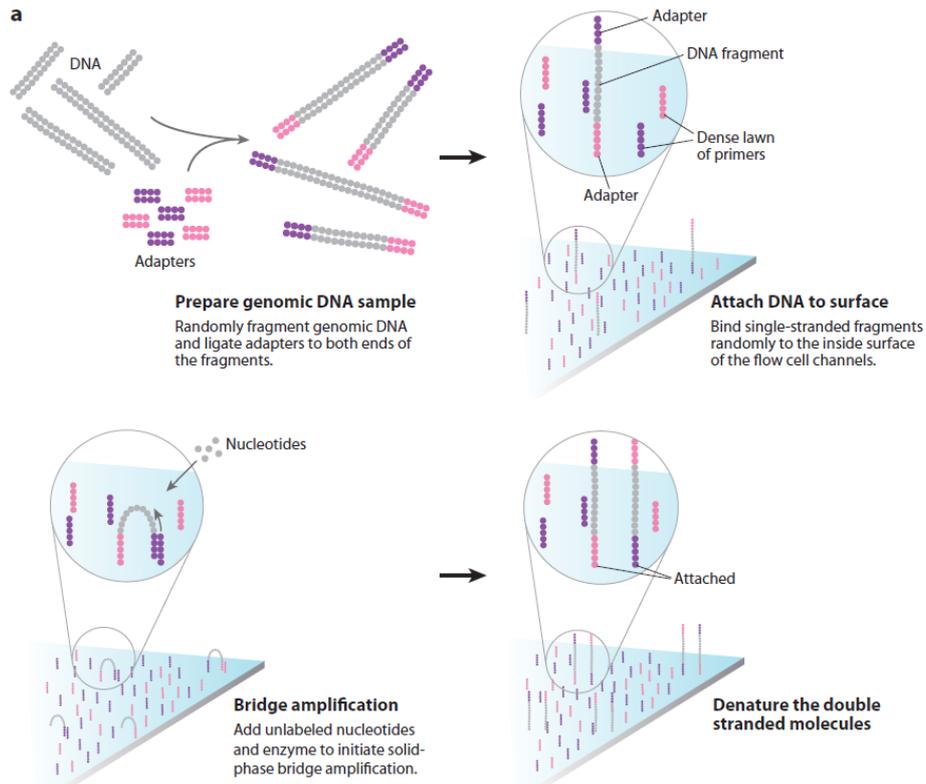


Abb. 15 Illumina *sequencing-by-synthesis*

Fragmentierte, genomische DNA wird mit spezifischen Adaptoren von beiden Seiten ligiert. Einzelsträngige Fragmente hybridisieren mit Oligos auf der *flow cell*-Oberfläche. Durch die Brücken-Amplifizierung mit unmarkierten Nucleotiden wird ein dichtes Netz von Clustern erhalten. Anschließend erfolgt die Sequenzierreaktion (Mardis, 2008).

Jeder Zyklus gliedert sich dabei wie folgt: Alle vier fluoreszenzmarkierten Nucleotide werden simultan auf die *flow cell* aufgebracht. Die DNA-Polymerase baut eines der vier Nucleotide ein. Durch die Konkurrenz der Nucleotide untereinander wird der Einbau einer falschen Base minimiert. Eine Schutzgruppe, welche die 3'OH-Stelle des Nucleotids blockiert, verhindert die Anlagerung weiterer Nucleotide. Es folgt die Signalintensitätsmessung des Fluorophors der eingebauten Base. Über die Fluoreszenz wird in einem nun nachfolgenden bildgebenden Schritt die eingebaute Nucleinsäure erfasst und anschließend werden der Fluorophor sowie die Schutzgruppe chemisch entfernt (reversible Terminator Chemie, RTC, Abb. 16).

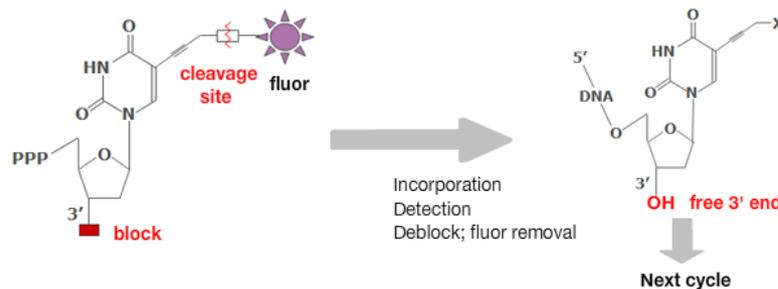


Abb. 16 Reversible Terminator-Chemie (RTC)

Bei jedem Zyklus wird ein fluoreszenzmarkiertes Desoxynucleotid (dNTP) eingebaut. Nach der Detektion erfolgt die Entfernung des Fluorophors und der Schutzgruppe (<http://www.illumina.com>).

Die entsprechende Base wird direkt über die Signalintensitätsmessungen des Fluorophors während jedes Zyklus erfasst (*base call*). Der *Base-Call-Algorithmus* ordnet die Sequenzen einander zu. Alle überschüssigen Desoxynucleotide und DNA-Polymerasen werden gewaschen und ein neuer Zyklus beginnt. Bei einer *paired-end read* (Ansequenzierung von beiden Seiten) von 150 bp wird dieser Zyklus insgesamt 300 Mal wiederholt. Eine integrierte Qualitätsprüfung wertet die Daten von jedem Lauf aus und entfernt automatisch Sequenzen, die der Referenzsequenz nicht zugeordnet werden konnten (Mardis, 2008).

3.10.2 Bioinformatische Datenanalyse

Nach der Sequenzierreaktion werden die kurzen DNA-Fragmente (*reads*) an einer Referenzsequenz ausgerichtet. Dieser Prozess des *Alignments* ist essentiell für das Ermitteln von Varianten in der Probensequenz (Ruffalo et al., 2011). Durch das Alignment werden *fastq-files* generiert, die dann in *BAM-files* konvertiert werden können. Anhand der *BAM-files* findet darauffolgend die eigentliche Identifizierung der genetischen Varianten (*variant calling*) statt. Hierfür wird nur die in der Manifest-Datei hinterlegte Zielregion berücksichtigt. Die Daten werden in generierten *VCF-files* gespeichert.

3.10.3 Varianten-Filterung

Die nach dem Lauf detektierten Varianten wurden mit dem *VariantStudio* (Illumina) weiter gefiltert. Die ersten drei Filterschritte zielen darauf ab, Artefakte aus der Analyse auszuschließen. In den anschließenden Schritten (4 und 5) sollen seltene und mutmaßlich pathogene Varianten herausgefiltert werden.

1. Qualität (Q > 30)

Die Qualität der Varianten, bezogen auf die Wahrscheinlichkeit, eine falsche Base ermittelt zu haben, wird in logarithmischen Q-Scores angegeben. Ein Q-Score von 30 bedeutet eine Wahrscheinlichkeit von 1 zu 1000, eine falsche Base zu ermitteln. Hier wurde ein Q-Score von > 30 gewählt. Somit liegt die Wahrscheinlichkeit, eine falsche Base ermittelt zu haben, bei < 0.1%.

2. Sequenziertiefe (*Read Depth*) (> 10)

Die *read depth* entspricht der Abdeckung einer bestimmten Position und beschreibt, wie häufig diese von einzelnen *reads* sequenziert wurde. Es wurden nur Bereiche mit einer *read depth* > 10 ausgewertet, da die Berechnung von Varianten bei einer niedrigen *read depth* zu einer hohen falsch-positiv Rate führt.

3. Anteil mutierter Reads (> 10%)

Der Anteil mutierter Reads beschreibt das Verhältnis mutierter zu nicht-mutierter Reads (Wildtyp). Da bei einer großen Zahl mutierter Reads von einem falsch-positiven Ergebnis auszugehen ist, wurde der Anteil mutierter Reads $\leq 10\%$ nicht bewertet.

4. Allelfrequenz (< 0,01% dominantes Modell, < 1% rezessives Modell)

Ausgehend von einem dominanten Modell wurden nur potentiell hochpenetrante, seltene Varianten, die mit einer Allelfrequenz < 0,01% in der Allgemeinbevölkerung auftreten, berücksichtigt. Für die Bestimmung der Allelfrequenz der Varianten wurden drei verschiedene Browser herangezogen (ExAC; 1000 Genomes; EVS), die auf weltweite Kollektive zurückgreifen. Für die Suche nach Varianten, die einem rezessiven Vererbungsmodell folgen könnten und daher als heterozygote Varianten häufiger in der Allgemeinbevölkerung vorkommen, wurde die Allelfrequenz auf < 1% erhöht.

5. Abschätzung der relativen Pathogenität der Varianten

Zur Abschätzung der relativen Pathogenität aller Arten von Punktmutationen oder kleinerer Indels wurde der „skalierte C-Score“ angewandt (Kircher et al., 2014). Er vereint verschiedene Annotationen aus Daten zur Konservierung, Genregulation, Transkription und Scores, die die Funktionalität der Variante auf Protein-Ebene abschätzen (SIFT und PolyPhen), zu einem einzigen Score. Er basiert auf dem Prinzip des „kombinierten annotationsabhängigen Verlusts“

(*Combined Annotation Dependent Depletion*, CADD). Für die Berechnung der CADD wurden 14,7 Millionen Allele aus Daten des *ENCODE-Projects* (Consortium, 2012) mit simulierten 8,6 Milliarden möglichen Varianten berücksichtigt. Die daraus abgeleiteten C-Scores vereinen Informationen zur Alleldiversität, Funktionalität, Pathogenität, Schwere der daraus entstehenden Erkrankung sowie experimentell ermittelte regulatorische Effekte und komplexer Assoziationen. Der C-Score gibt die „Schädlichkeit der Varianten“ (*deleteriousness*) wieder.

Für die einfache Interpretation wurde aus den ermittelten C-scores *phred-like scores (scaled C-scores)* definiert. Jede Variante wurde dabei in Relation zu allen 8,6 Milliarden möglichen SNVs betrachtet und einem Wert von 1 - 99 zugeordnet. Werte mit einem skalierten, logarithmischen C-Score > 10 beziehen sich auf die höchsten 10% der C-Scores ($-10\log_{10}$) und somit der am unwahrscheinlichsten zu beobachtenden Allele. Werte > 20 bezeichnen die höchsten 1% und Werte von > 30 bezeichnen die höchsten 0,1% unter allen C-Scores (Kircher et al., 2014). In die Auswertung mit einbezogen und als mutmaßlich schädlich definiert wurden alle Varianten mit einem log C-Score = CADD-Score von ≥ 20 .

Für die Einordnung von Missense-Varianten in MMR-Genen wurde zudem die Pathogenitätseinschätzung nach dem 5-Klassen-System des *Variante Interpretation Committees* (VIC) der *International Society for Gastrointestinal Hereditary Tumors* (InSiGHT) verwendet (<http://www.insight-group.org>) (Plazzer et al., 2013; Thompson et al., 2013). Sie basiert auf einer Kombination klinischer, familiärer, genetischer und laborchemischer Daten sowie den Ergebnissen funktioneller Tests. Anhand einer auf diesen Daten beruhenden Wahrscheinlichkeitsberechnung werden die Varianten einer von fünf Pathogenitätsklassen zugeordnet (Plon et al., 2008) (Tab. 13).

Tab. 13 Pathogenitätsklassen nach Plon et al.,2008

Klasse	Definition	Wahrscheinlichkeit Pathogenität
1	nicht pathogen/ keine klein. Signifikanz	0-0,1%
2	wahrscheinlich nicht pathogen / geringe klinische Signifikanz	0,1-5%
3	Varianten unklarer funktioneller Signifikanz (VUS)	6-94%
4	wahrscheinlich pathogen	95-99%
5	pathogen	99-100%

Material und Methoden

Varianten der Klasse 1 und 2 (nicht pathogen/wahrscheinlich nicht pathogen) wurden nicht berücksichtigt. Varianten der Klasse 4 und 5 der MMR-Gene wurden als pathogen betrachtet. Wurden bei Probanden im Laufe der Arbeit solche pathogenen Mutationen in MMR-Genen detektiert, galten diese Fälle als aufgeklärt und sie wurden von der weiteren Suche nach neuen Ursachen des LS ausgeschlossen. Die Varianten unklarer funktioneller Relevanz (*Variants of Unknown Significance, VUS*) bilden die Klasse 3. Sie wurden als „möglicherweise pathogen“ betrachtet und in den Ergebnissen dargestellt. Missense-Mutationen werden zum Großteil der Pathogenitätsklasse 3 zugeordnet. Da die Pathogenität von Missense-Mutationen häufig nicht klar bestimmt werden kann, stellen sie eine Herausforderung bei der Interpretation genetischer Varianten der MMR-Gene dar.

6. Validierung der gefilterten Varianten

Der *Integrative Genome Viewer* (IGV) diente der optischen Vorvalidierung der Varianten (Robinson et al., 2011). Hierfür werden die generierten BAM-files verwendet. Abb. 17 zeigt die IGV-Darstellung einer Missense-Mutation auf Chromosom 5. Die G>A-Transition liegt mit einem Anteil mutierter reads von 50% vor.



Abb. 17 Darstellung einer Missense-Mutation (G>A) auf Chromosom 5 im IGV

Alle relevanten Varianten nach Filterschritt Nr. 4 wurden zusätzlich mittels Sanger-Sequenzierung bestätigt.

4 Ergebnisse

4.1 Keimbahnmutationsanalyse

4.1.1 Voruntersuchungen und Ausschluss von Patienten mit ursächlichen Mutationen

Bei allen Patienten mit klinischem Verdacht auf Vorliegen eines LS war nach bestätigtem Ausfall von *MSH2/MSH6* im Tumorgewebe vor Durchführung dieser Studie bereits eine Mutationssuche durchgeführt worden. Die Mutationssuche zielte auf die pathogenen Punktmutationen und großen Deletionen/Duplikationen in *MSH2* und *MSH6* sowie auf eine Deletionsanalyse der letzten beiden Exons des *EPCAM*-Gens ab. Diese Keimbahn-Untersuchungen wurden an der Leukozyten-DNA der Patienten vorgenommen. Alle 95 Patienten ohne ursächliche Mutation wurden in diese Arbeit aufgenommen.

Bei 25 Patienten, die vor 2010 rekrutiert worden waren, erfolgte die Suche nach pathogenen Punktmutationen mit dem damals gebräuchlichen DHPLC-Screening-Verfahren und anschließender Sanger-Sequenzierung der auffälligen Exons. Diese heute in der Regel nicht mehr eingesetzte Methode ist aber weniger sensitiv als die direkte Sequenzierung des kompletten Gens. Aus diesem Grund wurde im Rahmen dieser Studie eine ergänzende Sanger-Sequenzierung des kompletten *MSH2*- und *MSH6*-Gens durchgeführt. Durch diese Vorgehensweise wurden vorher nicht detektierte pathogene *MSH2*- und *MSH6*-Keimbahnmutationen bei sieben der 25 Patienten identifiziert (Tab. 14) (Primer siehe Anhang Tab. B.1 und Tab. B.2).

Tab. 14 Nachträglich identifizierte pathogene *MSH2-/MSH6*-Mutationen des Kollektivs

Gen	Patienten-ID	IHC-Befund	Mutation	Konsequenz
<i>MSH2</i>	HNPCC_09	MSH2*	c.1276+1G>A	aberrantes Spleißen
<i>MSH6</i>	HNPCC_12	MSH2/MSH6	c.467C>G	p.Ser156Ter
<i>MSH2</i>	HNPCC_13	MSH2/MSH6	c.942+3A>T	aberrantes Spleißen
<i>MSH2</i>	HNPCC_14	MSH2*	c.942+3A>T	aberrantes Spleißen
<i>MSH6</i>	HNPCC_15	MSH2/MSH6	c.1190_1191delAT	p.Tyr397CysfsX3
<i>MSH2</i>	HNPCC_16	MSH2/MSH6	c.1165C>T	p.Arg389Ter
<i>MSH6</i>	HNPCC_17	MSH2/MSH6	c.3261dupC	p.Phe1088Leufs*5

* *MSH6* nicht untersucht

Neun Patienten des Gesamtkollektivs wiesen Varianten unklarer klinischer Signifikanz (VUS) (InSiGHT Klasse 3) in *MSH2* oder *MSH6* auf. Im Laufe der Analyse wurde bei vier Patienten mit unklaren Varianten eine Hochstufung dieser VUS auf Klasse 4 (wahrscheinlich pathogen) oder Klasse 5 (erwiesenermaßen pathogen) vorgenommen. Diese vier Patientenfälle gelten demnach ebenfalls als aufgeklärt. Bei einem weiteren Patienten bestand nach erneuter Aktensichtung eine *MUTYH*-assoziierte Polyposis; anschließend wurden bialele Keimbahnmutationen im *MUTYH*-Gen identifiziert (siehe Kap. 4.2 Differentialdiagnosen).

Die Daten dieser insgesamt 12 Patienten, für die nachträglich eine genetische Ursache der Erkrankung gefunden wurde, wurden aus dem Datensatz entfernt. Folglich standen noch 83 Patienten für die Suche nach neuen genetischen Ursachen für das LS zur Verfügung.

4.1.2 Untersuchung seltener Ursachen eines *MSH2*-Ausfalls

Zur Suche nach weiteren in der Literatur beschriebenen Veränderungen, die mit einem *MSH2*-Ausfall einhergehen, wie z.B. intronische *MSH2*-Mutationen, *MSH2*-Inversionen und Mutationen des *MSH2*-Promotors, führten wir im Rahmen eines Kooperationsprojektes mit der Humangenetik der Ludwig-Maximilians-Universität München und des Medizinisch Genetischen Zentrums München ergänzende Untersuchungen durch. Eingeschlossen wurden hierfür 84 HNPCC-Patienten, davon 52 aus diesem Kollektiv (Primer Tab. C.1 und Tab. C.2 im Anhang).

Die beschriebene intronische *MSH2*-Mutation sowie die 10 kb große *MSH2*-Inversion zwischen Exon 1 und 7 wurde bei keinem der Patienten gefunden. Einer der Patienten wies eine unklare Variante (c.-82G>C) in der *MSH2*-Promotorregion auf.

4.1.3 Kandidatengensequenzierung des Proteasom-Signalwegs

Der Proteasom-Signalweg beeinflusst unter anderem den Abbau von MSH2 in den Zellen. Somatische Deletionen der beteiligten Gene können in bestimmten Tumoren eine Degradierung von MSH2 herbeiführen (Diouf 2012). Als Pilotprojekt vor der CNV- und NGS-Analyse erfolgte deshalb eine Sanger-Sequenzierung der kodierenden Bereiche von vier Genen des Proteasom-Signalwegs (*HERC1*, *MTOR*, *PIK3C2B* und *PRKCZ*). Hierbei sollte untersucht werden, ob auch Keimbahnmutationen dieser vier Gene über den beschriebenen Interaktionsmechanismus zum MSH2-Ausfall in unserem Patientenkollektiv geführt haben könnten. Die Sequenzierung der vier Gene wurde an fünf nicht-verwandten Patienten (HNPCC_20; HNPCC_24; HNPCC_26; HNPCC_40 und HNPCC_41) mit besonders auffälliger Eigen- und Familiengeschichte durchgeführt.

Hierbei wurde lediglich bei einem Patienten (HNPCC_24) eine Missense-Variante in *PRKCZ* (c.520G>A;p.Gly174Ser) entdeckt. Diese Variante wird laut Exome Aggregation Consortium (ExAc) als seltene Variante geführt (MAF 0,048%) Der CADD-Score liegt bei 23. Er gehört damit zu den höchsten 1% aller ermittelten C-Scores und ist damit als potentiell schädlich anzusehen (vgl. Kap. 3.10.3 Varianten-Filterung „5. Abschätzung der Pathogenität“) (Primer siehe Anhang Tab. D.1. bis Tab. D.4).

4.2 Differentialdiagnosen

Bei allen Patienten des Bonner Kollektivs wurde nach biallelen *MUTYH*-Mutationen und nach pathogenen *POLE*- und *POLD1*-Mutationen gesucht, um der klinischen Differentialdiagnose einer MAP bzw. PPAP nachzugehen.

Bei einem Patienten wurden zwei wahrscheinlich biallele *MUTYH*-Keimbahnmutationen detektiert, sodass der hochgradige Verdacht auf Vorliegen einer *MUTYH*-assoziierten Polyposis (MAP) besteht (Tab. 15).

Vier unterschiedliche heterozygote Missense-Varianten in *POLE* wurden bei vier Patienten detektiert. Die CADD-Scores liegen zwischen 26 und 35, was für eine Pathogenität dieser

Ergebnisse

Varianten spricht. Keine der Mutationen befand sich jedoch in der Exonuklease-(Proofreading) Domäne des Proteins, die bisher als verantwortlich für die Ausbildung des Phänotyps gilt (Tab. 15).

Tab. 15 Missense Mutationen in *MUTYH* und *POLE*

Gen	Patienten-ID	Variante	vermutete Konsequenz	CADD-Score	HI-Score; RVIS %	ExAC MAF
<i>MUTYH</i>	HNPCC_30	c.504+19_31del13 c.1012C>T	aberrantes Spleißen p.Gln335X	-	64; 54	-
<i>POLE</i>	HNPCC_23	c.2770C>T	p.Arg924Cys	34		0,002%
<i>POLE</i>	HNPCC_37	c.274A>C	p.Ser92Arg	28	11; 3	0,006%
<i>POLE</i>	HNPCC_52	c.4709G>A	p.Arg1570Gln	35		-
<i>POLE</i>	HNPCC_65	c.5492T>C	p.Leu1831Pro	26		0,002%

4.2.1 Somatische Mutationen

Das Institut für Pathologie des Universitätsklinikums Bonn und das Klinikum der Universität zu Köln stellten für die Suche nach biallelen somatischen Mutationen in *MSH2* und *MSH6* Tumor-DNA von 11 Patienten bereit. Für diese Untersuchung wurde das NGS *TruSight Cancer Panel* (Illumina) verwendet. Alle Varianten in *MSH2* und *MSH6*, die in mindestens zehn reads vorkamen, wurden ausgewertet.

Es wurden drei heterozygote, potentiell pathogene, somatische Mutationen (CADD-Score > 20) bei zwei der elf untersuchten Patienten detektiert (Tab. 16). Keine der Varianten ist in der InSiGHT-LOVD-Datenbank oder der Adenokarzinom-Exom-Datenbank des *The Cancer Genome Atlas* (TCGA) aufgeführt. Bei einem der zwei Patienten (HNPCC_23) wurden zwei potentiell biallele Missense-Mutationen in *MSH2* gefunden. Ein CADD-Score von > 20 indiziert, dass die Varianten unter den höchsten 1% aller ermittelten C-Scores liegen und damit möglicherweise pathogen sind.

Für die Suche nach größeren Deletionen im Tumor, die zusammen mit einer pathogenen Punktmutation für ein bialleles Ereignis sprechen, wurde eine MLPA mit der Tumor-DNA der elf Patienten durchgeführt. Aufgrund der unzureichenden Qualität der Tumor-DNA konnten diese Daten jedoch nicht ausgewertet werden.

Tab. 16 Somatische Punktmutationen

Gen	Patienten-ID	Variante	Protein-Änderung	Mutationstyp	CADD-Score	FA
<i>MSH2</i>	HNPCC_23	c.1993C>G	p.His665Asp	Missense	25	Mutter „Ca“
<i>MSH2</i>		c.1835C>T	p.Ser612Leu	Missense	26	
<i>MSH2</i>	HNPCC_06	c.1100delT	p.Val367GlufsTer6	Frameshift	35	-

FA: Familienanamnese; Ca: Karzinom; CRC: Kolorektales Karzinom

4.3 CNV-Analyse

Eine Keimbahn-CNV-Analyse wurde durchgeführt, um seltene, hochpenetrante Mikrodeletionen und -duplikationen in bisher nicht mit LS assoziierten, potentiell ursächlichen Genen zu identifizieren. Die Analyse wurde an 83 nicht-verwandten Index-Patienten mit erfüllten klinischen Kriterien für ein HNPCC; MSH2-Proteinausfall im Tumorgewebe, aber nicht identifizierbarer Keimbahnmutation in *MSH2* durchgeführt. Alle Patienten wurden auf dem HumanOmniExpress Bead-Array von Illumina genotypisiert.

4.3.1 Suche nach Kandidatengen

Qualitätskontrolle

Als Qualitätskontrolle wurden für jede Probe die Standardabweichung Log R Ratio (LRR) und die Call-Rate bestimmt (Vgl. Kapitel 3.9.2). Zehn Patienten wiesen eine Standardabweichung der LRR von größer 0,3 auf. Da ihre Intensitätssignale die Genauigkeit der detektierten CNVs beeinflussen, wurden diese Proben aus der Analyse entfernt. Die eingeschlossenen 73 Patienten wiesen eine Standardabweichung der LRR zwischen 0,11 bis 0,29 auf. Die erforderliche Call-Rate von $\geq 98\%$ wurde bei allen Proben erreicht. Die durchschnittliche Call-Rate lag bei 99,7%.

CNV-Calling

Als *CNV-Calling* wird die Berechnung der CNVs über die Signalintensitäts- und Genotypisierungsdaten bezeichnet. Hierfür wurde der QuantiSNP-Algorithmus verwendet. Tab. 17 gibt eine Übersicht über die nach dem *CNV-Calling* erhaltenen CNVs.

Tab. 17 Berechnete CNVs (QuantiSNP) nach Qualitätskontrolle

Anzahl Patienten	Anzahl CNVs	Del	Dup	CNV/Patient (Durchschnitt)	CNV/Patient min.	CNV/Patient max.
73	8.460	2.025	6.435	116	26	190

Herausfiltern von Artefakten

Nach Anwendung der in Kapitel 3 dargestellten Filterkriterien („Filterkriterien der CNV-Analyse“, S. 38) (Größe der CNVs ≥ 10 kb; Anzahl konsekutiver SNP-Marker ≥ 5 SNPs; Log-Bayes-Faktor ≥ 10 für Deletionen bzw. ≥ 20 für Duplikationen) reduzierte sich die Zahl der CNVs von 8.460 auf 628. Zwei Patienten (HNPCC_81 und HNPCC_87) zeigten ab diesem ersten Filterschritt keine CNVs mehr (vgl. Tab. 18). Die größte Anzahl an mutmaßlich echten CNVs lag bei 17 CNVs pro Patient und wurde bei zwei Patienten (HNPCC_05 und HNPCC_45) ermittelt. Im Durchschnitt wiesen die Patienten 9 CNVs auf.

Tab. 18 CNVs nach erstem Filterschritt

	Anzahl Personen	Anzahl CNVs	Del.	Dup.	CNV/Patient (Durchschnitt)	CNV Größe
Patienten	71	628	367	261	9	10 kb– 3,9 Mb

4.3.2 CNVs in bekannten CRC-Genen und Interaktionspartnern von MSH2

Alle im Bereich der durch QuantiSNP berechneten, qualitätskontrollierten, ungefilterten 8.460 CNVs der Patienten wurden mit den 97 bekannten CRC-Genen und Interaktionspartnern von *MSH2* ± 50 kb verglichen (siehe Anhang Tab. F.1). Es wurden 171 Patienten-CNVs detektiert, welche sich im Genbereich (± 50 kb) von 22 der 97 Kandidatengene lokalisieren ließen. Nach Anwendung des Log Bayes Faktor-Filters ≥ 10 für den Ausschluss von Artefakten blieben noch 21 potentiell echte CNVs von 18 Patienten übrig. Diese 21 CNVs umfassten die Genbereiche ± 50 kb von *POLE*, *F10*, *MCM4* und *HRAS*. Nach Abgleich dieser CNVs mit den HNR-Kontrollen blieb eine 90 kb große Deletion übrig die mit keinem CNV der HNR-Kontrollen überlappte und den direkt kodierenden Bereich von *MCM4* betraf (siehe Tab. 19). Alle weiteren CNVs lagen als Duplikation (CNV=3) oder Multiplikation (CNV=4) vor und überlappten ≥ 35 Mal mit den CNVs der HNR-Kontrollen.

Tab. 19 Suche nach CNVs in bekannten CRC-Genen und Interaktionspartnern von MSH2

Gen	Patienten-ID	Chr	start (bp)	stop (bp)	Länge	Anzahl SNP-Marker	CNV	LBF
<i>MCM4</i>	HNPCC_41	8	48792199	48882899	90701	13	Deletion	43

4.3.3 Filterschritte auf Genebene

Seltene CNVs im Bereich proteinkodierender Gene

Für die Suche nach seltenen CNVs im Bereich proteinkodierender Gene wurden weitere Filterschritte angewandt (proteinkodierende CNVs, < 3-mal in 1320 *In-House*-Kontrollen; < 0.2% in der *Database of Genomic Variants*; Ausschluss von CNVs in Regionen segmentaler Duplikationen).

Durch die Beschränkung auf CNVs im Bereich proteinkodierender Gene und den Vergleich mit den *In-House*-Kontrollen wurde die Anzahl der CNVs von 628 auf 55 stark reduziert. Der Vergleich mit der *Database of Genomic Variants* brachte den Ausschluss von drei CNVs im Bereich von insgesamt vier proteinkodierenden Genen (Tab. 20). CNVs, die durch Überlappung mit Bereichen segmentaler Duplikationen ausgeschlossen wurden, sind in Tab. 21 aufgelistet. Hiervon waren sechs CNVs mit neun Genen betroffen.

Tab. 20 Ausgeschlossene CNVs durch den Vergleich mit der Database of Genomic Variants

Gen- Name	Patienten- ID	Chromosomale Position	CNV	Anzahl CNVs pro Anzahl Kontrollen	Studien
<i>CSMD1</i>	HNPCC_17	chr8:3683879-3711034	Del	39/270; 52/1109	(Redon et al., 2006; Vogler et al., 2010)
<i>CSMD1</i>	HNPCC_23	chr8:4555958-4597645	Dup	6/1109; 16/771	(Pinto et al., 2007; Vogler et al., 2010)
<i>C9orf66</i> <i>DOCK8</i>	HNPCC_74	chr9:185632-271455	Dup	8/270; 6/1109; 29/771; 10/2026; 8/1557	(Itsara et al., 2009; Pinto et al., 2007; Redon et al., 2006; Shaikh et al., 2009; Vogler et al., 2010)

Tab. 21 Ausgeschlossene CNVs aufgrund Überlappungen mit Bereichen segmentaler Duplikationen

Gen-Name	Patienten-ID	Chromosomale Position	CNV
<i>UGT2B10</i>	HNPCC_50	chr4:69678679-69706273	Dup
<i>OCLN; GTF2H2C; GUSBP3</i>	HNPCC_45	chr5:68749848-69175823	Dup
<i>TRIM49B; TRIM64C</i>	HNPCC_86	chr11:48747611-49124182	Dup
<i>TRIM48</i>	HNPCC_86	chr11:54794237-55049310	Dup
<i>CELA3B</i>	HNPCC_29	chr1:22296307-22310080	Del
<i>OCM</i>	HNPCC_45	chr7:5883522-5925156	Del

Validierung der CNVs mit GenomeStudio und qPCR

Die 45 übrigen CNVs wurden mit GenomeStudio und qPCR validiert. Zwei der CNVs stellten sich in der qPCR als falsch-positiv heraus (Tab. 22). Die proteinkodierende Region lag bei beiden Patienten und den vier Kontrollpersonen jeweils in der gleichen Kopienzahl (CNV = 2) vor. Die beiden CNVs waren insgesamt sehr klein (11 kb und 42 kb) und wiesen wenig Marker auf (5 und 7). Der niedrige LBF von 12 und 15 indiziert zudem eine geringe Wahrscheinlichkeit für die Echtheit der CNVs.

Tab. 22 Ausgeschlossene CNVs durch Validierung mit qPCR (falsch-positive CNVs)

Gen	Patienten-ID	chromosomale Position	CNV	Größe (kb)	Anzahl SNP-Markern	LBF
<i>TP63</i>	HNPCC_36	chr3:189364424-189379493	Del ^{intr}	42	7	15
<i>STARD13</i>	HNPCC_17	chr13:34140341-34151488	Del ^{intr}	11	5	12

Del: Deletion, intr: intronischer Bereich des Gens

Ergebnisse der Filterschritte

Nach Anwendung der genannten Filterschritte und der qPCR-Validierung verblieben 43 echte CNVs (21 heterozygote Deletionen und 22 Duplikationen) bei 30 der 73 Patienten (41%). Diese 43 CNVs umfassen 76 proteinkodierende Gene. Jeder CNV trat nur einmal auf und es gab keine Überschneidungen zwischen den CNVs. Auch war kein Gen doppelt / mehrfach von einem CNV betroffen. Alle CNVs waren im Vergleich zu den *In-House*-Kontrollen und DGV selten (< 0.2%). 28 CNVs wurden von keinem der *In-House*-Kontrollen geteilt und gelten somit als „neu“ (*novel*). Bei 15 Patienten-CNVs traten Überschneidungen zu ein bis zwei Kontroll-CNVs auf. Sie waren nach unserer Definition „seltene CNVs“. Tab. 23 gibt eine Gesamtübersicht über die Anzahl der

CNVs und proteinkodierenden Gene bezogen auf die Anzahl der Patienten pro Filterschritt. Abb. 18 zeigt das Fluss-Diagramm der Filterschritte mit der jeweiligen Anzahl der CNVs.

Tab. 23 Anzahl detektierter CNVs nach jedem Filterschritt

Filterschritte	QuantiSNP <i>calling</i> , QC	Größe, SNP- Marker, LBF	proteinkodierende Gene, HNRs	DGV, Segmentale Duplikationen	Validierung
# Patienten	73	71	36	32	30
CNVs	8460	628	56	45	43
Gene	n. u.	n. u.	92	78	76

n. u. nicht untersucht



Abb. 18 Flussdiagramm CNV-Analyse

QC: Qualitätskontrolle, LRR: Log R Ratio

4.3.4 Priorisierung der Gene

Die 76 gefilterten Gene wurden in zwei parallelen Schritten priorisiert, um funktionell relevante sowie mutationsensitive Gene aufzuspüren.

1. Art der Mutation und funktionelle Relevanz des betroffenen Gens (Kriterium I)

Für dieses erste Priorisierungskriterium wurde berücksichtigt, ob das vom CNV betroffene Gen in einem Zusammenhang zur Karzinogenese steht und ob die Art der Mutation (Deletion/Duplikation, partiell/komplett) und der daraus abzuleitende Effekt (*Loss of function/Gain of function*) mit der beschriebenen Funktion übereinstimmt (vgl. 2.2.4 Pathomechanismus von CNVs und Kap. 3.9.3 Filterkriterien der CNV-Analyse). Bei der Einordnung partieller Duplikationen wurde davon ausgegangen, dass potentiell immer auch ein LoF-Effekt denkbar ist.

Unter den 76 proteinkodierenden Genen befanden sich sechs Gene (8%), die vollständig von einer heterozygoten Deletion, 21 Gene (28%), die vollständig von einer Duplikation und 49 Gene (64%), die partiell von einem CNV (Deletion oder Duplikation) betroffen waren. Zehn dieser partiell betroffenen Gene lagen ausschließlich im Intronbereich des jeweiligen Gens (Tab. 24).

Unter den sechs Genen, die von einer kompletten heterozygoten Gen-Deletion (CNV = 1) betroffen waren, werden drei der Gene (*VWDE*, *ZNF154* und *ZNF671*) im direkten Zusammenhang mit Krebs erwähnt. *VWDE* (Von Willebrand Faktor D und EGF Domäne) ist allerdings als Onkogen beschrieben. Da dies nicht zu dem *Loss-of-Function*-Modell einer kompletten heterozygoten Gen-Deletion passt, wurden unter den kompletten Gen-Deletionen nur *ZNF154* und *ZNF671* als mögliche Kandidatengene berücksichtigt (Tab. 24, CNV: del, Kriterium: I). Unter den 21 kompletten Gen-Duplikationen wurden vier Gene (*DDX24*, *GNB2L1*, *LAMP3* und *PFKP*) als potentielle Onkogene eingestuft (Tab. 24, CNV: dup, Kriterium: I). Unter den Genen, welche partiell von einem CNV betroffen waren, wurde für 38% (19/49) ein Zusammenhang zu malignen Erkrankungen beschrieben (Tab. 24, CNV: del^{part} oder dup^{part} bzw. del^{intr} oder dup^{intr}, Kriterium: I).

2. Bioinformatische Scores zur Bewertung der Sensitivität von LoF-Mutationen (Kriterium II)

Der Haploinsuffizienz- und Intoleranz-Score wurde für die Bewertung der Sensitivität von LoF-Mutationen herangezogen (vgl. Kap. 3.9.3. Filterkriterien der CNV-Analyse).

Unter den 55 Genen, die von einer vollständigen Deletion oder partiellen Deletion/Duplikation betroffen waren, weisen 23 einen niedrigen Haploinsuffizienz-Score (HI-Score \leq 35%) und/ oder

Intoleranz-Score (RVIS \leq 25%) auf. Dies spricht dafür, dass bei diesen Genen der Verlust bzw. die mutationsbedingte Inaktivierung eines Allels kritische Auswirkungen auf die Proteinfunktion haben. Unter den dosissensitiven/variantenintoleranten Genen befinden sich auch solche Gene, die bisher in keinem Zusammenhang zu malignen Erkrankungen stehen. In Tab. 24 werden alle 76 detektierten CNV-Gene mit CNV-Status, Scores und ggf. Angabe des erfüllten Kriteriums (I/II) aufgelistet.

Solche Gene, die sowohl Kriterium I (relevante Funktion in Bezug auf die Art der Mutation) als auch Kriterium II (dosissensitive/variantenintolerante Gene) erfüllen, sind in Tab. 25 mit einer kurzen Funktionsbeschreibung aufgeführt. Dies trifft auf 13 der 76 Gene (17%) zu.

Tab. 24 Alle 76 Gene mit CNV-Status, Scores und Angabe des Kriteriums (I / II).

Gen	Patienten-ID	CNV	Hi-Score%	RVIS %	Kriterium
<i>ABCB1</i>	HNPCC_34	Del ^{intr}	14	23	II
<i>AGMO</i>	HNPCC_29	Del ^{part}	-	96	-
<i>ASB2</i>	HNPCC_01	Dup	(77)	(96)	-
<i>ATP12A</i>	HNPCC_97	Dup ^{part}	70	72	-
<i>B3GNT5</i>	HNPCC_73	Dup	(78)	(47)	-
<i>C9orf84</i>	HNPCC_07	Dup ^{part}	60	88	-
<i>CAPN13</i>	HNPCC_32	Dup ^{part}	91	96	-
<i>CHST2</i>	HNPCC_97	Dup	(61)	(31)	-
<i>DCUN1D1</i>	HNPCC_73	Dup ^{part}	31	56	I und II
<i>DDX24</i>	HNPCC_01	Dup	(50)	(79)	I
<i>DLG2</i>	HNPCC_28	Del ^{intr}	1	4	I und II
<i>DPP10</i>	HNPCC_72	Dup ^{intr}	30	50	I und II
<i>DUSP11</i>	HNPCC_45	Dup ^{part}	53	18	I und II
<i>ENOX1</i>	HNPCC_24	Dup ^{part}	22	9	I und II
<i>ESYT2</i>	HNPCC_92	Dup	-	(38)	-
<i>FAM181A</i>	HNPCC_01	Dup	-	(53)	-
<i>FBXO36</i>	HNPCC_24	Del ^{part}	53	89	-
<i>GNB2L1</i>	HNPCC_38	Dup	(60)	(25)	I
<i>GNG10</i>	HNPCC_07	Dup	(45)	(56)	-
<i>GRIK4</i>	HNPCC_94	Del ^{part}	41	16	II
<i>HMHB1</i>	HNPCC_40	Dup	(96)	(78)	-
<i>IFI27L1</i>	HNPCC_01	Dup ^{part}	-	86	-
<i>IQSEC3</i>	HNPCC_26	Dup ^{part}	17	30	II
<i>ITIH2</i>	HNPCC_89	Dup ^{part}	51	72	I
<i>ITIH5</i>	HNPCC_89	Dup	(91)	-	-
<i>KATNAL1</i>	HNPCC_90	Dup ^{part}	29	20	II
<i>KLHDC4</i>	HNPCC_33	Del ^{part}	36	96	-
<i>KLHL24</i>	HNPCC_73	Dup	(49)	(20)	-
<i>KLHL32</i>	HNPCC_91	Del ^{intr}	86	16	II
<i>KLHL6</i>	HNPCC_73	Dup	(46)	(35)	-
<i>LAMP3</i>	HNPCC_73	Dup	(70)	(71)	I
<i>LCLAT1</i>	HNPCC_32	Dup ^{part}	-	34	-

Ergebnisse

<i>MCCC1</i>	HNPCC_73	Dup	(28)	(83)	-
<i>MCF2L2</i>	HNPCC_73	Dup	(12)	(38)	-
<i>MCM4</i>	HNPCC_41	Del ^{part}	2	4	I und II
<i>MTHFD1L</i>	HNPCC_50	Dup ^{part}	86	43	-
<i>MYH6</i>	HNPCC_18	Dup ^{part}	5	10	II
<i>MYH7</i>	HNPCC_18	Dup ^{part}	19	0.3	II
<i>NCAPG2</i>	HNPCC_92	Dup	(9)	(46)	-
<i>NMNAT3</i>	HNPCC_18	Del ^{intr}	36	50	-
<i>NOSIP</i>	HNPCC_18	Del	51	25	II
<i>NPS</i>	HNPCC_45	Del	100	76	-
<i>NRG3</i>	HNPCC_38	Del ^{intr}	62	30	I
<i>OTUB2</i>	HNPCC_01	Dup	(9)	(47)	-
<i>PAG1</i>	HNPCC_47	Del ^{part}	35	39	I und II
<i>PDZRN4</i>	HNPCC_41	Dup ^{intr}	71	68	I
<i>PFKP</i>	HNPCC_28	Dup	(45)	(5)	I
<i>PITPNC1</i>	HNPCC_21	Del ^{intr}	15	19	II
<i>PITRM1</i>	HNPCC_28	Dup ^{part}	81	99	-
<i>PLEKHG1</i>	HNPCC_50	Dup ^{part}	61	52	-
<i>PRIMA1</i>	HNPCC_01	Dup ^{part}	56	48	I
<i>PRKCA</i>	HNPCC_42	Del ^{intr}	5	9	I und II
<i>PRKDC</i>	HNPCC_41	Del ^{part}	9	-	I und II
<i>PRRG2</i>	HNPCC_18	Del ^{part}	72	80	-
<i>RAP1GDS1</i>	HNPCC_38	Del ^{part}	44	35	-
<i>RBM19</i>	HNPCC_38	Dup ^{part}	40	88	I
<i>RUNDC3B</i>	HNPCC_34	Del ^{part}	66	19	I und II
<i>SLC9A9</i>	HNPCC_97	Dup	(43)	(7)	-
<i>SLCO4C1</i>	HNPCC_31	Dup ^{part}	38	21	I und II
<i>SRBD1</i>	HNPCC_31	Del ^{intr}	53	83	-
<i>TPRKB</i>	HNPCC_45	Dup	(43)	(54)	-
<i>TRIM41</i>	HNPCC_38	Dup ^{part}	83	62	I
<i>TRIM52</i>	HNPCC_38	Dup	(88)	(45)	-
<i>TRMT44</i>	HNPCC_43	Del ^{part}	75	91	-
<i>TSPAN5</i>	HNPCC_38	Del ^{part}	32	23	I und II
<i>TSPAN9</i>	HNPCC_18	Dup ^{part}	59	39	I
<i>VWDE</i>	HNPCC_32	Del	-	-	-
<i>WDR60</i>	HNPCC_92	Dup ^{part}	76	91	-
<i>WFDC10A</i>	HNPCC_55	Del	92	69	-
<i>WFDC11</i>	HNPCC_55	Del ^{part}	88	61	-
<i>WFDC9</i>	HNPCC_55	Del ^{part}	99	75	-
<i>YEATS2</i>	HNPCC_73	Dup ^{part}	74	42	-
<i>ZNF154</i>	HNPCC_79	Del	66	92	I
<i>ZNF551</i>	HNPCC_79	Del ^{part}	94	35	-
<i>ZNF671</i>	HNPCC_79	Del	89	22	I und II
<i>ZNF776</i>	HNPCC_79	Del ^{part}	94	18	II

Del: Deletion; Dup: Duplikation

intr: CNV betrifft nur intronischen Bereich;

part: kodierender Gen-Bereich partiell vom CNV betroffen

Kriterium I: Funktionell relevant in Bezug zur Art der Mutation;

Kriterium II: Gene mit einem Hi-Score ≤ 35 oder RVIS ≤ 25

Werte in Klammern: Komplette Gen-Duplikation, Scores nicht vollständig anwendbar

Tab. 25 Funktionell relevante Gene mit zusätzlicher Haploinsuffizienz/Variantenintoleranz

Gen	Patienten-ID	CNV-Art	Funktion/involverter Signalweg
<i>DCUN1D1</i>	HNPCC_73	Dup ^{part}	Plattenepithelzellkarzinom der Lunge und Adrenocorticales-Karzinom. Bewirkt Zelltransformation und Fortschreiten der Karzinogenese
<i>DLG2</i>	HNPCC_28	Del ^{intr}	Mögliches TSG. An der Regulation der Zellproliferation und Apoptose beteiligt.
<i>DPP10</i>	HNPCC_72	Dup ^{intr}	CRC-Gen. Expressionsverlust von DPP10 in CRC wird in Zusammenhang mit dem Fortschreiten der Karzinogenese gebracht
<i>DUSP11</i>	HNPCC_45	Dup ^{part}	Inhibiert Mitglieder des MAPK-Signalwegs
<i>ENOX1</i>	HNPCC_24	Dup ^{part}	Zellverteidigung, Zellwachstumskontrolle, Zellüberleben. Assoziiert mit Prostata-Karzinom
<i>MCM4</i>	HNPCC_41	Del ^{part}	Initialisierung der Genom-Replikation. Überexpression in verschiedenen Karzinomen (Brust, Ösophagus, Meningeom)
<i>PAG1</i>	HNPCC_47	Del ^{part}	Lymphom; Prostata-Ca; mitogen-activated protein kinase (MAPK)/phosphatidylinositol 3-kinase (PI3K) Signalweg
<i>PRKCA</i>	HNPCC_42	Del ^{intr}	TSG, Fördert β -Catenin-Degradierung/-Phosphorylierung; Colon-, Prostata-, Leber-Karzinom
<i>PRKDC</i>	HNPCC_41	Del ^{part}	Reparatur von Doppelstrangbrüchen und Rekombination. Funktionsverlust führt zu vermehrten Falschpaarungen in der DNA der betroffenen Zellen
<i>RUNDC3B</i>	HNPCC_34	Del ^{part}	Assoziiert mit Brust- und Lungenkrebs, interagiert mit der MAPK-Kaskade Zellproliferation und -differenzierung. Karzinogenese in Lymphdrüsenkrebs
<i>SLCO4C1</i>	HNPCC_31	Dup ^{part}	NOTCH und TP53 Singnalweg. Inaktiviert über Methylierung / somatische Mutationen in Tumorzellen
<i>TSPAN5</i>	HNPCC_38	Del ^{part}	Signaltransduktionsmediation für die Zellregulation, Zellaktivierung und Wachstum
<i>ZNF671</i>	HNPCC_79	Del	Transkriptionsregulator, Zellwachstum und -differenzierung, Hypermethylierung in Tumorzellen von Nieren-; Cervix- und Urothel-Karzinom

Del: Deletion; Dup: Duplikation; TSG: Tumorsuppressorgen;
intr: CNV betrifft nur intronischen Bereich, part: kodierender Gen-Bereich partiell vom CNV betroffen

4.3.5 Suche nach Long Range Regulatoren von MSH2

Um nach Veränderungen in *long range* Regulatoren von *MSH2* zu suchen (Kap 2.2), wurden alle CNVs $\pm 1,5$ Mb um *MSH2* separat untersucht. Nach Abgleich dieser CNVs gegen die *In-House*-Kontrollen konnte kein seltener CNV in dieser Region festgestellt werden.

4.3.6 Netzwerk-Analyse

Für die Suche nach direkten Interaktionen zwischen den bei der CNV-Analyse erhaltenen Genen (bzw. ihren daraus kodierenden Proteinen) und MSH2, wurde eine Netzwerk-Analyse mit Hilfe der STRING-Datenbank (*Search Tool for the Retrieval of Interacting Genes/Proteins*) durchgeführt. Für die Analyse wurden alle 76 Gene gemeinsam mit *MSH2* in die Datenbank hochgeladen. Es wurden insgesamt 12 der 76 Gene (16%) als direkte oder indirekte Interaktionspartner von MSH2 ausfindig gemacht (*ABCB1*, *ATP12A*, *GNB2L1*, *MCM4*, *NRG3*, *PAG1*, *PRKCA*, *PRKDC*, *RUNDC3B*, *SLC9A9*, *TRIM41* und *TRIM52*) (Abb. 20). Eine experimentell begründete Evidenz für die vorliegenden Interaktionen liegt für sieben der zwölf Proteine (*GNB2L1*, *MCM4*, *PAG1*, *PRKCA*, *PRKDC*, *TRIM41* sowie *TRIM52*) vor. Die Protein-Interaktion zwischen *MCM4* und *MSH2* welche zuvor bei der Interaktionspartnersuche festgestellt wurde (Kap. 4.3.2) konnte über die Netzwerk-Analyse bestätigt werden.

Mit *PRKDC* wurde ein weiterer direkter Interaktionspartner ausfindig gemacht. *PRKDC* ist laut STRING-Analyse ein Bindungspartner von *MSH2*. *MCM4* ist für die Initiierung der genomischen Replikation verantwortlich. *PRKDC* ist an der Reparatur von Doppelstrangbrüchen involviert. Die 91 kb große Deletion überspannt beide Gene bis zur Hälfte. Abb. 19 zeigt die UCSC-GenomeBrowser-Abbildung der Gene *MCM4* und *PRKDC* und des zugehörigen CNVs (in rot).

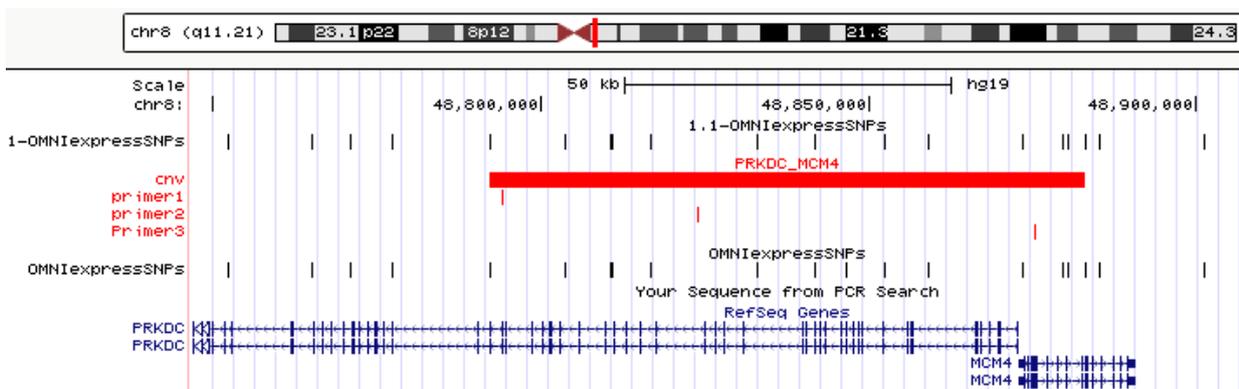


Abb. 19 *MCM4*_PRKDC Deletion HNPCC_41 (UCSC GenomeBrowser hg19)

Über *PRKDC* und *PRKCA*, welche viele verschiedene Verbindungen zueinander aufweisen (gegenseitige Aktivierung oder Inhibierung, Expressionsbeeinflussung und Bindungspartnerschaften), sind weitere Gene an das Netzwerk angeschlossen. Tab. 26 listet alle 12 Proteine auf.

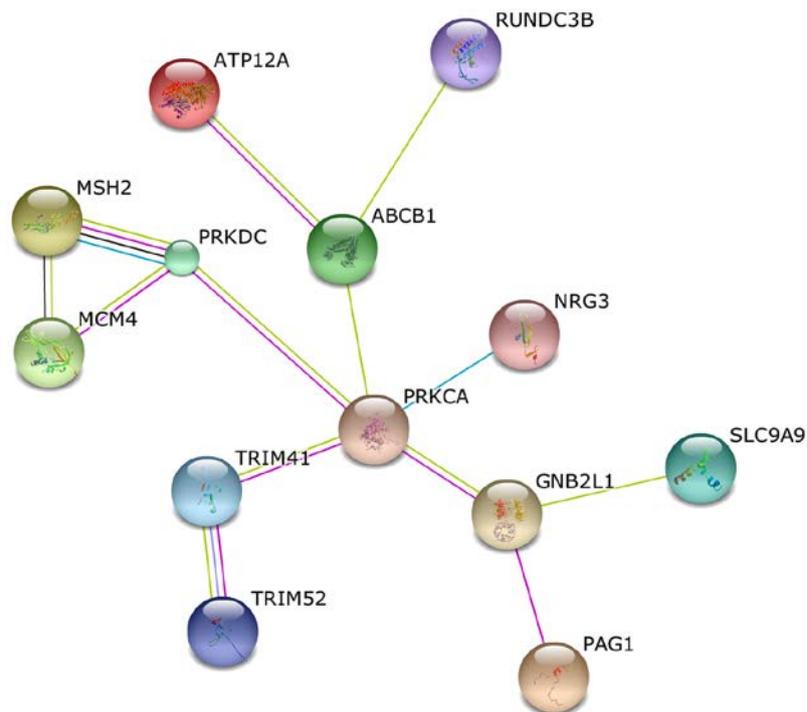


Abb. 20 Netzwerkpartner von MSH2 nach STRING-Netzwerk-Analyse – Evidenz

Linienfarbe: gelb: algorithmus-basiert, schwarz: Co-expression, pink: experimentell, hellblau: Datenbank
 Einstellungen STRING: Spezies: Homo sapiens, Interaktionsdatenbank-Referenzen: Alle; Min. geforderter Interaktions-Score: Mittleres Konfidenzlevel (0,4); Nur Interaktionen zwischen eingegebenen Proteinen

Tab. 26 Netzwerkpartner von MSH2

Protein	CNV	Patientencode	HI-Score%	RVIS%
ABCB1	Del ^{intr}	HNPCC_34	14	23
ATP12A	Dup ^{part}	HNPCC_97	70	72
GNB2L1	Dup	HNPCC_38	(60)	(25)
MCM4	Del ^{part}	HNPCC_41	2	4
NRG3	Del ^{intr}	HNPCC_38	62	30
PAG1	Del ^{part}	HNPCC_47	35	39
PRKCA	Del ^{intr}	HNPCC_42	5	9
PRKDC	Del ^{part}	HNPCC_41	9	-
RUNDC3B	Del ^{part}	HNPCC_34	66	19
SLC9A9	Dup	HNPCC_97	(43)	(7)
TRIM41	Dup ^{part}	HNPCC_38	83	62
TRIM52	Dup	HNPCC_38	(88)	(45)

Del: Deletion; Dup: Duplikation

intr: CNV betrifft nur intronischen Bereich

part: Gen partiell vom CNV betroffen (CNV überlappt Exon-Bereiche);

Werte in Klammern: Komplette Gen-Duplikation, Scores nicht genügend anwendbar

4.3.7 Pathway-Analyse

Die Ingenuity Pathway-Analyse (Qiagen) wurde verwendet, um eine signifikante Anreicherung der betroffenen Gene der CNV-Analyse in bestimmten Signalwegen oder eine Beteiligung dieser Gene an Signalwegen bekannter CRC-Gene aufzudecken. Hierfür wurden alle 76 CNV-Gene zusammen mit 38 relevanten CRC-Genen analysiert. Insgesamt wurden 38 signifikante (p-Werte < 0,05) kanonische Signalwege, in denen insgesamt 13 CNV-Gene (18%) und 29 CRC-Gene (76%) gemeinsam involviert sind, ermittelt (siehe Tab. H.1 im Anhang). Von allen detektierten Signalwegen wurden 29 als relevant für die Entwicklung von Malignomen eingeschätzt (76%) (Tab. 27). Unter diesen 29 relevanten Signalwegen sind sieben der 76 CNV-Gene beteiligt (*DLG2*, *DUSP11*, *GNB2L1*, *MCM4*, *NRG3*, *PRKCA*, *PRKDC*).

Tab. 27 Relevante Signalwege und betroffene Gene, Ingenuity Pathway Analyse

kanonische Signalwege	Gene	Anzahl CRC-Gene (+ CNV-Gene)
Colorectal Cancer Metastasis Signaling	<i>TP53, MAP2K4, PIK3C2B, SMAD2, AXIN1, GNB2L1, MLH1, APC, CDH1, AKT1, MSH2, MSH6, SMAD4, GSK3B, TCF7L2</i>	14 (+1)
Molecular Mechanisms of Cancer	<i>TP53, MAP2K4, PIK3C2B, PRKDC, SMAD2, AXIN1, PRKCZ, APC, CDH1, AKT1, BMPR1A, SMAD4, GSK3B, CHEK2, PRKCA</i>	13 (+2)
ErbB Signaling	<i>MAP2K4, PIK3C2B, MTOR, AKT1, NRG3, GSK3B, PRKCZ, PRKCA</i>	6 (+2)
Glioma Signaling	<i>TP53, PIK3C2B, MTOR, AKT1, PDGFRA, PRKCZ, PRKCA, PTEN</i>	7 (+1)
14-3-3-mediated Signaling	<i>MAP2K4, TSC1, PIK3C2B, AKT1, TSC2, GSK3B, PRKCZ, PRKCA</i>	7 (+1)
mTOR Signaling	<i>TSC1, PIK3C2B, MTOR, AKT1, PPP2CA, TSC2, PPP2R1B, PRKCZ, PRKCA</i>	8 (+1)
p53 Signaling	<i>TP53, PIK3C2B, PRKDC, AKT1, GSK3B, CHEK2, PTEN</i>	6 (+1)
p70S6K Signaling	<i>PIK3C2B, MTOR, AKT1, PPP2CA, PPP2R1B, PRKCZ, PRKCA</i>	6 (+1)
HIPPO Signaling	<i>SMAD2, PPP2CA, DLG2, SMAD4, PPP2R1B, PRKCZ</i>	5 (+1)
EGF Signaling	<i>MAP2K4, PIK3C2B, MTOR, AKT1, PRKCA</i>	4 (+1)
ErbB2-ErbB3 Signaling	<i>PIK3C2B, AKT1, NRG3, GSK3B, PTEN</i>	4 (+1)
Gαq Signaling	<i>PIK3C2B, AKT1, GNB2L1, GSK3B, PRKCZ, PRKCA</i>	4 (+2)
CXCR4 Signaling	<i>MAP2K4, PIK3C2B, AKT1, GNB2L1, PRKCZ, PRKCA</i>	4 (+2)
ErbB4 Signaling	<i>PIK3C2B, AKT1, NRG3, PRKCZ, PRKCA</i>	3 (+2)
Role of Tissue Factor in Cancer	<i>TP53, PIK3C2B, MTOR, AKT1, PRKCA, PTEN</i>	5 (+1)
Protein Kinase A Signaling	<i>PTPRJ, GNB2L1, SMAD4, DUSP11, GSK3B, TCF7L2, PRKCZ, PRKCA, PTEN</i>	7 (+3)
P2Y Purigenic Receptor Signaling Pathway	<i>PIK3C2B, AKT1, GNB2L1, PRKCZ, PRKCA</i>	4 (+2)

Gβγ Signaling	<i>AKT1, GNB2L1, PRKCZ, PRKCA</i>	2 (+2)
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	<i>TP53, PRKDC, PRKCZ, CHEK2</i>	3 (+1)
HGF Signaling	<i>MAP2K4, PIK3C2B, AKT1, PRKCZ, PRKCA</i>	4 (+1)
PPARα/RXRα Activation	<i>MAP2K4, SMAD2, SMAD4, PRKCA</i>	4 (+1)
PDGF Signaling	<i>MAP2K4, PIK3C2B, PDGFRA, PRKCA</i>	3 (+1)
Signaling by Rho Family GTPases	<i>MAP2K4, PIK3C2B, CDH1, GNB2L1, PRKCZ</i>	4 (+1)
Wnt/Ca+ pathway	<i>AXIN1, GSK3B, PRKCA</i>	2 (+1)
Cell Cycle Control of Chromosomal Replication	<i>MCM4, CHEK2</i>	1 (+1)
Apoptosis Signaling	<i>TP53, MAP2K4, PRKCA</i>	2 (+1)
VEGF Signaling	<i>PIK3C2B, AKT1, PRKCA</i>	2 (+1)
RhoGDI Signaling	<i>CDH1, GNB2L1, PRKCA</i>	1 (+2)
ERK/MAPK Signaling	<i>PIK3C2B, PPP2CA, PPP2R1B, PRKCA</i>	3 (+1)
Hervorgehobene Gene: CNV-Gene		

4.3.8 Segregationsanalyse

Bei vorhandener Segregation treten CNV und Phänotyp innerhalb der betroffenen Familien gemeinsam auf, sie segregieren. Die Stärke der Segregation hängt insbesondere von der Zahl untersuchter Patienten bzw. gesunder Kontrollpersonen ab und ist ein wichtiger Hinweis zur kausalen Bedeutung einer Keimbahnvariante für den vorliegenden Phänotyp. Es wurden alle mutmaßlichen LoF-Gene, wie auch potentielle, von einer kompletten Duplikation betroffene Onkogene berücksichtigt. Mit eingeschlossen wurden darüber hinaus solche CNV-Gene, die durch die Netzwerk- und Pathway-Analyse besonders relevant erschienen. Dies betraf 23 CNV-Gene, welche in der Keimbahn von 14 Indexpatienten mutiert vorliegen. Tab. 28 listet die Index-Patienten mit allen CNVs und den darin betroffenen Genen auf.

Tab. 28 Index-Patienten für die Segregationsanalyse

#	Patienten-ID	FA	Gen	Art der Mutation
1	HNPCC_01	Mutter† CRC 50 J., 70 J	DDX24	Dup
		Mutter BRCA 79 J	ENOX1	Dup ^{part}
		Großvater CRC 42 J		
2	HNPCC_24	Onkel Prostata-Ca 72 J	FBX036	Del
		Cousine 2x CRC		
		Cousin† CRC 30 J		
			DLG2	Del ^{intr}
3	HNPCC_28	Schwester EC 50 J.	PFKP	Dup
			PITRM1	Dup
			SLCO4C1	Dup ^{part}
4	HNPCC_31	Onkel mit CRC 37 J.	SRBD1	Del ^{intr}
			ABCB1	Del ^{intr}
5	HNPCC_34	unbekannt	RUNDC3B	Del ^{part}
			GNB2L1	Dup
			NRG3	Del ^{intr}
			RAP1GDS1	Del ^{part}
6	HNPCC_38	Muttter 81 J. EC-Ca	RBM19	Dup ^{part}
			TRIM41	Dup ^{part}
			TRIM52	Dup
			TSPAN5	Del ^{part}
		Schwester EC 30 J, CRC 60J	PDZRN4	Dup ^{intr}
7	HNPCC_41	Mutter Leberkrebs 76 J.	MCM4	Del ^{part}
		Vater Leukämie 37 J.	PRKDC	Del ^{part}
8	HNPCC_42	Vater† CRC 55 J.	PRKCA	Del ^{intr}
			DUSP11	Dup ^{part}
9	HNPCC_45	Mutter „Ca“ 45 J.	NPS	Del
		Großvater CRC 83 J	TPRKB	Dup
10	HNPCC_47	Onkel „Malignome“	PAG1	Del ^{part}
		Cousin CRC, Magen-Ca		
11	HNPCC_72	Vater Gehirntumor 62 J.**	DPP10	Dup ^{intr}
			B3GNT5	Dup
			DCUN1D1	Dup ^{part}
			KLHL24	Dup
12	HNPCC_73	Mutter CRC 67 J.	KLHL6	Dup
		Vater CRC 62 J., CRC 77 J.**	LAMP3	Dup
			MCCC1	Dup
			MCF2L2	Dup
			YEATS2	Dup ^{part}
			ZNF154	Del
			ZNF551	Del
13	HNPCC_79	Vater CRC 32 J.**	ZNF671	Del
			ZNF776	Del ^{part}

14	HNPCC_97	Schwester Ovarial-Ca 63 J.**	ATP12A	Dup ^{part}
			CHST2	Dup
			SLC9A9	Dup

BRCA: Brust-und Eierstockkrebs; CA: Karzinom; CRC: Kolorektales Karzinom, Del: Deletion; Dup: Duplikation; EC: Endometriumkarzinom, FA: Familienanamnese; † bereits verstorben, hervorgehoben: relevante CNV-Gene, die für die Segregationsanalyse berücksichtigt wurden, intr: CNV betrifft nur intronischen Bereich, part: kodierender Gen-Bereich partiell vom CNV betroffen, *Patient hinterher im NGS als ursächlich aufgeklärt (pathogene *MSH6*-Mutation), **Patienten des HNPCC-Konsoriums, keine Kontaktaufnahme mit Angehörigen möglich

Dreizehn der 14 Index-Patienten wiesen eine positive Familienanamnese auf. Allerdings konnten nur bei sechs der Familien betroffene, noch lebende Angehörige ausfindig gemacht werden. Lediglich von einer der sechs kontaktierten Familien erhielten wir eine Rückmeldung eines weiteren erkrankten Angehörigen. Daraufhin wurde die Blutprobe der metachron an Darmkrebs erkrankten Cousine der Index-Patientin angefordert. Der Stammbaum dieser Familie ist in Abb. 21 dargestellt. Die Index-Patientin (HNPCC_24) weist eine partielle Duplikation in *ENOX1* auf.

Die Kopienzahl wurde mittels einer qPCR überprüft. Hierbei stellte sich heraus, dass die Cousine eine normale Kopienzahl im *ENOX1*-Gen aufwies. Eine Co-Segregation dieser partiellen Duplikation mit dem Phänotyp konnte hier somit nicht festgestellt werden. *ENOX1* ist somit mit hoher Wahrscheinlichkeit kein hochpenetrantes Kandidatengen für den vorliegenden Phänotyp.

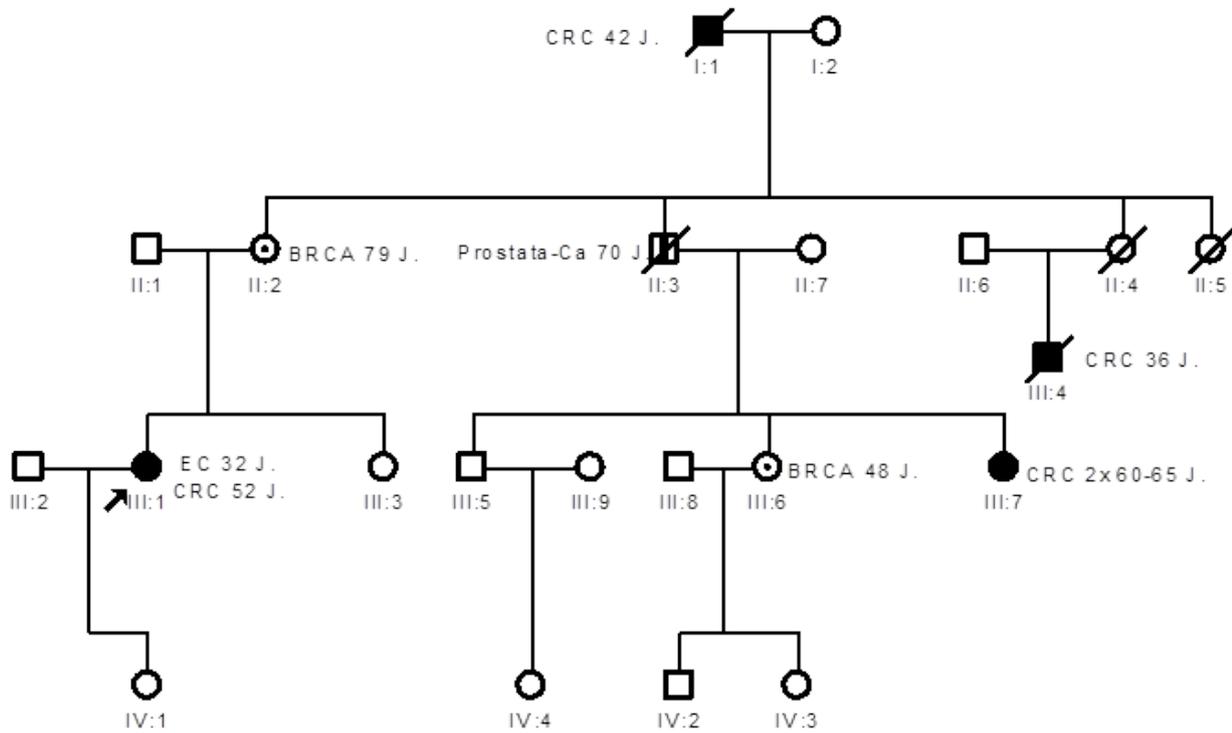


Abb. 21 Stammbaum von HNPCC_24 ENOX1 (partielle Duplikation)

Die Indexpatientin trägt den Pfeil. Untersucht wurde die betroffene Cousine (III:7). BRCA: Brust- und Eierstockkrebs; Ca: Karzinom; CRC: Kolorektales Karzinom; EC: Endometriumkarzinom

4.4 Next-Generation-Sequencing

Für die Suche nach putativ pathogenen Punktmutationen (CADD-Score ≥ 20) in den CNV-Kandidatengenomen dieser Studie sowie bekannten CRC-Genen wurden alle Patienten des Kollektivs, von denen genügend DNA in ausreichender Qualität zur Verfügung stand ($n = 44$), mit einem CustomTarget NGS-Panel von Illumina sequenziert. Das Target-Panel bestand aus 106 Genen: 68 Gene der CNV-Analyse, 30 CRC-Gene und 8 Gene des Proteasom-Signalweges (siehe Anhang Tab. G.1 „Genliste NGS“). Aufgrund von limitierten finanziellen Mitteln konnten nicht alle Gene berücksichtigt werden, daher wurde eine Auswahl getroffen. CNV-Gene, die funktionell und aufgrund der Voruntersuchungen keinerlei Hinweis auf eine Kausalität zur Erkrankung lieferten, wurden ausgeschlossen, ebenso bekannte CRC-Gene, die einen bekanntermaßen unterschiedlichen Phänotyp hervorrufen.

Hierbei wurden insgesamt 12.956 Varianten bei den 44 Patienten detektiert. Nach dem Herausfiltern potentieller Artefakte durch Anwendung mehrerer Filterschritte ($Q > 30$, Read Depth > 10 , Alt Variant Freq > 10) wurden für die Suche nach dominanten Mutationen häufige Varianten mit einer Allelfrequenz $> 0,01\%$ (TGP, EVS und ExAc) ausgeschlossen. Hierbei blieben 112 seltene Varianten in 63 der 106 Gene übrig. Für die Suche nach rezessiven Varianten wurde der Allelfrequenz-Filter von $< 0,01\%$ (sehr seltene Varianten) auf $< 1\%$ hochgesetzt, da rezessive Varianten in heterozygotem Zustand weitaus häufiger in der Allgemeinbevölkerung auftreten. Mit dieser Allelfrequenz von $< 1\%$ wurde nach zwei Varianten pro Gen (homozygot oder potentiell compound heterozygot) desselben Patienten gesucht. Es konnten dadurch keine rezessiven Varianten detektiert werden.

Zur Beurteilung der negativen Auswirkungen (*deleteriousness*) der Varianten wurden „skalierte C-Scores / PHRED-like scores“ verwendet (vgl. Kap. 3.10.3 Varianten-Filterung, S.49). Bewertet wurden alle Varianten mit einem CADD-Score ≥ 20 , was für eine beeinträchtigende Auswirkung der Variante auf die Proteinfunktion spricht.

4.4.1 Bekannte CRC-Gene der Literatur

Das Custom-Panel umfasste unter anderem die für das LS verantwortlichen MMR-Gene. In den MMR-Genen *MLH1*, *MSH2* und *MSH6* wurden sechs bisher nicht detektierte heterozygote, pathogene Mutationen gefunden (Tab. 29). Die Patienten wurden daraufhin nachträglich von

Ergebnisse

der CNV-Studie ausgeschlossen. In den MMR-Genen wurden zudem Missense-Varianten unklarer Signifikanz (VUS) nachgewiesen (Tab. 30).

Tab. 29 Pathogene Punktmutationen der MMR-Gene

Gen	Patienten-ID	cDNA-Änderung	Protein Änderung	Mutations- typ	InSiGHT Klasse	CADD- Score
<i>MLH1</i>	HNPCC_27	c.62C>A	p.Ala21Glu	Missense	5	35
<i>MSH2</i>	HNPCC_35	c.943-2A>G	aberrantes Spleißen	Intronisch	4	25
<i>MSH2</i>	HNPCC_20	c.1147C>T	p.Arg383Ter	Stopp	5	40
<i>MSH6</i>	HNPCC_02	c.650_651insT	p.Lys218Ter	Frameshift	5	26
<i>MSH6</i>	HNPCC_07	c.2401_2042delCT	p.Ser682CysfsTer15	Frameshift	5	34
<i>MSH6</i>	HNPCC_62	c.3227delG	p.Arg107Profs*3	Frameshift	5	34

Tab. 30 MMR-Gen-Varianten unklarer funktioneller Signifikanz (InSiGHT Klasse 3)

Gen	Patienten-ID	cDNA-Änderung	Protein Änderung	Mutation	CADD- Score
<i>MSH2</i>	HNPCC_60	c.818T>A	p.Val273Glu	Missense	32
<i>MSH2</i>	HNPCC_22	c.1864C>A	p.Pro622Thr	Missense	28
<i>MSH6</i>	HNPCC_46	c.722G>T	p.Ser24Ile	Missense	24
<i>MSH6</i>	HNPCC_65	c.1372C>T	p.His458Tyr	Missense	26

Weiterhin wurde eine klar pathogene Stopp-Mutation im *APC*-Gen, welches für die familiäre adenomatöse Polyposis verantwortlich ist ausfindig gemacht (HNPCC_63: c.6344T>G;p.Leu2115Ter). Der Patient wurde ebenfalls von der Studie ausgeschlossen. Zusätzlich wurden im *APC*-Gen VUS bei zwei Patienten (HNPCC_22 und HNPCC_24) detektiert (Tab. 31).

In weiteren, für erbliche Darmkrebsformen verantwortlichen Genen, wurden Varianten von unklarer klinischer Signifikanz (VUS) entdeckt. In *CHEK2* wurde eine Missense-Mutation und in *MAP2K4* eine intronische Variante detektiert (Tab. 31). Da diese Mutationen derzeit nicht sicher eingeordnet werden können, wurden diese Patienten nicht von der Studie ausgeschlossen.

Tab. 31 Punktmutationen in bekannten Genen für erbliche Darmkrebsformen

Gen	Patienten-ID	Variante	vermutete Konsequenz	Mutations- typ	CADD- Score	HI- Score; RVIS %
<i>APC</i>	HNPCC_63	c.6344T>G	p.Leu2115Ter	Stopp	40	
<i>APC</i>	HNPCC_22	c.1703G>A	p.Ser568Asn	Missense	28	17; 0,2
<i>APC</i>	HNPCC_24	c.8425G>A	p.Val2809Met	Missense	25	
<i>CHEK2</i>	HNPCC_47	c.1441G>T	p.Asp481Tyr	Missense	33	1; 70
<i>MAP2K4</i>	HNPCC_25	c.-3A>G	aberrantes Spleißen	Intronisch	20	21; 26

4.4.2 Gene des Proteasom-Signalwegs

Unter den Genen des Proteasom-Signalwegs konnten sechs seltene, heterozygote Varianten in vier der acht Gene identifiziert werden (Tab. 32). Diese Gene stehen in Verdacht, bei einem Funktionsverlust des kodierten Proteins eine Downregulation von MSH2 zu bewirken (siehe Kap. 2.4.5). *HERC1* wies hierbei drei heterozygote Varianten in drei verschiedenen Patienten auf; darunter zwei Varianten mit CADD-Scores von 35, was für eine relative Pathogenität der Varianten spricht. In *AKT1*, *TSC1* und *TSC2* wurde jeweils eine Variante entdeckt.

Tab. 32 Punktmutationen in Genen des Proteasom-Signalwegs

Gen	Patienten-ID	Variante	vermutete Konsequenz	Mutationstyp	CADD- Score	HI-Score; RVIS %
<i>AKT1</i>	HNPCC_06	c.361C>T	p.Arg121Trp	Missense	26	0,4; 8
	HNPCC_10	c.9508C>T	p.Arg3170Cys	Missense	35	
<i>HERC1</i>	HNPCC_43	c.10660C>T	p.Arg3554Trp	Missense	35	45; 0,05
	HNPCC_65	c.14297A>G	p.Asn4766Ser	Missense	24	
<i>TSC1</i>	HNPCC_23	c.532G>A	p.V1178Ile	Missense	24	6; 19
<i>TSC2</i>	HNPCC_22	c.4545C>G	p.Asn1515Lys	Missense	24	4; 1

4.4.3 CNV-Gene

Unter den 68 auf dem Panel befindlichen CNV-Genen wurden 15 seltene heterozygote Varianten mit einem CADD-Score ≥ 20 ermittelt (Tab. 33).

Die 15 Punktmutationen bestehen zum überwiegenden Teil (75%) aus putativen *Missense*-Varianten. Die übrigen Varianten verteilen sich auf zwei intronische Varianten außerhalb der hochkonservierten Spleiß-Konsensus-Regionen eine synonyme Variante und zwei *Inframe*-

Deletionen. Die Signifikanz dieser Varianten ist unsicher. Sicher pathogene trunkierende Mutationen wurden nicht gefunden.

Die meisten Varianten (4 Missense-Varianten) wurden in *GRIK4* gefunden. In *TRIM41* wurden zwei Missense-Varianten entdeckt. Alle weiteren Gene waren von je einer Variante betroffen. Eine der Missense-Varianten in *TRIM41* wurde bei demselben Patienten (HNPCC_38) gefunden, welcher auch eine partielle Duplikation im *TRIM41*-Gen aufwies. Die Variante liegt in dem Bereich des Gens, der von der Duplikation betroffen ist. Dies könnte auf eine rezessive Form der Vererbung hindeuten; die Compound-Heterozygotie lies sich allerdings nicht prüfen. Weitere Hinweise auf eine rezessive Vererbung (zwei pathogene Punktmutationen im selben Patienten mit einer Allelfrequenz < 1%) wurden nicht detektiert.

Tab. 33 Punktmutationen in CNV-Genen

Gene	Patienten-ID NGS	Variante	vermutete Konsequenz	Mutationstyp	Patienten-ID CNV	CNV	CADD- Score
<i>GRIK4</i>	HNPCC_08	c.514C>T	p.Leu172Phe	Missense	HNPCC_94	Del ^{part}	29
	HNPCC_36	c.569C>T	p.Ser190Phe	Missense			24
	HNPCC_39	c.2219T>C	p.Ile740Thr	Missense			26
	HNPCC_46	c.896C>G	p.Thr299Ser	Missense			20
<i>KATNAL1</i>	HNPCC_22	c.1082C>T	p.Pro361Leu	Missense	HNPCC_90	Dup ^{part}	29
<i>LAMP3</i>	HNPCC_52	c.607C>T	p.Arg203Cys	Missense	HNPCC_73	Dup	23
<i>MTHFD1L</i>	HNPCC_55	c.2924T>C	p.Val975Ala	Missense	HNPCC_50	Dup ^{part}	26
<i>NOSIP</i>	HNPCC_44	c.277G>A	p.Gly93Ser	Missense	HNPCC_18	Del	23
<i>OTUB2</i>	HNPCC_38	c.146G>A	p.Gly49Glu	Missense	HNPCC_01	Dup	28
<i>PRKCA</i>	HNPCC_45	c.1658_1660delACA	p.Asn554del	Inframe-Deletion	HNPCC_42	Del ^{intr}	23
<i>PRKDC</i>	HNPCC_31	c.8811G>A	p.Thr2937	Synonym	HNPCC_41	Del ^{part}	22
<i>PRRG2</i>	HNPCC_55	c.254C>A	p.Thr85Asn	Missense	HNPCC_18	Del ^{part}	22
<i>TRIM41</i>	HNPCC_38	c.1084C>T	p.Arg362Cys	Missense	HNPCC_38	Dup ^{part}	24
	HNPCC_65	c.1561G>T	p.Gly521Cys	Missense			24
<i>VWDE</i>	HNPCC_47	c.4430G>A	p.Arg1477His	Missense	HNPCC_31	Del	23

Del: Deletion, Dup: Duplikation; intr. CNV betrifft nur intronischen Bereich; part: kodierender Gen-Bereich partiell vom CNV betroffen

hervorgehobene Gene: CNV-Kandidatengene mit zusätzlichen Varianten im Gen

4.5 Zusammenfassung der Ergebnisse aus CNV- und NGS-Analyse

Unter Anwendung stringenter Filter auf CNV- und Gen-Ebene und unter Berücksichtigung aller nachfolgenden Analysen wurden die zehn spannendsten Kandidatengene ermittelt (*DLG2*, *DUSP11*, *GNB2L1*, *MCM4*, *PAG1*, *PRKCA*, *PRKDC*, *RUNDC3B*, *TRIM41* und *TRIM52*) (Tab. 34). Für die Einschätzung berücksichtigt wurden dabei: die Genfunktion, der Hi- und RVI-Score, die Ergebnisse der Netzwerk-, Pathway-, Segregations- und NGS-Analyse, der Vergleich mit den Ergebnissen vorangegangener CNV-Studien zu ungeklärten Tumorsyndromen (Tab. 35) und das Vorliegen einer Expression in der Darmschleimhaut sowie das Vorliegen als somatische Mutationen $\geq 3\%$ in 295 nicht-hypermutierten Adenokarzinomen der TCGA-Datenbank.

Die selektierten Gene sind funktionell relevant; sieben der Gene sind potentielle TSG, drei potentielle Onkogene. Dabei stimmt die Art des CNVs (Del/Dup, partiell/komplett) mit dem daraus abzuleitenden Effekt auf die Genfunktion (*LoF/gain of Function*) überein (Kriterium I, vgl. Kap. 4.3.4). Alle sieben potentiellen LoF-Gene weisen zudem einen niedrigen Hi-/RVI-Score auf, sie sind folglich dosissensitiv/variantenintolerant gegenüber Mutationen. Acht der Gene (*GNB2L1*, *MCM4*, *PAG1*, *PRKCA*, *PRKDC*, *RUNDC3B*, *TRIM41*, *TRIM52*) sind in einem gemeinsamen Netzwerk mit *MSH2* beschrieben. Sechs der Gene (*DLG2*, *DUSP11*, *GNB2L1*, *MCM4*, *PRKCA*, *PRKDC*) sind in einem Pathway beschrieben, der als relevant für die Entwicklung von Malignomen eingeschätzt wurde und an dem weitere bekannte CRC-Gene beteiligt sind. In drei der Gene (*PRKCA*, *PRKDC*, *TRIM41*) wurden zusätzliche Keimbahnvarianten identifiziert. Für fünf Gene (*DUSP11*, *MCM4*, *PRKCA*, *TRIM41* und *TRIM52*) wurde in nah verwandten Genen anderer CNV-Studien bei Patienten mit ungeklärtem familiärem CRC bereits Kopienzahlveränderungen identifiziert. Die berücksichtigten Studien sind in Tab. 35 aufgelistet. Keines der Gene wies somatische Mutationen $\geq 3\%$ in nicht-hypermutierten Adenokarzinomen der TCGA-Datenbank auf.

Eine komplette Auflistung aller CNV-Gene befindet sich im Anhang in Tab. E.1.

Tab. 34 Zusammenfassung der Ergebnisse aller Analysen (CNV und NGS)

Gen	ID	Art des CNVs	LoF-/Onkogen	Hi-/ RVI-Score	Kriterium	Netzwerk-Analyse	Pathway-Analyse	NGS	Literaturvergleich	Expression in Darmschleimhaut
DLG2	28	Del ^{intr}	LoF-Gen	1; 4	I und II	-	x	-	-	x
DUSP11	45	Dup ^{part}	LoF-Gen	53; 18	I und II	-	x	-	Masson et al., 2013 (<i>DUSP8</i> , Dup)	x
GNB2L1	38	Dup	Onkogen	(60); (25)	I und -	x	x	-	-	x
MCM4	41	Del ^{part}	LoF-Gen	2; 4	I und II	x	x	-	Horpaopan et al., 2015 (<i>MCM3AP</i> , Dup ^{part})	x
PAG1	47	Del ^{part}	LoF-Gen	35; 39	I und II	x	-	-	-	x
PRKCA	42	Del ^{intr}	LoF-Gen	5; 9	I und II	x	x	x	Masson et al., 2013 (<i>PRKCI</i> , Dup ^{part})	x
PRKDC	41	Del ^{part}	LoF-Gen	9; -	I und II	x	x	x	-	x
RUNDC3B	34	Del ^{part}	LoF-Gen	66; 19	I und II	x	-	-	-	x
TRIM41	38	Dup ^{part}	Onkogen	83; 62	I und -	x	-	x	Masson et al., 2013 (<i>TRIM69</i> , 2x Dup ^{part})	x
TRIM52	38	Dup	Onkogen	(88); (45)	I und -	x	-	-	Masson et al., 2013 (<i>TRIM69</i> , 2x Dup ^{part})	x

Del: Deletion; Dup: Duplikation; intr: CNV betrifft nur intronischen Bereich;

part: kodierender Gen-Bereich partiell vom CNV betroffen

Kriterium I: Funktionell relevant in Bezug zur Art der Mutation;

Kriterium II: Gene mit einem Hi-Score ≤ 35 oder RVI ≤ 25

x: Proteinprodukt dieses Gens beteiligt an Netzwerk- / Pathwayanalyse; Variante im NGS detektiert; exprimiert in Darmschleimhaut

Werte in Klammern: Komplette Gen-Duplikation, Scores nicht vollständig anwendbar

Tab. 35 Berücksichtigte CNV-Studien zu familiären Tumorsyndromen

Krebskohorte (familiär / früh- manifestiert)	# Pat.	# Kontr.	# CNVs	Keimbahn CNVs (Details der Studie)	Referenzen
Pankreas-Ca	57	607	56	nichtrekurrente, seltene CNVs (nicht in Kontrollpersonen)	(Lucito et al., 2007)
MSS CRC ohne Polyposis	41	1600	7	nichtrekurrente, seltene CNVs (in keinen Kontrollen und DGV)	(Venkatachalam et al., 2011)
Brust- und Eierstockkrebs	68	258	26	nichtrekurrente, seltene CNVs (in keinen Kontrollen und DGV)	(Krepischi et al., 2012)
CRC	371	1262	1	Seltene 12p12.3 Deletion in 2 Patienten und keinen Kontrollen (DGV)	(Yang et al., 2014)
Pankreas-Ca	223	169	2	nichtrekurrente, seltene CNVs (in keinen Kontrollen DGV)	(Willis et al., 2014)
Melanom	62	17	1	nichtrekurrente, seltene Duplikation auf 4q13 in 3 Patienten einer Familie und keinen Kontrollen	(Yang et al., 2012)
HNPCC* und LS- Patienten	35 und 65	384	1	seltene Duplikation auf 7q11.21 (in 28% der Patienten und in keinem der Kontrollen)	(Talseth-Palmer et al., 2013)
HNPCC*	125	40	207	rekurrente seltene CNVs (nicht in Kontrollen)	(Masson et al., 2013)
Polyposis	221	531	125	nichtrekurrente, seltene CNVs (nicht in Kontrollen und DGV)	(Horpaopan et al., 2015)
Brust- und Eierstockkrebs	81	36	35	nichtrekurrente, seltene CNVs (in keinen Kontrollen und DGV)	(Kuusisto et al., 2013)
HNPCC*	45	100	35	nichtrekurrente, seltene CNVs (in keinen Kontrollen, DGV, Affymetrix Database of Variants)	(Villacis et al., 2015)

* Erfüllte Amsterdam/Bethesda-Kriterien, keine Mutation in einem der MMR-Gene

CRC: Kolorektales Karzinom; Ca: Karzinom; LS: Lynch Syndrom

hervorgehobene Studien: im Vergleich mit unserer Studie übereinstimmendes Gen von CNV betroffen

5 Diskussion

Bei 30 bis 50% aller Patienten, welche die klinischen Kriterien für das LS erfüllen und die typischen Merkmale im Tumor aufweisen (Ausfall eines Mismatch-Repair-Proteins, hohe Mikrosatelliteninstabilität), lässt sich keine ursächliche pathogene Keimbahnmutation nachweisen. Aufgrund des Auftretens typischer, frühmanifestierender HNPCC-assoziiierter Tumoren (meist kolorektale Karzinome oder Endometriumkarzinome im Alter < 50 Jahre) oder dem zusätzlich gehäuftem Auftreten dieser Tumoren innerhalb der Familien, ist jedoch eine erbliche genetische Komponente zu vermuten. Dieser fehlende genetische Faktor könnte auf eine noch unbekanntes monogene, hochpenetrante Mutation zurückzuführen sein, oder auf mehreren milder penetranten Varianten beruhen, welche in Form eines komplexen Zusammenspiels die Erkrankung verursachen. Bei einer Reihe erblicher Tumorsyndrome konnten seltene CNVs, insbesondere heterozygote Deletionen, als Teil des Mutationsspektrums nachgewiesen werden (Krepischi et al., 2012; Manolio et al., 2009). Weitere Mikrodeletionen oder -duplikationen könnten demnach für die noch ungeklärten Fälle beim HNPCC verantwortlich sein. Um nach neuen hochpenetranten monogenen Varianten zu suchen, führten wir eine genomweite CNV-Analyse in einem ausgewählten HNPCC-Kollektiv mit MSH2-Verlust durch.

5.1 Ursächlich aufgeklärte Patienten des Kollektivs

Durch die gründliche Aufarbeitung des Kollektivs konnte für insgesamt 19 (20%) der ursprünglich 95 HNPCC-Patienten nachträglich eine pathogene Mutation in einem bekannten, etablierten Gen nachgewiesen werden. Für 17 der 19 Patienten wurde hierbei die Diagnose eines LS gestellt. Ein Patient konnte einer FAP, ein anderer einer MAP zugeteilt werden.

Vier der 19 Patienten erfuhren eine Hochstufung von MMR-Gen-Varianten der Klasse 3 (unklare klinische Signifikanz) zu Varianten der Klasse 4 (wahrscheinlich pathogen) oder Klasse 5 (erwiesenermaßen pathogen). Da sich die Einstufungen des InSiGHT-VIC je nach Datenlage ändern können, sind Neuklassifizierungen der VUS möglich.

Bei 13 der 19 Patienten konnten vorher nicht detektierte pathogene Keimbahnmutationen in MMR-Genen identifiziert werden. Hierunter befanden sich sieben Patienten, bei denen zuvor nur die DHPLC-Methode zum Mutations-Screening eingesetzt worden war. Heute ist bekannt,

dass durch die Anwendung dieser Methode häufig Mutationen übersehen werden (Medlock et al., 2012). Durch eine im Rahmen dieser Studie ergänzende Sanger-Sequenzierung des kompletten *MSH2*- und *MSH6*-Gens wurden bei diesen Patienten vier Keimbahnmutationen in *MSH2* und drei Keimbahnmutationen in *MSH6* identifiziert (siehe Kap. 4.1.1). Bei den übrigen sechs der 13 Patienten wurden pathogene MMR-Gen-Mutation über das NGS-Panel detektiert. Darunter befanden sich zwei *MSH2*-Mutationen, die über die vorausgegangene Sanger-Sequenzierung übersehen worden waren, sowie weiterhin drei *MSH6*- und eine *MLH1*-Mutation (Kap. 4.4.1). Der Nachweis der *MSH6*-Mutationen und der *MLH1*-Mutation war überraschend, da aufgrund des immunohistochemischen Befundes (*MSH2*-Ausfall) nur *MSH2*-Mutationen zu erwarten waren und daher nur nach diesen gesucht worden war. Die Immunohistochemie ist eine in der Diagnostik etablierte Methode, welche die Suche nach pathogenen Keimbahnmutationen erleichtert. Zum einen erspart man sich so die Kosten für die Sequenzierung aller vier MMR-Gene, zum anderen vereinfacht sie auch die Interpretation der funktionellen Relevanz einer detektierten Variante. Sensitivität und Spezifität der Methode liegen allerdings nicht bei 100%, sodass falsch positive und negative Befunde vorkommen (Steinke et al., 2010). Bei ursächlichen *MSH2*-Mutationen fällt immer auch *MSH6* im Tumorgewebe mit aus, da *MSH2* der einzige Bindungspartner von *MSH6* ist. Bei Mutationen im *MSH6*-Gen hingegen, bleibt die *MSH2*-Expression üblicherweise erhalten, da *MSH2* noch weitere Bindungen mit *MSH3* eingehen kann (vgl. Kapitel 2.3.6). Da dies in der Praxis jedoch nicht immer zutrifft ist man inzwischen in der Diagnostik dazu übergegangen auch bei alleinigem Ausfall von *MSH2* (bzw. *MLH1*), den Komplexbpartner *MSH6* (bzw. *PMS2*) mitzusequenzieren.

Bei zwei Patienten konnte molekulargenetisch eine MAP bzw. FAP diagnostiziert werden. Der Verdacht einer MAP ergab sich durch erneute Aktensichtung. Der Patient erkrankte mit 40 Jahren an Dickdarmkrebs. Koloskopisch ließen sich 23 kolorektale Adenome nachweisen. Zudem war sein Bruder im Alter von 42 Jahren an Dickdarmkrebs erkrankt. Die anschließende Sanger-Sequenzierung offenbarte zwei wahrscheinlich biallele Keimbahnmutationen im *MUTYH*-Gen. Bei dem weiteren Patienten wurde eine pathogene *APC*-Mutation detektiert. Dies war ein Zufallsbefund der über das NGS-Panel detektiert wurde. Die Familien- und Eigengeschichte des Patienten sowie der Tumorbefund lieferten keinen Hinweis auf eine FAP. Der Patient wurde infolge eines Dünndarmkarzinoms mit einer geringen Anzahl kolorektaler Adenome zusammen mit den typischen Veränderungen im Tumorgewebe (hohe Mikro-

satelliteninstabilität und Ausfall von MSH2) phänotypisch als HNPCC-Patient klassifiziert. Zudem war seine Mutter mit 49 Jahren an einem Magenkarzinom, einem zum HNPCC-Spektrum gehörenden Tumor, erkrankt. Aufgrund der geringen Polypenzahl und der MSI-H wurde klinisch keine FAP vermutet. Bei der klassischen FAP können bis zu tausende Polypen im gesamten Darm auftreten. In Bezug auf Anzahl, Größe und Verteilung der Polypen gibt es bei der FAP allerdings eine weite interfamiliäre und interindividuelle Variabilität (Giardiello et al., 1994). Generell rufen Mutationen am 3'-Ende des *APC*-Gens, wie hier vorliegend, ein attenuiertes Krankheitsbild mit wenigen Polypen hervor. Die Polypen sind in der Regel mikrosatellitenstabil. Das Auftreten von MSI bei FAP-Patienten ist bislang nicht näher untersucht. Es könnte sich um zufällige somatische Mutationen im Tumor mit dadurch bedingter MSI gehandelt haben. Die weitere Krankheitsgeschichte des Patienten verlief untypisch: Mit dreieinhalb Jahren erkrankte er an einer akuten lymphatischen Leukämie, mit 21 Jahren an einem Akustikusneurinom und mit 29 Jahren an einem Zungengrundkarzinom. Diese Tumoren gehören nicht zu dem Spektrum der HNPCC- oder FAP-Tumoren. Die Ursachen für die extraintestinalen Krebserkrankungen des Patienten sind durch die *APC*-Mutation nicht zu erklären. Hierfür kann das Vorliegen weiterer genetischer Risikofaktoren nicht ausgeschlossen werden. Der Patient wies keine seltenen CNVs und keine weiteren pathogenen Punktmutationen in den untersuchten Genen auf.

Die Daten demonstrieren, dass der etablierte diagnostische Algorithmus (Einteilung nach Familiengeschichte und Eigenanamnese als auch Tumorbefund) nicht vollständig sensitiv ist. Eine Umstellung auf NGS-Panel, mit denen eine Vielzahl kausaler Gene synchron untersucht werden können, könnte dieses Problem in Zukunft lösen. Vorteil dieser Methode ist die schnelle, kostengünstige Sequenzierung mehrerer Gene gleichzeitig unter Einsatz einer nur geringen DNA-Menge. Durch die hohe Abdeckung (*coverage*) könnte die Detektionsrate beim NGS höher sein als bei der Sanger-Sequenzierung. In der vorliegenden Arbeit war die Detektion zweier pathogener *MSH2*-Mutationen erst mittels NGS möglich. Andererseits ist bekannt, dass homopolymere Bereiche, wie sie auch im *MSH2*-Gen vorkommen, über das NGS nicht sicher abgedeckt werden (Simbolo et al., 2015). Somit ist auch mit dieser Technik keine vollständige Sensitivität zu gewährleisten. Des Weiteren weist das NGS eine gewisse falsch-positiv Rate auf, weshalb momentan noch für diagnostische Zwecke alle mit NGS detektierten Varianten mit Sanger validiert werden müssen.

5.2 CNV-Analyse

Die für diese Arbeit durchgeführte genomweite CNV-Analyse von 83 nicht-verwandten, mutationsnegativen HNPCC-Patienten hatte zum Ziel, neue ursächliche Gene für das LS aufzudecken. Da bei dem zugrundeliegenden Phänotyp der in dieser Studie eingeschlossenen Patienten (junges Erkrankungsalter, über die Hälfte der ausgewählten Patienten mit positiver Familienanamnese) von einer monogenen genetischen Ursache auszugehen ist, wurde nach seltenen und somit potenziell hochpenetranten CNVs gesucht. Insbesondere bei erfüllten Amsterdam-Kriterien (Tumoren aus dem engeren HNPCC-Spektrum bei mindestens drei erstgradig Verwandten) ist von einer monogen-dominanten Vererbung auszugehen.

Die Anwendung des Hi- und RVI-Scores zur Bewertung der Dosissensitivität und Variantenintoleranz, die Berücksichtigung der Art des CNVs (Deletion/Duplikation, partiell/komplett) und seiner beschriebenen Funktion, eine *in-silico* Netzwerk-, Pathway- und Segregations-Analyse sowie der Vergleich mit bereits durchgeführten CNV-Studien bei Patienten mit ungeklärten Tumorsyndromen dienten der Selektion möglicher ursächlicher Kandidatengene. Weiterhin wurde berücksichtigt, ob die Gene eine zusätzliche Keimbahnvariante aufwiesen, ob sie in Darmschleimhaut exprimiert werden, oder ob sie bei Adenokarzinomen des TCGA von somatischen Mutationen betroffen sind.

Das Patientenkollektiv der CNV-Analyse wurde in sieben Zentren aus Deutschland und den Niederlanden rekrutiert. Die CNV-Daten von zehn der insgesamt 83 Patienten (12%) mussten zu Beginn der Analyse aufgrund einer zu großen Standardabweichung des Intensitätssignals Log R Ratio ($LRR > 0,3$) entfernt werden. Die Ursache für eine zu große Standardabweichung der LRR ist in den meisten Fällen eine minderwertige DNA-Qualität. Qualitätsverluste der DNA sind bei langer oder unsachgemäßer Lagerung oder bei unsauberer Extraktion festzustellen. Auch mechanische Einwirkungen, wie zu starkes Vortexen oder zu hohe Zentrifugalkräfte, können zu Brüchen in der DNA-Struktur führen, die eine nachfolgende sensitive Untersuchung verhindern.

Für die Berechnung der CNVs wurde der QuantiSNP-Algorithmus beruhend auf dem *Hidden Markov Model* verwendet. Programme wie *cnvPartion* oder *PennCNV* sind gebräuchliche Alternativen, die in Tests unterschiedliche *Calling*-Ergebnisse erzielten (Marenne et al., 2011). Die Anwendung mehrerer Algorithmen minimiert die Anzahl an falsch-positiven *calls* (erhöhte Spezifität). Gleichzeitig führt dies zu Einbußen in der Sensitivität. In einer vergleichenden Studie

schnitt QuantiSNP als leistungsfähigster Algorithmus unter den sieben gebräuchlichsten CNV-Algorithmen ab (Dellinger et al., 2010). Wir verwendeten daher QuantiSNP als einziges Programm und benutzten weitere Filterkriterien (Größenfilter, Anzahl konsekutive SNPs, Log Bayes Faktor), um die Zahl an falsch-positiven CNVs auf ein Minimum zu reduzieren.

Die 73 ausgewerteten Patienten wiesen nach dem QuantiSNP *calling* 8.460 CNVs durchschnittlich 116 CNVs pro Patient, auf. Duplikationen sind meistens mit einem milderem Phänotyp verbunden und stehen unter weniger negativem Selektionsdruck (Zarrei et al., 2015). Duplikationen kommen daher im Allgemeinen häufiger im Genom vor, sind jedoch schwieriger zu detektieren und daher häufig in den CNV-Analysen unterrepräsentiert (Buizer-Voskamp et al., 2011; Dauber et al., 2011). In unserer Analyse wurden hingegen deutlich mehr Duplikationen als Deletionen detektiert (2.025 Deletionen zu 6.435 Duplikationen). Unter diesen *gecallten* Duplikationen waren jedoch > 90% artifizieller Natur. Von den 6.435 Duplikationen wiesen 5.603 Duplikationen einen Log Bayes Faktor (LBF) < 10 auf und wurden im ersten Filterschritt ausgeschlossen. Die anschließende Validierung der CNVs ergab, dass selbst Duplikationen mit einem LBF zwischen 10 und 20 nicht nachweisbar waren, woraufhin dieser Validitätsfilter für Duplikationen auf ≥ 20 hochgesetzt wurde. Das Verhältnis Deletionen zu Duplikationen war daraufhin mit 367 Deletionen zu 261 Duplikationen annähernd ausgeglichen. Zwischen den Deletionen und Duplikationen ließen sich Größenunterschiede feststellen. So waren unter den Duplikationen deutlich größere CNVs (> 500 kb) als unter den Deletionen. Dies lässt sich wiederum mit der geringeren phänotypischen Auswirkung von Duplikationen auf den Organismus erklären. Große (≥ 1 Mb) Deletionen in genhaltigen Bereichen resultieren häufig in phänotypisch schwerwiegenden Erkrankungen (z.B. angeborene Fehlbildungen) (Watson et al., 2014). Da unsere Patienten eine spätmanifeste Erkrankung aufweisen und weder eine mentale Retardierung, noch bekannte Dysmorphien oder Fehlbildungen zeigten, war der Nachweis von kleineren Deletionen (≤ 500 kb) mit meist nur ein bis drei gemeinsam betroffenen Genen entsprechend zu erwarten.

Ähnlich wie SNPs haben häufig auftretende CNVs (CNPs) meistens keine nennenswerten phänotypischen Auswirkungen auf den Organismus oder spielen lediglich als niedrig-penetrante Risikofaktoren eine Rolle (Shlien und Malkin, 2010). Selten auftretende CNVs jedoch können als kausal relevante monogene Faktoren potenziell krankheitsverursachend sein. Für die Suche nach hochpenetranten Risikogenen für das HNPCC waren daher seltene CNVs von Bedeutung.

Neben dem Vergleich der Ergebnisse mit CNVs von *In-House*-Kontrollen bietet insbesondere die *Database of Genomic Variants* eine gute Möglichkeit, die Häufigkeit von CNVs in gesunden Kontrollen mit der in Patientenkollektiven zu vergleichen.

Neben der Anwendung von Häufigkeitsfiltern wurden CNVs in nicht-kodierenden Regionen (siehe Abschnitt 5.5 Limitationen) sowie in Regionen segmentaler Duplikationen ausgeschlossen. Segmentale Duplikationen sind duplizierte Genregionen > 1 kb mit einer hohen Sequenzidentität. Sie befinden sich häufig in Telomer- oder Chromatinregionen (Bailey et al., 2001, 2002). Der Nachweis dieser CNVs ist durch unspezifische Primerbindung in diesen Bereichen hoher Sequenzidentität schwierig durchzuführen.

Nach Eingrenzung der Daten auf CNV-Ebene erfolgte eine Eingrenzung auf Gen-Ebene. Es wurden zunächst über die CNV-Analyse detektierte Pseudogene sowie olfaktorische Gene ausgeschlossen. Im menschlichen Genom kommen über 1.000 olfaktorische Rezeptorgene vor, von denen etwa 60% mutiert vorliegen (Gilad et al., 2003; Glusman et al., 2001; Zozulya et al., 2001). Diese Klasse von Genen weist einen niedrigen Selektionsdruck auf und ist sehr tolerant gegenüber Mutationen (Intoleranz-Score > 90%). Denn trotz des gehäuften Auftretens von Mutationen, darunter auch Stopp-Mutationen, lassen sich keine klinischen Konsequenzen daraus ableiten (Petrovski et al., 2013).

Unter Ausschluss der anschließend über das NGS detektierten sieben mutationspositiven Patienten (vgl. Kap. 4.4.1) wurden zusammenfassend 42 seltene CNVs (21 Deletionen und 21 Duplikationen), welche insgesamt 74 Gene betreffen, in 29 von 66 (44%) mutationsnegativen Patienten detektiert. Jeder CNV kam nur bei einem Patienten im Kollektiv vor und es fanden sich keine Überschneidungen unter den CNVs oder den betroffenen Genen innerhalb der Patientengruppe. Diese Befunde sind vergleichbar mit den wenigen bisher durchgeführten genomweiten CNV-Analysen zu ungeklärten familiären Tumorsyndromen, welche ebenfalls individuelle, nicht-rekurrente CNVs identifizierten (Horpaopan et al., 2015; Krepischi et al., 2012; Kuusisto et al., 2013; Lucito et al., 2007; Venkatachalam et al., 2011; Villacis et al., 2015; Willis et al., 2014; Yang et al., 2012). Dies schließt eine Ursächlichkeit der nur von einzelnen Patienten betroffenen Gene allerdings nicht aus. Es ist davon auszugehen, dass bisher nicht identifizierte Gene, die für familiäre Tumorprädispositions-Syndrome ursächlich sind, sehr selten vorkommen. Die 2013 von Palles et. al. detektierte Keimbahnmutation in dem

Polymerase-Gen *POLD1*, die zusammen mit Mutationen in *POLE* für die PPAP verantwortlich gemacht wird, wurde einmalig in einem Kollektiv aus 3.805 Patienten (0,03%) mit ungeklärten multiplen kolorektalen Adenomen und CRC nachgewiesen (Palles et al., 2013). Masson et al. hingegen konnten in ihrem Kollektiv aus 125 HNPCC Patienten einzelne Gene nachweisen, die in bis zu vier Patienten überschneidend als Deletionen oder Duplikationen vorlagen (Masson et al., 2013). Unter diesen Genen fand sich auch *NRG3*, welches bei einem unserer Patienten eine intronische Deletion aufwies (s.u.).

Unter den 66 ausgewerteten mutationsnegativen Patienten der CNV-Analyse wies etwas mehr als die Hälfte (56%) keine seltenen CNVs in proteinkodierenden Genen auf. Für diese Patienten mit ungeklärtem familiärem CRC sind womöglich andere genetische Faktoren, wie Punktmutationen oder häufige niedrigpenetrante Varianten, ursächlich für die CRC-Prädisposition.

Eine Segregationsanalyse war aufgrund der geringen Anzahl noch lebender und als erkrankt einzustufender Familienmitglieder sowie der schwierigen Kontaktaufnahme nur in einem Fall möglich. Die Index-Patientin wies eine partielle Duplikation in *ENOX1* auf. Das kodierende Protein, ein Plasmamembran-Elektronentransporter, erfüllt weitere Funktionen in der Kontrolle des Zellwachstums und Zellüberlebens. In der Arbeit von Kluth et al. wurden 0,5% der untersuchten Prostatakarzinomzellen mutiert vorgefunden (Kluth et al., 2015). Aufgrund seiner Funktion bei der Kontrolle des Zellwachstums und seiner niedrigen Hi- und RVI-Scores (22% und 9%), die auf eine Dosissensitivität und Variantenintoleranz hindeuten, wurde es als interessantes Kandidatengen gewertet. Bei der Segregationsanalyse konnte allerdings keine Segregation mit dem entsprechenden CNV bei der zweimal hintereinander an CRC betroffenen Cousine festgestellt werden. Dies kann entweder bedeuten, dass kein Zusammenhang zwischen dem CNV und der Erkrankung besteht, oder dieser CNV nur als moderater Risikofaktor (neben weiteren) zu werten ist. *ENOX1* wurde daher nicht als final hochpenetrantes Kandidatengen gewertet. Einige Studien konnten eine Segregation einzelner CNVs in betroffenen Familienangehörigen feststellen (Villacis et al., 2015; Yang et al., 2012, 2014). Wir fanden keine Überschneidungen mit den in dieser Studie nachgewiesenen CNVs in den beschriebenen Regionen (12p12.3, 4q13) (Yang et al., 2012, 2014) bzw. den betroffenen Genen (*GALNT1*, *KMT2C*) (Villacis et al., 2015).

Aus der Gesamtheit der durchgeführten Untersuchungen der CNV-Gene (Literaturrecherche, Anwendung bioinformatischer Scores, Segregations-, Pathway-, Netzwerk- und NGS-Analyse, dem Literaturvergleich mit vorangegangenen CNV-Studien bei Patienten mit ungeklärten Tumorsyndromen) wurden die zehn überzeugendsten Kandidatengene (*DLG2*, *DUSP11*, *GNB2L1*, *MCM4*, *PAG1*, *PRKCA*, *PRKDC*, *RUNDC3B*, *TRIM41* und *TRIM52*) selektiert. Nah verwandte Gene von *DUSP11*, *MCM4*, *PRKCA*, *TRIM41*, *TRIM52* wurden bereits in früheren Studien bei Patienten mit ungeklärtem familiärem CRC als Kopienzahlveränderung detektiert (Horpaopan et al., 2015, Masson et al. 2013, Villacis et al., 2015). Hierbei stimmte die Art des CNVs (Del/Dup) und dem daraus abzuleitenden denkbaren Effekt (LoF/*Gain of Function*) bis auf eine Ausnahme mit den bei uns betroffenen Genen überein. Acht der zehn Gene (*GNB2L1*, *MCM4*, *PAG1*, *PRKCA*, *PRKDC*, *RUNDC3B*, *TRIM41* und *TRIM52*) sind in einem gemeinsamen Netzwerk mit MSH2 beschrieben (vgl. Kap. 4.3.6). Sie könnten über einen noch unbekanntem Mechanismus potentiell ursächlich für den MSH2-Verlust sein. *DUSP11*, *GNB2L1*, *PRKCA*, *PRKDC* und *TRIM41* weisen funktionell eine Verbindung zur Proteinkinase A oder C auf, die u.a. an der Regulation des Zellwachstums, der Zelldifferenzierung, der Zelladhäsion und Angiogenese sowie der Tumorgenese beteiligt sind. Drei dieser zehn Gene (*PRKCA*, *PRKDC*, *TRIM41*) sind von zusätzlichen, potentiell pathogenen (CADD-Score ≥ 20) Keimbahnvarianten betroffen. Alle zehn Kandidatengene werden in der Darmschleimhaut exprimiert. Keiner der Kandidatengene wies somatische Mutationen $\geq 3\%$ in nicht-hypermutierten Adenokarzinomen der TCGA-Datenbank auf.

Die final interessantesten Kandidatengene wurden aufgrund der Gesamtheit aller Analyseergebnisse und unter Berücksichtigung der Informationen aus Literatur- und Datenbankrecherche ausgewählt. Weitere zunächst interessant erscheinene Gene wurden letztlich nicht berücksichtigt, da sie nur einzelne Kriterien erfüllten, in der Summe der Analysen aber widersprüchliche Ergebnisse lieferten.

ATP12A (*ATPase H⁺/K⁺ Transporting Non-Gastric Alpha2 Subunit*) und *NRG3* (Neuregulin 3) beispielsweise erschienen zunächst relevant, da CNVs dieser Gene bereits in vorherigen Studien bei Kollektiven mit genetisch ungeklärtem familiärem CRC detektiert wurden (Masson et al., 2013; Villacis et al., 2015; unveröffentlichte Daten unserer Arbeitsgruppe). *ATP12A* reguliert mit Hilfe von ATP den Austausch von H⁺ und K⁺ über die Plasmamembran. Das Protein wird

hauptsächlich in der Luftröhre exprimiert. *NRG3* ist ein Ligand des transmembranen Tyrosinkinase-Rezeptors ERBB4, welcher zur Familie der EGFR (*Epidermal Growth Factor*)-Rezeptoren gehört. Somatische Mutationen in ERBB4 wurden mit Prostata-, Ovarial- und Brustkrebs assoziiert (Hashemi et al., 2016; Kim et al., 2016; Saglam et al., 2017). Aus diesem Grund erschien *NRG3* funktionell relevant für die Entstehung des LS zu sein. Die vorliegende tief-intronische Deletion ist jedoch mit hoher Wahrscheinlichkeit nicht pathogen, da im tief-intronischen Genbereich selten Spleißvorgänge stattfinden. *NRG3* ist ein bekanntes Suszeptibilitätsgen für Schizophrenie und wird primär im Gehirn exprimiert. *ATP12A* und *NRG3* weisen hohe HI- und RVI-Scores auf. Sie sind somit nicht dosissensitiv oder variantenintolerant. Daher wurden sie nicht in die Liste der finalen Kandidatengene mit aufgenommen. *ATP12A* steht zudem funktionell in keinem Zusammenhang zu malignen Erkrankungen.

GRIK4 erschien zunächst besonders relevant, da bei fünf Patienten unserer Studie mögliche pathogene Varianten in diesem Gen identifiziert wurden: Neben der bei uns detektierten partiellen Deletion wurden zusätzlich vier Missense-Mutationen detektiert. Das Gen kodiert für den ionotropen Kainat-4-Glutamatrezeptor. Glutamat ist der bedeutendste exzitatorische Neurotransmitter im zentralen Nervensystem. Eine balancierte Translokation zwischen Exon 2 und 3 in *GRIK4* wurde mit Schizophrenie assoziiert (Pickard et al., 2006). *GRIK4* wird wie *NRG3* im Gehirn exprimiert. Aufgrund seiner Genfunktion und mangels seiner primären Expression im Gehirn wurde es letztlich nicht als finales Kandidatengen gewertet.

Die finalen Kandidatengene und ihre Proteinfunktionen werden im Weiteren näher diskutiert:

DLG2

Discs Large Homolog 2 (DLG2) ist ein Mitglied der Membran-assoziierten Guanylat-Kinase-Familie. Das Protein bildet mit Mitgliedern der DLG-Familie Heterodimere aus, die an postsynaptischen Seiten mit Rezeptoren, Ionenkanälen und assoziierten Signalproteinen interagieren. Eine mögliche Rolle als TSG wird diskutiert, die genaue Funktion des Proteins ist noch unklar. Der in dieser Studie vorliegende betroffene Patient erkrankte im Alter von 49 Jahren an CRC, die Schwester mit 50 Jahren an einem Endometriumkarzinom. Über die CNV-Analyse wurde eine tief-intronische Deletion von 31 kb in *DLG2* detektiert. Eine Beeinträchtigung der Proteinfunktion wäre somit vor allem über ein aberrantes Spleißen oder betroffene regulatorischer Bereiche zu erklären. Tief-intronische Spleißstellen sind

unvollständig erfasst und wenig beschrieben, da häufig funktionelle Studien fehlen, die ihre Auswirkungen auf die Proteinfunktion untersuchen (Svaasand, 2015). *In silico*-Programme zur Spleiß-Prädiktion beschränken sich meist nur auf die ersten 15 bis 41 Basen ab der kanonischen 5', 3' Spleißstelle (Baralle und Baralle, 2005). *DLG2* weist einen niedrigen HI- und RVI-Score auf (1% und 4%), die einen Funktionsverlust des Proteins bei Vorliegen einer Deletion/Mutation anzeigen. Bei der Pathway-Analyse wurde *DLG2* im *Hippo-Pathway* gemeinsam mit den CRC-Kandidatengen *SMAD2* und *SMAD4*, *PPP2CA* und *PPP2R1B* und *PRKCZ* beschrieben. Dieser Signalweg ist für die Regulation der Zellproliferation und Apoptose verantwortlich. Es konnten keine zusätzlichen Punktmutationen in *DLG2* in unserem sequenzierten Kollektiv nachgewiesen werden. Auch wenn *DLG2* in keinem direkten Zusammenhang zu *MSH2* steht, ist die Funktion des Gens als möglicher Tumorsuppressor interessant. Mögliche aberrante Spleißprodukte die durch die intronische Deletion auftreten könnten, müssten mittels einer RNA-Untersuchung (Transkriptanalyse) untersucht werden. Eine Transkriptanalyse war für den hier vorliegenden Fall allerdings nicht möglich, da kein entsprechendes Material von dem Patienten erhalten werden konnte.

DUSP11

DUSP11 ist ein Mitglied der dual-spezifischen Phosphatase-Familie. Sie inaktivieren ihre Zielkinasen durch Dephosphorylierung der Phosphoserin/-threonin und Phosphotyrosinreste. Sie inhibieren Mitglieder des MAPK-Signalwegs (MAPK/ERK, SAPK/JNK, p38), welcher mit zellulärer Proliferation und Differenzierung assoziiert und bei der Entwicklung von Krebserkrankungen involviert ist (Beeram et al., 2005). Eine erhöhte Expression von *DUSP11* wurde in einem *in vitro* Modell zur Nachbildung der *in vivo* Tumorformation bei Kolonkarzinomzellen vorgefunden (Dardousis et al., 2007). Im hier untersuchten Kollektiv liegt *DUSP11* bei einem Patienten partiell dupliziert vor. Unter der Annahme eines LoF-Effekts passt die beschriebene Funktion von *DUSP11* (Inhibitor des MAPK-Signalweges) zu der Hypothese einer kausalen Beteiligung am Phänotyp des Patienten. Der Patient erkrankte im Alter von 50 Jahren an einem CRC, seine Mutter bereits im Alter von 45 Jahren an einem Karzinom unklarer Ätiologie. Die Pathway-Analyse ergab eine Beteiligung von *DUSP11* am Proteinkinase-A-Signalweg. Dieser Signalweg ist assoziiert mit der Entstehung und dem Fortschreiten verschiedener Tumorarten (Caretta und Mucignat-Caretta, 2011). Eine weitere CNV-Studie an HNPCC-Patienten detektierte eine vollständige Duplikation in *DUSP8*, einem nah verwandten

Gen mit identischer Funktion. Bei einer vollständigen Duplikation ist von einer Überexpression des Proteins auszugehen, was wiederum nicht zu der bekannten Proteinfunktion passt. Die Ursächlichkeit von *DUSP11* ist damit derzeit unklar.

GNB2L1 (RACK1)

Das *Guanin Nukleotide Binding Protein, Receptor of Activated Protein Kinase C; Human Lung Cancer Oncogene, Proliferation Inducing Gene* (GNB2L1) taucht bei einer Reihe relevanter Signalwege, wie dem Proteinkinase-A-Signalweg, CRC-Metastase-Signalweg, Gαq-Signalweg, CXCR4-Signalweg, PSY-Purigenic-Rezeptor-Signalweg, Gβγ-Signalweg und *Signaling by Rho-Family GTPase* auf. Aus den Daten der Netzwerk-Analyse ist abzuleiten, dass GNB2L1 über PRKCA und PRKDC mit MSH2 in Verbindung steht. Diese Verbindung ist experimentell begründet. Sie beruht auf der gemeinsamen Beteiligung dieser Proteine am Integrin-Adhensionskomplex (Byron et al., 2012). Dieser Komplex reguliert sowohl die Zell-Zell-Interaktion als auch die Interaktion zwischen Zellen und der extrazellulären Matrix. Hierüber werden wichtige Signaltransduktionsvorgänge und die Rekrutierung verschiedener Proteine vermittelt (Byron et al., 2012). GNB2L1 weist eine veränderte Expression (Über- als auch Unterexpression) in einer Reihe von Tumoren (u.a. Magenkarzinom, Speiseröhrenkrebs, CRC, Ovarial-Karzinom) auf (Chen et al., 2015; Cheng et al., 2016; Jin et al., 2014; Li und Xie, 2015; Lin et al., 2014; Wang et al., 2015). Je nach Zellkontext werden ihm sowohl apoptotische als auch antiapoptotische Zellwachstumkontroll- oder Zellwachstumsinhibitions-Funktionen zugeschrieben. In Kolonkrebszellen wurde eine Interaktion zwischen GNB2L1 und Fem1b, einem pro-apoptotischen Protein, nachgewiesen. Erhöhte GNB2L1-Level führen demnach zur Degradierung von Fem1b. Durch die Degradierung wird der induzierte Zelltod verhindert und eine Progression der Krebszellen ist möglich (Subauste et al., 2009). In der vorliegenden Studie liegt *GNB2L1* bei dem mit 49 Jahren an einem CRC erkrankten Patienten als vollständige heterozygote Duplikation vor, die zu einer Überexpression des Proteins führen könnte. Zudem besteht über den Integrin-Adhensions-Komplex eine Verbindung von GNB2L1 zu MSH2. Wie genau die einzelnen Proteine dabei miteinander in Kontakt treten, ist noch ungeklärt.

MCM4

MCM4 gehört zum Minichromosomalen-Maintenance-Komplex, der für die Initiierung der eukaryotischen Genomreplikation essentiell ist. Punktmutationen in diesem Gen werden mit

Brustkrebs in Verbindung gebracht. Bei Mäusen führte die Mutation zu einer beeinträchtigten Proteinfunktion, was mit einem gestörten Zellwachstum und der Entwicklung von genomischer Instabilität einherging (Chuang et al., 2012; Shima et al., 2007). In diesem Gen wurde bei einem Patienten des Kollektivs eine partielle Deletion von 91 kb detektiert, die sowohl den 3'-Bereich von *PRKDC* als auch von *MCM4*, welche auf Chromosom 8 Kopf an Kopf liegen, umfasst. Die betroffene Patientin erkrankte mit 48 Jahren an einem EC und mit 72 Jahren an einem CRC. Die Schwester erkrankte mit 30 Jahren an einem EC und mit 60 Jahren an einem CRC. Die Mutter hatte Leberkrebs und der Vater Leukämie. Die dem Gen zugeordneten Scores (HI-Score 2%, RVI-Score 4%) indizieren, dass Mutationen in diesem Gen äußerst selten sind und eine Auswirkung auf die Proteinfunktion zu erwarten ist. Das Protein steht im engen Netzwerk mit MSH2 und interagiert dort zusammen mit PRKDC, einer DNA-aktivierten Proteinkinase, die an der Reparatur von DNA-Doppelstrangbrüchen involviert ist (Karmakar et al., 2010). Über die Pathway-Analyse wurde *MCM4* im Zell-Zyklus-Kontroll-Signalweg gemeinsam mit *CHEK2*, einem wichtigen Tumorsuppressorgen, aufgefunden. In einer weiteren CNV-Studie wurde bei Patienten mit ungeklärter adenomatöser kolorektaler Polyposis eine partielle Duplikation in *MCM3AP* sowie eine Frameshift-Mutation und zwei vermutlich pathogene Missense-Mutationen nachgewiesen (Horpaopan et al. 2015). Alle detektierten Varianten sind somit kompatibel mit einem erwarteten LOF-Effekt. *MCM3AP* bewirkt die Acetylierung von *MCM3*, einem weiteren Protein des Minichromosom-Maintenance Komplexes. Acetyliertes *MCM3* bewirkt eine Inhibierung der DNA-Replikation. Damit verhindert es das Fortschreiten des Zellzyklus (Takei et al., 2002). Die reduzierte Expression eines MCMs bewirkt zudem eine reduzierte Expression aller Proteine des Komplexes.

PAG1

In der CNV-Analyse wurde bei einem mit 40 Jahren an einem Hodgkin Lymphom und mit 44 Jahren an CRC erkranktem Patienten in *PAG1* eine partielle Deletion nachgewiesen, welche mit einem Funktionsverlust des Proteins einhergehen könnte. *Phosphoprotein Associated with Glycosphingolipid-enriched microdomains 1* (*PAG1*) ist ein potentiell TSG, welches in vielen Krebszellen herunterreguliert ist (Suzuki et al., 2011). Das Protein wirkt inhibierend auf seine Zielkinasen und unterdrückt dadurch die Zellproliferation und Antigenstimulation (Rolland et al., 2014). In Tumorzellen konnte eine Inhibierung der Invasion und Metastasierung von Krebszellen durch *PAG1* gezeigt werden (Yu et al., 2010). Die bei uns nachgewiesene Mutation

bei dem Patienten passt somit zu der beschriebenen Funktion von *PAG1* als TSG. In der Netzwerk-Analyse zeigte sich eine entfernte Verbindung zu MSH2.

PRKCA

Die Proteinkinase C Alpha (PRKCA) gehört zur Familie der Serin- und Threonin-spezifischen Proteinkinasen. *PRKCA* gilt als Tumorsuppressorgen, da es die Aktivierung des EGFR (*Epidermal Growth Factor*)-Rezeptors verhindern kann (Koese et al., 2013). Der EGFR-Rezeptor ist ein Transmembranrezeptor, welcher im aktivierten Zustand über eine Kaskade biochemischer Prozesse zur weiteren Aktivierung von *downstream* Signalwegen wie dem Ras/Raf MAP-Kinase-, PI3K/Akt- und Jak2/STAT3-Signalweg führt. Am Ende der Kaskade werden Zellproliferation, Angiogenese und Metastasierung angeregt und der programmierte Zelltod verhindert (Prenzel et al., 2001). Der EGFR-Rezeptor wird von normalen als auch von Tumorzellen exprimiert. Er liegt bei einer Reihe epithelialer Tumoren, wie dem Nicht-Kleinzelligen Bronchialkarzinom (40-80%), CRC (72-82%), Kopf-Hals-Karzinom (95-100%), Brustkrebs (14-91%) und Nierenkrebs (50-90%), überexprimiert vor. Die antiproliferative Aktivität der Proteinkinase C beruht auf der Hemmung von EGFR und wurde in Kolonepithelzellen nachgewiesen. Eine erniedrigte Aktivität von PRKCA wurde in kleinen Adenomzellen beschrieben (Assert et al., 1999). PRKCA gilt als dosissensitiv und mutationsintolerant (Hi-score 5%, RVI-Score 9%). Die bei uns vorliegende intronische Deletion könnte über ein aberrantes Spleißen (Überspringen eines Exons, Einbehalten des Introns, Einfügen einer neuen Spleißstelle) zu einem Funktionsverlust dieses Proteins geführt haben. Der Patient, welcher die *PRKCA*-Deletion aufwies, entwickelte mit 43 Jahren, sein Vater mit 55 Jahren ein CRC. Eine Interaktion mit MSH2 wurde über die *in-silico* Netzwerk-Analyse beschrieben. Mittels NGS konnte neben dem CNV bei einem weiteren Patienten eine Punktmutation (*Inframe*-Deletion) nachgewiesen werden, die als potentiell pathogen einzuschätzen ist (CADD-Score 23). In einem vergleichbaren Kollektiv von HNPCC-Patienten wurde eine partielle Duplikation eines nah verwandten Gens (*PRKCI*) detektiert. Auch bei dieser genetischen Alteration ist von einem LoF-Effekt auszugehen.

PRKDC

PRKDC kodiert für die *DNA-dependent protein kinase catalytic subunit* (DNA-PKcs), dem Schlüsselprotein des NHEJ (Dietlein et al., 2014). PRKDC gehört zur Familie der Pi3/Pi4-Kinasen und ist eine Untereinheit der DNA-abhängigen Serin/Threonin-Kinase DNA-PK. Im Darmgewebe

konnte gezeigt werden, dass eine funktionelle Beeinträchtigung von DNA-PKcs zu einer signifikanten Erhöhung von Fehlpaarungen führt. Der Signalweg tritt als Antwort von DNA-Schädigungen auf und ist zelltypspezifisch. Für PRKDC wurde über die Netzwerk-Analyse eine direkte Interaktion zu MSH2 aufgezeigt (von Eyss et al., 2012). Das Gen ist im *Damage-Checkpoint-Regulation*-Signalweg gemeinsam mit wichtigen CRC-Genen wie *TP53*, *PRKCZ* und *CHEK2* involviert. In der vorliegenden Studie wurde eine partielle Deletion und eine synonyme Variante detektiert. Die Deletion umfasst *PRKDC* und *MCM4*, welche auf Chromosom 8 Kopf an Kopf nebeneinander liegen. Die Deletion überdeckt beide Gene zur Hälfte. Bei der betroffenen Patientin wurde EC und CRC im Alter von 48 und 72 Jahren diagnostiziert. Die Schwester entwickelte ebenfalls ein EC und CRC, die Mutter hatte Leberkrebs und der Vater Leukämie. Durch den niedrigen HI-Score von *PRKDC* (9%) ist von einer beeinträchtigten Proteinfunktion bei Ausfall eines Allels auszugehen.

RUNDC3B

Das Proteinprodukt von *RUN Domain Containing 3B (RUNDC3B)* interagiert über seine N-terminale Region mit Rap2, einer wichtigen Komponente der mitogen-aktivierten Proteinkinase (MAPK)-Kaskade, welche die zelluläre Proliferation und Differenzierung reguliert. RUNDC3B dient möglicherweise als Mediator zwischen der Rap2 und der MAPK-Signalkaskade (Burmeister et al., 2015). Es soll an der Transformation und Progression von Brustkrebs beteiligt sein (Raguz et al., 2005). Über die Netzwerk-Analyse wurde es als entfernter Interaktionspartner von MSH2 ermittelt. Ein Patient unseres Kollektivs wies eine partielle heterozygote Deletion von 34 kb auf, welche die Exons 3 und 4 (von insg. 11 Exons) des Gens einschließt. Der Patient erkrankte mit 36 Jahren an einem Appendixkarzinom. Die Familiengeschichte ist unauffällig.

TRIM41 (RINCK) und TRIM52

Tripartite Motif Containing 41 (TRIM41) kodiert für eine E3-Ligase der Proteinkinase C und spielt eine mögliche Rolle bei der Ubiquitinierung und Proteindegradierung von PKC. Da PKC als Tumorsuppressor eingeordnet wird, kommt *TRIM41* somit eine onkogene Funktion zuteil. Für TRIM41 konnte über die Netzwerk-Analyse eine experimentell bewiesene Verbindung zu MSH2 hergestellt werden (vgl. Kap. 4.3.6). Neben der partiellen Duplikation wurden zwei Missense-Mutationen in *TRIM41* detektiert. Als zusätzliche Besonderheit wurde dabei eine der

Punktmutationen in demselben Patienten (HNPCC_38) nachgewiesen. Dies könnte auf eine rezessive Vererbung hindeuten. Die Missense-Mutation liegt auf Höhe des CNVs, wobei nicht bekannt ist, ob sie auf dem duplizierten oder dem gegenüberliegenden Allel liegt. Die hier vorliegende Duplikation umfasst weiterhin vollständig die Gene *GNB2L1* (s.o.) und *TRIM52*. Mit der Duplikation auf *TRIM52* liegen somit zwei Duplikationen in einem Gen der TRIM-Familie vor. Die genaue Funktion von *TRIM52* ist noch unbekannt. *TRIM52* ist ebenfalls in dem Netzwerk mit MSH2 beteiligt. Es wurden keine weiteren Punktmutationen in *TRIM52* nachgewiesen, auch war es nicht an der Pathway-Analyse beteiligt. In einer vorherigen CNV-Studie bei HNPCC-Patienten wurden weitere Duplikationen in einem Gen der TRIM-Familie detektiert. Dort wurden zwei vollständige Gen-Duplikationen für *TRIM69* beschrieben (Masson et al., 2013). *TRIM41*, *TRIM52* und *TRIM69* werden in der Darmschleimhaut exprimiert. Die Familie der TRIM-Proteine ist an der Zellproliferation, Zellzyklusregulation, Apoptose und Immunantwort beteiligt (Chen et al., 2007). Aufgrund der vorliegenden Befunde sind die Gene *TRIM41* und *TRIM52* spannende Kandidaten. Ob sie ursächlich für den MSH2-Ausfall sein könnten, müsste durch nachfolgende Analysen (Überexpressionsstudie in Zellmodellen) näher untersucht werden.

5.3 Sequenzierung der CRC-Gene der Literatur

Zusätzlich zu der Suche nach weiteren Punktmutationen in CNV-Genen wurden bekannte CRC-Gene bei der zielgerichteten Sequenzierung mitberücksichtigt. In diesen CRC-Genen fanden sich weitere Missense- und Spleiß-Mutationen (Kap. 4.4.1). In *APC* wurden zwei Missense-Mutationen entdeckt. Missense-Mutationen sind für *APC* ungewöhnlich; bei den allermeisten handelt es sich tatsächlich um an den Exon-Enden gelegene Spleißmutationen (Friedl und Aretz, 2005; Kaufmann et al., 2009). Die in dieser vorliegenden Studie Missense-Mutationen befinden sich nicht an den Exon-Enden von *APC*, daher ist hier am ehesten von seltenen Polymorphismen oder niedrig-penetrante Varianten auszugehen, die keine oder allenfalls geringe Auswirkung auf die Proteinfunktion haben. Die Patienten wiesen zudem klinisch keinerlei Anzeichen einer polypösen Erkrankung auf.

Weitere Missense-Mutationen wurden in *CHEK2*, *MAP2K4*, und *POLE* detektiert. *MAP2K4* (*Mitogen-activated protein kinase*) ist bei einer Reihe zellulärer Prozesse wie Proliferation, Differenzierung und Transkriptionsregulation beteiligt. Das Gen liegt bei zahlreichen

Krebserkrankungen mutiert vor. Die biochemischen Eigenschaften der mutierten Kinase und ihre Rolle bei der Tumorgenese sind noch nicht vollständig bekannt. In *CHEK2* und *POLE* wurden Missense-Mutationen mit besonders hohen CADD-Scores (> 30) detektiert. Sie zählen damit zu den höchsten 0,1% der C-Scores aller genomweit berücksichtigten Mutationen. Dies spricht für eine hohe Wahrscheinlichkeit eines krankheitsverursachenden Effekts dieser Mutationen. *CHEK2* (Checkpoint Kinase 2) ist ein Zellzyklus-Checkpoint-Regulator und ein potenzieller Tumorsuppressor. Es ist ein bekanntes, mäßig penetrantes Krebsgen, dessen Zusammenhang mit Darmkrebs kontrovers diskutiert wird. Heterozygote Mutationen in *CHEK2* werden mit einem erhöhten Risiko für Brustkrebs in Verbindung gebracht (Bell et al., 1999; Cybulski et al., 2006; Walsh et al., 2006). Das Proteinprodukt von *CHEK2* reguliert über die Phosphorylierung von p53/TP53 die Apoptose. Bei Abwesenheit des Proteins im Tumor z.B. durch somatische Mutationen beobachtet man eine chromosomale Instabilität des Tumors. Die bei uns detektierte Missense-Mutation (*CHEK2*:c.1441G>T;p.Asp481Tyr) ist in der LOVD-Datenbank (*Leiden Open Variation Database*) als sehr seltene Mutation mit einer Frequenz von 4/13006, hinterlegt. Der Effekt dieser Missense-Mutation ist derzeit unklar. Bei zwei Patienten wurden *POLE*-Mutationen mit einem CADD-Score > 30 detektiert (c.2770C>T;p.Arg924Cys und c.4709G>A;p.Arg1570Gln) (Kap. 4.2 „Differentialdiagnosen“). *POLE*-Mutationen sind als Ursache für die sehr seltene Polymerase-Proofreading-Assoziierte-Polyposis beschrieben (Vgl. Kap. 2.3.5) (Palles et al., 2013). Bisher ist allerdings nur eine Mutation in der hochkonservierten Exonuklease-Domain in Exon 13 (c.1270C>G;p.Leu424Val) als ursächlich charakterisiert worden (Elsayed et al., 2014; Palles et al., 2013; Spier et al., 2015; Valle et al., 2014). Die im Rahmen dieser Untersuchung gefundenen *POLE*-Mutationen befanden sich nicht in der Exonuklease-Domäne weswegen man zum heutigen Zeitpunkt nicht von einer pathogenen Ursache dieser Mutationen ausgehen kann.

In den Genen des Proteasom-Signalwegs (Kap. 4.4.2) fanden sich sechs Mutationen in vier der acht untersuchten Gene (in *AKT1*, *HERC1*, *TSC1* und *TSC2*). Zwei der drei Missense-Mutationen in *HERC1* (c.9508C>T;p.Arg3170Cys und c.10660C>T;p.Arg3554Trp) sind aufgrund der CADD-Scores von 35 von besonderer Relevanz. Kein Patient trug mehr als eine Mutation in einem der Proteasom-Pathway-Gene, was mit der Theorie übereinstimmt, dass für die Tumorentstehung in der Regel nur ein entscheidendes Gen innerhalb eines Pathways mutiert sein muss. Es wurde beschrieben, dass somatische Deletionen, welche zu einem Ausfall eines Gens des Proteasom-

Pathways führen, eine MSH2/MSH6-Downregulation zur Folge haben. Dies wurde bei akuter myelotischer Leukämie und sporadischem Darmkrebs nachgewiesen (Diouf et al., 2011). Dieser Mechanismus wurde bisher nicht für Keimbahn-, bzw. Punktmutationen beschrieben. Um die Plausibilität dieser Hypothese zu überprüfen, sollten daher weitere Untersuchungen, wie Expressionsstudien, durchgeführt werden. Da der beschriebene Mechanismus posttranskriptionell abläuft, sind hohe mRNA-Expressionslevel von MSH2 und MSH6 im Tumor bei gleichzeitigem Proteinausfall zu erwarten. Heterozygote Keimbahnmutationen der genannten Proteasom-Pathway-Gene sind außer bei dem beschriebenen Interaktionsmechanismus mit MSH2 (Diouf et al., 2011) auch für definierte Tumor-Syndrome verantwortlich.

Eine spezifische, aktivierende Mutation im *AKT1*-Gen ist als somatisches Mosaik Ursache des Proteus-Syndroms (Lindhurst et al., 2011). Bei diesem Syndrom erfahren bestimmte Körperteile im Laufe des kindlichen Wachstums eine asymmetrische Vergrößerung und Malformationen. Im späteren Alter entwickeln sich Tumoren (Cohen, 2014). Trunkierende Keimbahnmutationen in *TSC1* und *TSC2* sind verantwortlich für die tuberöse Sklerose, die von Tumoren des Gehirns, Hautveränderungen, epileptischen Anfällen und geistiger Retardierung gekennzeichnet ist. Die Erkrankung wird autosomal-dominant vererbt, etwa die Hälfte der Fälle entsteht *de-novo* (Crino et al., 2006). Der Phänotyp unserer Patienten passt nicht zu diesen Krankheitsbildern. Die Wirkung der vorliegenden Mutationen auf die Funktion der Proteine lässt sich nicht vorhersagen. Eine hochpenetrante aktivierende bzw. trunkierende Mutation von *AKT1* bzw. *TSC1/-2* ist jedoch auszuschließen.

5.4 Sequenzierung von Tumor-DNA

Bei der Sequenzierung der Tumor-DNA von insgesamt elf Patienten konnten drei somatische *MSH2*-Mutationen bei zwei der elf Patienten identifiziert werden (Kap. 4.2.1). Einer dieser beiden Patienten wies hierbei zwei somatische, möglicherweise biallele, *MSH2*-Missense-Mutationen auf (c.1993C>G;p.His665Asp und c.1835C>T;p.Ser612Leu). Dadurch ergibt sich für diesen Patienten ein Hinweis auf eine sporadische Erkrankungsursache. Der Patient erkrankte mit 81 Jahren an einem Dünndarmkarzinom. Aus der Familiengeschichte ist wenig bekannt: Die Mutter erkrankte in einem unbekanntem Alter an einem „Karzinom“. Die revidierten Bethesda-Kriterien sind für diesen Patienten somit vermutlich nicht erfüllt. Da der Patient erst in hohem Alter erkrankte, ist eine sporadische, nicht-erbliche Ursache plausibel. Bei einem weiteren

Patienten (HNPCC_06) wurde eine heterozygote somatische *MSH2*-Frameshift-Mutation detektiert. Diese sicher pathogene Mutation könnte, zusammen mit einer weiteren somatischen Mutation (beispielsweise einer Deletion), zu einer sporadischen Tumorerkrankung geführt haben. In der Familiengeschichte des Betroffenen sind keine HNPCC-typischen Tumoren bekannt. Kürzlich veröffentlichte Studien fanden bei über 50% der untersuchten, mutationsnegativen Patienten mit einem Verdacht auf ein LS zwei somatische Ereignisse (Geurts-Giele et al., 2014; Haraldsdottir et al., 2014; Mensenkamp et al., 2014). Die Suche nach größeren Deletionen konnte aufgrund von unzureichender Menge und Qualität der Tumor-DNA (archiviertes FFPE-Material) nicht ausgewertet werden. Da insgesamt nur von elf Patienten Tumor-DNA zur Verfügung stand, war eine Bestimmung der Rate an somatischen Ereignissen in dem Kollektiv außerdem nicht möglich. Interessanterweise wiesen sowohl die Patienten mit nachträglich ursächlich bestätigter Keimbahnmutation in den MMR-Genen, als auch die Patienten mit möglicher sporadischer Erkrankungsursache, mit einer einzigen Ausnahme, keine seltenen CNVs auf. Dies unterstützt die Hypothese, dass die detektierten, seltenen CNVs einen kausalen Beitrag zur Entstehung des Phänotyps der betroffenen Patienten leisten. Somatische Mutationen in MMR-Genen sind allerdings kein Beweis für ein sporadisches CRC. Auch bei bestimmten erblichen CRC-Syndromen wie der MAP oder PPAP können als Folge der gestörten DNA-Reparatur gehäuft somatische Mutationen in MMR-Genen auftreten.

5.5 Limitationen

Wie bei anderen, vergleichbaren wissenschaftlichen Studien war es trotz gründlicher Untersuchung und der Verwendung neuester Technik nicht möglich, das gesamte Spektrum molekulargenetischer Veränderungen zu erfassen. Mittels CNV-Analyse wurden unbalancierte Strukturvarianten detektiert. Balancierte Strukturvarianten, wie Inversionen oder Translokationen, konnten durch die angewandte Technik nicht erfasst und ausgewertet werden. Ebenso war es aufgrund des verwendeten SNP-Arrays technisch nicht möglich, CNVs kleiner als 10 kb sicher zu erfassen und abzubilden. CNVs in nicht-kodierenden Regionen wurden, mit Ausnahme von intronischen Genbereichen und der Region 1,5 Mb um *MSH2* herum, nicht berücksichtigt. Zum einen wiesen die SNP-Marker des Arrays in den nicht-kodierenden Regionen große Lücken auf, zum anderen ist die Interpretation von CNVs in genfreien Bereichen äußerst schwierig und bedarf weiterer funktioneller Aufarbeitung die im Rahmen dieser Arbeit nicht möglich gewesen wäre. Durch die Anwendung stringenter

Filterkriterien wurde die Zahl häufig auftretender CNVs stark reduziert. Für die Suche nach dominanten Varianten wurde die Allelfrequenz auf $\leq 0,01\%$ gesetzt. In der Allgemeinbevölkerung häufig auftretende CNVs würden allenfalls moderat und gemeinsam mit Umweltfaktoren zur Erkrankung beitragen; sie wurden deshalb in dieser Arbeit nicht berücksichtigt. Mosaik, d.h. Varianten, die nur in einem Teil der Zellen vorkommen, wurden ebenso nicht berücksichtigt. Es ist davon auszugehen, dass niedriggradige (*low-level*) Mosaik eine untergeordnete Rolle für die Krankheitsentstehung spielen. Des Weiteren ist eine Validierung und Unterscheidung zwischen niedrig auftretenden Mosaiken und Artefakten schwierig.

6 Zusammenfassung

Bei bis zu 50% der Patienten, welche die klinischen Kriterien für das Lynch-Syndrom (LS) erfüllen, lässt sich keine ursächliche Keimbahnmutation in den bekannten Mismatch-Repair (MMR)-Genen nachweisen. Das frühe Auftreten von Tumoren (primär kolorektales Karzinom (CRC) oder Endometriumkarzinom) bzw. das gehäufte Auftreten dieser Tumoren innerhalb der betroffenen Familien, deuten zusammen mit den typischen Tumormerkmalen (Ausfall eines MMR-Gens und hohe Mikrosatelliteninstabilität) auf eine erbliche Erkrankungsursache hin. Das Ziel unserer Arbeit war es daher, (i) neue genetische Faktoren für das LS mit MSH2-Defekt aufzudecken, (ii) in der Literatur beschriebene alternative Ursachen für den MSH2-Ausfall zu untersuchen und (iii) die klinische Relevanz der Kandidatengene durch weitere Untersuchungen zu erforschen (Segregationsanalyse, Screening nach Keimbahnpunktmutationen, Netzwerk- und Pathway-Analyse).

In der Ätiologie einer Vielzahl von Tumorsyndromen spielen konstitutive Kopienzahlvarianten (*Copy Number Variations*, CNVs) eine bedeutende Rolle. Für die Suche nach weiteren ursächlichen Genen führten wir deshalb eine genomweite CNV-Analyse in einem Kollektiv mutationsnegativer LS-Patienten durch. Nach Anwendung stringenter Filterschritte (Ausschluss von Artefakten, Vergleich mit CNVs gesunder Kontrollen) wurden 43 seltene Keimbahn-CNVs (21 Deletionen und 22 Duplikationen) bei 30 Patienten (45%) identifiziert, welche 76 proteinkodierende Gene betreffen. Übereinstimmend mit publizierten CNV-Studien verwandter Phänotypen trat jeder CNV nur einmal innerhalb des Patientenkollektivs auf. Es gab keine Überschneidungen zwischen den betroffenen Genen. Keiner der CNVs betraf die regulatorischen Bereiche von *MSH2*.

Segregationsanalysen zur weiteren Abklärung der pathogenen Relevanz der CNVs konnten abgesehen von einem Fall mit fehlender Segregation (*ENOX1*) aufgrund fehlender weiterer betroffener Verwandter oder mangels entsprechendem Material nicht durchgeführt werden.

Ein zielgerichtetes Sequenzieren mittels *Next-Generation-Sequencing* zur Suche nach zusätzlichen Punktmutationen in den CNV-Kandidatengenen identifizierte 15 weitere potentiell pathogene Keimbahnmutationen in elf der 68 untersuchten CNV-Gene. Hierbei traten zum überwiegenden Teil Missense-Mutationen unklarer Signifikanz auf; trunkierende Mutationen

wurden nicht detektiert. Lediglich im *TRIM41-Gen* fand sich bei einem Patienten ein Hinweis auf eine rezessive Vererbung.

Unter Berücksichtigung aller durchgeführten Untersuchungen der CNV-Gene (Literaturrecherche, Anwendung bioinformatischer Scores, Segregations-, Pathway-, Netzwerk- und NGS-Analyse, dem Literaturvergleich mit vorangegangenen CNV-Studien bei Patienten mit ungeklärten Tumorsyndromen) konnten die zehn überzeugendsten Kandidatengene (*DLG2*, *DUSP11*, *GNB2L1*, *MCM4*, *PAG1*, *PRKCA*, *PRKDC*, *RUNDC3B*, *TRIM41*, *TRIM52*) selektiert werden. Die Gene *DUSP11*, *GNB2L1*, *PRKCA*, *PRKDC* und *TRIM41* weisen funktionell eine Verbindung zur Proteinkinase A oder C auf. Diesen Proteinkinasen kommt eine wichtige Bedeutung bei der Regulation des Zellwachstums, der Zelldifferenzierung, der Zelladhäsion und Angiogenese sowie der Tumorgenese zu. *MCM4* und *PRKDC* sind an der DNA-Reparatur involviert; zusammen mit *GNB2L1*, *PAG1*, *PRKCA*, *RUNDC3B*, *TRIM41* und *TRIM52* sind sie zudem in einem gemeinsamen Netzwerk mit *MSH2* beschrieben. Sie könnten aufgrund noch unbekannter Mechanismen potentiell ursächlich für den *MSH2*-Ausfall sein. Nah verwandte Gene von *DUSP11*, *MCM4*, *PRKCA*, *TRIM41* und *TRIM52* wurden bereits bei vorherigen CNV-Studien zu ungeklärtem familiärem CRC detektiert (Horpaopan et al., 2015, Masson et al. 2013, Villacis et al., 2015). Dies bestärkt den Verdacht, dass diese Gene ursächlich für die Tumorprädisposition sein könnten.

Zur weiteren differentialdiagnostischen Abklärung wurde die DNA der Patienten auf Keimbahnpunktmutationen in 30 bekannten CRC-Genen untersucht. Hierbei ließen sich 23 Varianten in zehn CRC-Genen identifizieren. Bei acht Patienten wurde eine ursächliche Keimbahnmutation detektiert. Darunter befanden sich sechs Patienten mit einer pathogenen MMR-Gen-Mutation, ein Patient mit einer *APC*-Stopp-Mutation und ein Patient mit einer vermutlich biallelen *MUTYH*-Mutation. Weiterhin wurden Keimbahnvarianten unklarer Signifikanz in den Genen *APC*, *BUB1B*, *CHEK2*, *MAP2K4*, *PPP2R1B* und *POLE* detektiert.

Ergänzend wurde nach Mutationen in den Genen des Proteasom-Pathways gesucht. Diese Gene stehen in Verdacht, bei Ausfall die Degradierung von *MSH2* herbeizuführen. Hierbei ließen sich Missense-Mutationen in *AKT1*, *HERC1*, *TSC1* und *TSC2* sowie eine synonyme Mutation in *MTOR* nachweisen. Zwei der Missense-Mutationen in *HERC1* sind aufgrund des hohen CADD-Scores von 35 von besonderer Relevanz.

Da davon auszugehen ist, dass bei einem Teil der Patienten ohne Nachweis einer MMR-Keimbahnmutation somatische Mutationen im Tumorgewebe für den MSH2-Ausfall verantwortlich sind, führten wir eine Suche nach somatischen Mutationen in *MSH2* und *MSH6* durch. Die Sequenzierung der Tumor-DNA detektierte bei zwei der elf Patienten drei heterozygote, potentiell pathogene somatische Mutationen. Bei einem dieser zwei Patienten wurden dabei zwei potentiell biallele somatische Missense-Mutationen in *MSH2* detektiert.

Insgesamt konnten 19 der ursprünglich 95 eingeschlossenen Patienten (20%) im Verlauf dieser Studie durch eine ursächliche pathogene Mutation in etablierten Genen aufgeklärt werden.

Die vorliegende Arbeit stellt die erste systematische genomweite CNV-Analyse bei Patienten mit klinischem Verdacht auf ein LS mit MSH2-Ausfall ohne nachgewiesener MMR-Keimbahnmutation dar.

Mittels CNV-Analyse und unter Einbeziehung aller nachfolgenden Analysen sowie Literatur- und Datenbankrecherchen, konnte eine Gruppe von Kandidatengenen selektiert werden, für die eine Beteiligung an der Entstehung der Erkrankung denkbar ist. Eine monogene Ursache für die Erkrankung konnte in dieser Studie nicht sicher nachgewiesen werden. In Übereinstimmung mit früheren Studien traten innerhalb unseres Kollektivs kaum rekurrent betroffene Gene auf; dies spricht für eine große Heterogenität der genetischen Faktoren. Um die kausale Relevanz der Kandidatengene als monogene, hoch-penetrante Risikofaktoren weiter abzuklären, wären unter anderem Segregationsanalysen in geeigneten Familien, der Nachweis rekurrent mutierter Gene in großen Patientenkollektiven internationaler Konsortien und funktionelle Analysen notwendig.

7 Ausblick

Die CNV-Analyse deckte eine Reihe interessanter Kandidatengene auf. Um die Relevanz dieser Gene zu bewerten, sollte nach wiederkehrenden Mutationen bei Patienten mit demselben Phänotyp gesucht werden. Hierfür wäre die Einbeziehung größerer Kollektive hilfreich, was aufgrund der Seltenheit der Erkrankung eine Herausforderung darstellt. EU-weite Kooperationen mit europaweiten Kollektiven sind ein nächstes Ziel.

Ausgehend von einem monogenen Ansatz mit einzelnen hoch-penetranten Mutationen bietet sich eine Exom- (oder Genom-)sequenzierung an, um innerhalb des Kollektivs nach pathogenen Punktmutationen zu suchen.

Um die Ursachen der gefundenen Mutationen sicher bewerten zu können, müssen in einem nächsten Schritt funktionelle Analysen folgen. Hierfür bieten sich experimentell durchgeführte Expressionsstudien, Interaktions- und Pathway-Analysen in Zellkulturen an.

Sollte eine Ursächlichkeit für einige der hier mutierten Gene bestätigt werden, so ergeben sich über die Beteiligung neuer Biosynthesewege auch weitere Therapieoptionen für die Patienten. Ein effizienterer Einsatz von Zytostatika die zielgerichtet in den betroffenen Signalweg eingreifen, wäre somit denkbar.

8 Literaturverzeichnis

- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi:10.1038/nature11632
- Aaltonen, L.A., Peltomäki, P., Leach, F.S., Sistonen, P., Pylkkänen, L., Mecklin, J.P., Järvinen, H., Powell, S.M., Jen, J., Hamilton, S.R., 1993. Clues to the pathogenesis of familial colorectal cancer. *Science* 260, 812–816.
- Aceto, G., Cristina Curia, M., Veschi, S., De Lellis, L., Mammarella, S., Catalano, T., Stuppia, L., Palka, G., Valanzano, R., Tonelli, F., Casale, V., Stigliano, V., Cetta, F., Battista, P., Mariani-Costantini, R., Cama, A., 2005. Mutations of APC and MYH in unrelated Italian patients with adenomatous polyposis coli. *Hum. Mutat.* 26, 394–394. doi:10.1002/humu.9370
- Acharya, S., Wilson, T., Gradia, S., Kane, M.F., Guerrette, S., Marsischky, G.T., Kolodner, R., Fishel, R., 1996. hMSH2 forms specific mismatch-binding complexes with hMSH3 and hMSH6. *Proc. Natl. Acad. Sci.* 93, 13629–13634.
- Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., Kirfel, J., Tafazzoli, A., Horpaopan, S., Uhlhaas, S., Stienen, D., Friedrichs, N., Altmüller, J., Laner, A., Holzapfel, S., Peters, S., Kayser, K., Thiele, H., Holinski-Feder, E., Marra, G., Kristiansen, G., Nöthen, M.M., Büttner, R., Möslein, G., Betz, R.C., Brieger, A., Lifton, R.P., Aretz, S., 2016. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *Am. J. Hum. Genet.* doi:10.1016/j.ajhg.2016.06.015
- Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., Dermitzakis, E., Bonnen, P.E., Altshuler, D.M., Gibbs, R.A., de Bakker, P.I.W., Deloukas, P., Gabriel, S.B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Yu, F., Chang, K., Hawes, A., Lewis, L.R., Ren, Y., Wheeler, D., Gibbs, R.A., Marie Muzny, D., Barnes, C., Darvishi, K., Hurles, M., Korn, J.M., Kristiansson, K., Lee, C., McCarroll, S.A., Nemes, J., Dermitzakis, E., Keinan, A., Montgomery, S.B., Pollack, S., Price, A.L., Soranzo, N., Bonnen, P.E., Gibbs, R.A., Gonzaga-Jauregui, C., Keinan, A., Price, A.L., Yu, F., Anttila, V., Brodeur, W., Daly, M.J., Leslie, S., McVean, G., Moutsianas, L., Nguyen, H., Schaffner, S.F., Zhang, Q., Ghorri, M.J.R., McGinnis, R., McLaren, W., Pollack, S., Price, A.L., Schaffner, S.F., Takeuchi, F., Grossman, S.R., Shlyakhter, I., Hostetter, E.B., Sabeti, P.C., Adebamowo, C.A., Foster, M.W., Gordon, D.R., Licinio, J., Cristina Manca, M., Marshall, P.A., Matsuda, I., Ngare, D., Ota Wang, V., Reddy, D., Rotimi, C.N., Royal, C.D., Sharp, R.R., Zeng, C., Brooks, L.D., McEwen, J.E., 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58. doi:10.1038/nature09298
- Aretz, S., Uhlhaas, S., Goergens, H., Siberg, K., Vogel, M., Pagenstecher, C., Mangold, E., Caspari, R., Propping, P., Friedl, W., 2006. MUTYH-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. *Int. J. Cancer* 119, 807–814. doi:10.1002/ijc.21905
- Assert, R., Kötter, R., Bisping, G., Scheppach, W., Stahlnecker, E., Müller, K.M., Dusel, G., Schatz, H., Pfeiffer, A., 1999. Anti-proliferative activity of protein kinase C in apical

- compartments of human colonic crypts: evidence for a less activated protein kinase C in small adenomas. *Int. J. Cancer* 80, 47–53.
- Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., Eichler, E.E., 2002. Recent segmental duplications in the human genome. *Science* 297, 1003–1007. doi:10.1126/science.1072047
- Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., Eichler, E.E., 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* 11, 1005–1017.
- Baralle, D., Baralle, M., 2005. Splicing in action: assessing disease causing sequence changes. *J. Med. Genet.* 42, 737–748. doi:10.1136/jmg.2004.029538
- Beeram, M., Patnaik, A., Rowinsky, E.K., 2005. Raf: a strategic target for therapeutic development against cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 23, 6771–6790. doi:10.1200/JCO.2005.08.036
- Bell, D.W., Varley, J.M., Szydlo, T.E., Kang, D.H., Wahrer, D.C., Shannon, K.E., Lubratovich, M., Verselis, S.J., Isselbacher, K.J., Fraumeni, J.F., Birch, J.M., Li, F.P., Garber, J.E., Haber, D.A., 1999. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science* 286, 2528–2531.
- Belmont, J.W., Boudreau, A., Leal, S.M., Hardenbol, P., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Gao, Y., others, 2005. A haplotype map of the human genome. *Nature* 437, 1299–1320.
- Bisgaard, M.L., Fenger, K., Bülow, S., Niebuhr, E., Mohr, J., 1994. Familial adenomatous polyposis (FAP): Frequency, penetrance, and mutation rate. *Hum. Mutat.* 3, 121–125. doi:10.1002/humu.1380030206
- Boland 2010 MSI in CRC.pdf, n.d.
- Boland, C.R., Goel, A., 2010. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138, 2073–2087.e3. doi:10.1053/j.gastro.2009.12.064
- Boland, C.R., Koi, M., Chang, D.K., Carethers, J.M., 2008. The biochemical basis of microsatellite instability and abnormal immunohistochemistry and clinical behavior in Lynch Syndrome: from bench to bedside. *Fam. Cancer* 7, 41–52. doi:10.1007/s10689-007-9145-9
- Bonis, P.A., Trikalinos, T.A., Chung, M., Chew, P., Ip, S., DeVine, D.A., Lau, J., 2007. Hereditary nonpolyposis colorectal cancer: diagnostic strategies and their implications. *Evid. ReportTechnology Assess.* 1–180.
- Briggs, S., Tomlinson, I., 2013. Germline and somatic polymerase ϵ and δ mutations define a new class of hypermutated colorectal and endometrial cancers: DNA polymerase mutations in colorectal and endometrial cancers. *J. Pathol.* 230, 148–153. doi:10.1002/path.4185
- Buizer-Voskamp, J.E., Muntjewerff, J.-W., Genetic Risk and Outcome in Psychosis (GROUP) Consortium Members, Strengman, E., Sabatti, C., Stefansson, H., Vorstman, J.A.S., Ophoff, R.A., 2011. Genome-wide analysis shows increased frequency of copy number variation deletions in Dutch schizophrenia patients. *Biol. Psychiatry* 70, 655–662. doi:10.1016/j.biopsych.2011.02.015
- Burmeister, D.W., Smith, E.H., Cristel, R.T., McKay, S.D., Shi, H., Arthur, G.L., Davis, J.W., Taylor, K.H., 2015. The expression of RUNDC3B is associated with promoter methylation in lymphoid malignancies. *Hematol. Oncol.* doi:10.1002/hon.2238

- Byron, A., Humphries, J.D., Craig, S.E., Knight, D., Humphries, M.J., 2012. Proteomic analysis of $\alpha 4\beta 1$ integrin adhesion complexes reveals α -subunit-dependent protein recruitment. *Proteomics* 12, 2107–2114. doi:10.1002/pmic.201100487
- Caretta, A., Mucignat-Caretta, C., 2011. Protein Kinase A in Cancer. *Cancers* 3, 913–926. doi:10.3390/cancers3010913
- Chan, T.L., Yuen, S.T., Kong, C.K., Chan, Y.W., Chan, A.S.Y., Ng, W.F., Tsui, W.Y., Lo, M.W.S., Tam, W.Y., Li, V.S.W., Leung, S.Y., 2006. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat. Genet.* 38, 1178–1183. doi:10.1038/ng1866
- Chen, D., Gould, C., Garza, R., Gao, T., Hampton, R.Y., Newton, A.C., 2007. Amplitude Control of Protein Kinase C by RINCK, a Novel E3 Ubiquitin Ligase. *J. Biol. Chem.* 282, 33776–33787. doi:10.1074/jbc.M703320200
- Chong, S., Whitelaw, E., 2004. Epigenetic germline inheritance. *Curr. Opin. Genet. Dev.* 14, 692–696. doi:10.1016/j.gde.2004.09.001
- Chuang, C.-H., Yang, D., Bai, G., Freeland, A., Pruitt, S.C., Schimenti, J.C., 2012. Post-transcriptional homeostasis and regulation of MCM2-7 in mammalian cells. *Nucleic Acids Res.* 40, 4914–4924. doi:10.1093/nar/gks176
- Chung, D.C., 2000. The genetic basis of colorectal cancer: Insights into critical pathways of tumorigenesis. *Gastroenterology* 119, 854–865. doi:10.1053/gast.2000.16507
- Clendenning, M., Buchanan, D.D., Walsh, M.D., Nagler, B., Rosty, C., Thompson, B., Spurdle, A.B., Hopper, J.L., Jenkins, M.A., Young, J.P., 2011. Mutation deep within an intron of MSH2 causes Lynch syndrome. *Fam. Cancer* 10, 297–301. doi:10.1007/s10689-011-9427-0
- Cohen, M.M., 2014. Proteus syndrome review: molecular, clinical, and pathologic features. *Clin. Genet.* 85, 111–119. doi:10.1111/cge.12266
- Colebatch, A., Hitchins, M., Williams, R., Meagher, A., Hawkins, N.J., Ward, R.L., 2006. The role of MYH and microsatellite instability in the development of sporadic colorectal cancer. *Br. J. Cancer* 95, 1239–1243. doi:10.1038/sj.bjc.6603421
- Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., Ragoussis, J., 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* 35, 2013–2025. doi:10.1093/nar/gkm076
- Conrad, D.F., Andrews, T.D., Carter, N.P., Hurler, M.E., Pritchard, J.K., 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 38, 75–81. doi:10.1038/ng1697
- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., MacArthur, D.G., MacDonald, J.R., Onyiah, I., Pang, A.W.C., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., Hurler, M.E., 2010. Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712. doi:10.1038/nature08516
- Consortium, T.E.P., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi:10.1038/nature11247
- Crino, P.B., Nathanson, K.L., Henske, E.P., 2006. The tuberous sclerosis complex. *N. Engl. J. Med.* 355, 1345–1356. doi:10.1056/NEJMra055323
- Cybulski, C., Wokołorczyk, D., Huzarski, T., Byrski, T., Gronwald, J., Górski, B., Debniak, T., Masojć, B., Jakubowska, A., Gliniewicz, B., Sikorski, A., Stawicka, M., Godlewski, D.,

- Kwias, Z., Antczak, A., Krajka, K., Lauer, W., Sosnowski, M., Sikorska-Radek, P., Bar, K., Klijer, R., Zdrojowy, R., Małkiewicz, B., Borkowski, A., Borkowski, T., Szwiec, M., Narod, S.A., Lubiński, J., 2006. A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer. *J. Med. Genet.* 43, 863–866. doi:10.1136/jmg.2006.044974
- Dardousis, K., Voolstra, C., Roengvoraphoj, M., Sekandarzad, A., Mesghenna, S., Winkler, J., Ko, Y., Hescheler, J., Sachinidis, A., 2007. Identification of Differentially Expressed Genes Involved in the Formation of Multicellular Tumor Spheroids by HT-29 Colon Carcinoma Cells. *Mol. Ther.* 15, 94–102. doi:10.1038/sj.mt.6300003
- Dauber, A., Yu, Y., Turchin, M.C., Chiang, C.W., Meng, Y.A., Demerath, E.W., Patel, S.R., Rich, S.S., Rotter, J.I., Schreiner, P.J., Wilson, J.G., Shen, Y., Wu, B.-L., Hirschhorn, J.N., 2011. Genome-wide association of copy-number variation reveals an association between short stature and the presence of low-frequency genomic deletions. *Am. J. Hum. Genet.* 89, 751–759. doi:10.1016/j.ajhg.2011.10.014
- DeAngelis, M.M., Wang, D.G., Hawkins, T.L., 1995. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* 23, 4742.
- Dellinger, A.E., Saw, S.M., Goh, L.K., Seielstad, M., Young, T.L., Li, Y.J., 2010. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res.* 38, e105–e105. doi:10.1093/nar/gkq040
- Dietlein, F., Thelen, L., Jokic, M., Jachimowicz, R.D., Ivan, L., Knittel, G., Leiser, U., van Oers, J., Edelmann, W., Heukamp, L.C., Reinhardt, H.C., 2014. A functional cancer genomics screen identifies a druggable synthetic lethal interaction between MSH3 and PRKDC. *Cancer Discov.* 4, 592–605. doi:10.1158/2159-8290.CD-13-0907
- Diouf, B., Cheng, Q., Krynetskaia, N.F., Yang, W., Cheok, M., Pei, D., Fan, Y., Cheng, C., Krynetskiy, E.Y., Geng, H., Chen, S., Thierfelder, W.E., Mullighan, C.G., Downing, J.R., Hsieh, P., Pui, C.-H., Relling, M.V., Evans, W.E., 2011. Somatic deletions of genes regulating MSH2 protein stability cause DNA mismatch repair deficiency and drug resistance in human leukemia cells. *Nat. Med.* 17, 1298–1303. doi:10.1038/nm.2430
- Elsayed, F.A., Kets, C.M., Ruano, D., van den Akker, B., Mensenkamp, A.R., Schrupf, M., Nielsen, M., Wijnen, J.T., Tops, C.M., Ligtenberg, M.J., others, 2014. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. *Eur. J. Hum. Genet.*
- Emanuel, B.S., Saitta, S.C., 2007. From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. *Nat. Rev. Genet.* 8, 869–883. doi:10.1038/nrg2136
- Epstein, D.J., 2009. Cis-regulatory mutations in human disease. *Brief. Funct. Genomic. Proteomic.* 8, 310–316. doi:10.1093/bfgp/elp021
- Fearon, E.R., 2011. Molecular Genetics of Colorectal Cancer. *Annu. Rev. Pathol. Mech. Dis.* 6, 479–507. doi:10.1146/annurev-pathol-011110-130235
- Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. doi:10.1038/nrg1767
- Fishel 1995 MMR genes role in cancer development.pdf, n.d.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., Jensen, L.J., 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi:10.1093/nar/gks1094

- Friedl, W., Aretz, S., 2005. Familial Adenomatous Polyposis: Experience from a Study of 1164 Unrelated German Polyposis Patients. *Hered. Cancer Clin. Pract.* 3, 95–114. doi:10.1186/1897-4287-3-3-95
- Galiatsatos, P., Foulkes, W.D., 2006. Familial Adenomatous Polyposis. *Am. J. Gastroenterol.* 101, 385–398. doi:10.1111/j.1572-0241.2006.00375.x
- Geurts-Giele, W.R., Leenen, C.H., Dubbink, H.J., Meijssen, I.C., Post, E., Sleddens, H.F., Kuipers, E.J., Goverde, A., van den Ouweland, A.M., van Lier, M.G., Steyerberg, E.W., van Leerdam, M.E., Wagner, A., Dinjens, W.N., 2014. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers: Somatic MMR aberrations cause Lynch syndrome-like tumours. *J. Pathol.* 234, 548–559. doi:10.1002/path.4419
- Giardiello, F.M., Allen, J.I., Axilbund, J.E., Boland, C.R., Burke, C.A., Burt, R.W., Church, J.M., Dominitz, J.A., Johnson, D.A., Kaltenbach, T., Levin, T.R., Lieberman, D.A., Robertson, D.J., Syngal, S., Rex, D.K., 2014. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on Colorectal Cancer. *Dis. Colon Rectum* 57, 1025–1048. doi:10.1097/DCR.0000000000000000
- Giardiello, F.M., Krush, A.J., Petersen, G.M., Booker, S.V., Kerr, M., Tong, L.L., Hamilton, S.R., 1994. Phenotypic variability of familial adenomatous polyposis in 11 unrelated families with identical APC gene mutation. *Gastroenterology* 106, 1542–1547.
- Gilad, Y., Bustamante, C.D., Lancet, D., Pääbo, S., 2003. Natural Selection on the Olfactory Receptor Gene Family in Humans and Chimpanzees. *Am. J. Hum. Genet.* 73, 489–501.
- Glusman, G., Yanai, I., Rubin, I., Lancet, D., 2001. The complete human olfactory subgenome. *Genome Res.* 11, 685–702. doi:10.1101/gr.171001
- Gu, W., Zhang, F., Lupski, J.R., 2008. Mechanisms for human genomic rearrangements. *PathoGenetics* 1, 4. doi:10.1186/1755-8417-1-4
- Gunderson, K.L., Steemers, F.J., Ren, H., Ng, P., Zhou, L., Tsan, C., Chang, W., Bullis, D., Musmacker, J., King, C., Lebruska, L.L., Barker, D., Oliphant, A., Kuhn, K.M., Shen, R., 2006. Whole-genome genotyping. *Methods Enzymol.* 410, 359–376. doi:10.1016/S0076-6879(06)10017-8
- Hampel, H., Frankel, W.L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Clendenning, M., Sotamaa, K., Prior, T., Westman, J.A., Panescu, J., Fix, D., Lockman, J., LaJeunesse, J., Comeras, I., de la Chapelle, A., 2008. Feasibility of Screening for Lynch Syndrome Among Patients With Colorectal Cancer. *J. Clin. Oncol.* 26, 5783–5788. doi:10.1200/JCO.2008.17.5950
- Hampel, H., Stephens, J.A., Pukkala, E., Sankila, R., Aaltonen, L.A., Mecklin, J.-P., de la Chapelle, A., 2005. Cancer Risk in Hereditary Nonpolyposis Colorectal Cancer Syndrome: Later Age of Onset. *Gastroenterology* 129, 415–421. doi:10.1053/j.gastro.2005.05.011
- Haraldsdottir, S., Hampel, H., Tomsic, J., Frankel, W.L., Pearlman, R., de la Chapelle, A., Pritchard, C.C., 2014. Colon and Endometrial Cancers With Mismatch Repair Deficiency Can Arise From Somatic, Rather Than Germline, Mutations. *Gastroenterology* 147, 1308–1316.e1. doi:10.1053/j.gastro.2014.08.041
- Hashemi, M., Moradi, N., Rezaei, M., Sanaei, S., Ziaee, S. a. M., Narouie, B., Sotoudeh, M., Bahari, G., Ghavami, S., 2016. ERBB4 gene polymorphisms and the risk of prostate cancer in a sample of Iranian Population. *Cell. Mol. Biol. Noisy-Gd. Fr.* 62, 43–48.
- Hastings, P.J., Lupski, J.R., Rosenberg, S.M., Ira, G., 2009. Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10, 551–564. doi:10.1038/nrg2593

- Hegde, M., Ferber, M., Mao, R., Samowitz, W., Ganguly, A., 2014. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). *Genet. Med.* 16, 101–116. doi:10.1038/gim.2013.166
- Hendriks, Y.M.C., Wagner, A., Morreau, H., Menko, F., Stormorken, A., Quehenberger, F., Sandkuijl, L., Møller, P., Genuardi, M., van Houwelingen, H., Tops, C., van Puijenbroek, M., Verkuijlen, P., Kenter, G., van Mil, A., Meijers-Heijboer, H., Tan, G.B., Breuning, M.H., Fodde, R., Winjen, J.T., Bröcker-Vriends, A.H.J.T., Vasen, H., 2004. Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance. *Gastroenterology* 127, 17–25. doi:10.1053/j.gastro.2004.03.068
- Hernandez-Pigeon, H., Quillet-Mary, A., Louat, T., Schambourg, A., Humbert, O., Selves, J., Salles, B., Laurent, G., Lautier, D., 2005. hMutS α is Protected from Ubiquitin-proteasome-dependent Degradation by Atypical Protein Kinase C ζ Phosphorylation. *J. Mol. Biol.* 348, 63–74. doi:10.1016/j.jmb.2005.02.001
- Hitchins, M., Williams, R., Cheong, K., Halani, N., Lin, V.A.P., Packham, D., Ku, S., Buckle, A., Hawkins, N., Burn, J., Gallinger, S., Goldblatt, J., Kirk, J., Tomlinson, I., Scott, R., Spigelman, A., Suter, C., Martin, D., Suthers, G., Ward, R., 2005. MLH1 Germline Epimutations as a Factor in Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology* 129, 1392–1399. doi:10.1053/j.gastro.2005.09.003
- Horpaopan, S., Spier, I., Zink, A.M., Altmüller, J., Holzapfel, S., Laner, A., Vogt, S., Uhlhaas, S., Heilmann, S., Stienen, D., Pasternack, S.M., Keppler, K., Adam, R., Kayser, K., Moebus, S., Draaken, M., Degenhardt, F., Engels, H., Hofmann, A., Nöthen, M.M., Steinke, V., Perez-Bouza, A., Herms, S., Holinski-Feder, E., Fröhlich, H., Thiele, H., Hoffmann, P., Aretz, S., 2015. Genome-wide CNV analysis in 221 unrelated patients and targeted high-throughput sequencing reveal novel causative candidate genes for colorectal adenomatous polyposis: CNV analysis polyposis. *Int. J. Cancer* 136, E578–E589. doi:10.1002/ijc.29215
- lafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951. doi:10.1038/ng1416
- International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature* 437, 1299–1320. doi:10.1038/nature04226
- Jacobs, P.A., Browne, C., Gregson, N., Joyce, C., White, H., 1992. Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *J. Med. Genet.* 29, 103–108.
- Jiricny, J., 2006. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* 7, 335–346. doi:10.1038/nrm1907
- Jones, S., Emmerson, P., Maynard, J., Best, J.M., Jordan, S., Williams, G.T., Sampson, J.R., Cheadle, J.P., 2002. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G: C \rightarrow T: A mutations. *Hum. Mol. Genet.* 11, 2961–2967.
- Karmakar, S., Mahajan, M.C., Schulz, V., Boyapaty, G., Weissman, S.M., 2010. A multiprotein complex necessary for both transcription and DNA replication at the β -globin locus. *EMBO J.* 29, 3260–3271. doi:10.1038/emboj.2010.204
- Kaufmann, A., Vogt, S., Uhlhaas, S., Stienen, D., Kurth, I., Hameister, H., Mangold, E., Kötting, J., Kaminsky, E., Propping, P., Friedl, W., Aretz, S., 2009. Analysis of rare APC variants at the

- mRNA level: six pathogenic mutations and literature review. *J. Mol. Diagn. JMD* 11, 131–139. doi:10.2353/jmoldx.2009.080129
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. *Genome Res.* 12, 996–1006.
- Kim, J.-Y., Jung, H.H., Do, I.-G., Bae, S., Lee, S.K., Kim, S.W., Lee, J.E., Nam, S.J., Ahn, J.S., Park, Y.H., Im, Y.-H., 2016. Prognostic value of ERBB4 expression in patients with triple negative breast cancer. *BMC Cancer* 16, 138. doi:10.1186/s12885-016-2195-3
- Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi:10.1038/ng.2892
- Klein, D., 2002. Quantification using real-time PCR technology: applications and limitations. *Trends Mol. Med.* 8, 257–260.
- Kleinjan, D.A., van Heyningen, V., 2005. Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. *Am. J. Hum. Genet.* 76, 8–32.
- Klopocki, E., Ott, C.-E., Benatar, N., Ullmann, R., Mundlos, S., Lehmann, K., 2008. A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *J. Med. Genet.* 45, 370–375. doi:10.1136/jmg.2007.055699
- Knudson, A.G., 2001. Two genetic hits (more or less) to cancer. *Nat. Rev. Cancer* 1, 157–162.
- Koese, M., Rentero, C., Kota, B.P., Hoque, M., Cairns, R., Wood, P., Vilà de Muga, S., Reverter, M., Alvarez-Guaita, A., Monastyrskaya, K., Hughes, W.E., Swarbrick, A., Tebar, F., Daly, R.J., Enrich, C., Grewal, T., 2013. Annexin A6 is a scaffold for PKC α to promote EGFR inactivation. *Oncogene* 32, 2858–2872. doi:10.1038/onc.2012.303
- Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer, A., Saunders, A.C.E., Chi, J., Yang, F., Carter, N.P., Hurler, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M., Snyder, M., 2007. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science* 318, 420–426. doi:10.1126/science.1149504
- Krepischi, A.C., Achatz, M.I., Santos, E.M., Costa, S.S., Lisboa, B.C., Brentani, H., Santos, T.M., Gonçalves, A., Nóbrega, A.F., Pearson, P.L., others, 2012. Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res* 14, R24.
- Krepischi, A.C.V., Pearson, P.L., Rosenberg, C., 2012. Germline copy number variations and cancer predisposition. *Future Oncol.* 8, 441–450. doi:10.2217/fon.12.34
- Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strömbom, L., Ståhlberg, A., Zoric, N., 2006. The real-time polymerase chain reaction. *Mol. Aspects Med.* 27, 95–125. doi:10.1016/j.mam.2005.12.007
- Kuusisto, K.M., Akinrinade, O., Vihinen, M., Kankuri-Tammilehto, M., Laasanen, S.-L., Schleutker, J., 2013. Copy Number Variation Analysis in Familial BRCA1/2-Negative Finnish Breast and Ovarian Cancer. *PLoS ONE* 8, e71802. doi:10.1371/journal.pone.0071802
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., others, 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., Ishino, F., 2002. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* 129, 1807–1817.

- Lee, J.A., Carvalho, C.M.B., Lupski, J.R., 2007. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. *Cell* 131, 1235–1247. doi:10.1016/j.cell.2007.11.037
- Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., others, 2007. The diploid genome sequence of an individual human. *PLoS Biol* 5, e254.
- Li, J.-J., Xie, D., 2015. RACK1, a versatile hub in cancer. *Oncogene* 34, 1890–1898. doi:10.1038/onc.2014.127
- Ligtenberg, M.J.L., Kuiper, R.P., Chan, T.L., Goossens, M., Hebeda, K.M., Voorendt, M., Lee, T.Y.H., Bodmer, D., Hoenselaar, E., Hendriks-Cornelissen, S.J.B., Tsui, W.Y., Kong, C.K., Brunner, H.G., van Kessel, A.G., Yuen, S.T., van Krieken, J.H.J.M., Leung, S.Y., Hoogerbrugge, N., 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat. Genet.* 41, 112–117. doi:10.1038/ng.283
- Lindhurst, M.J., Sapp, J.C., Teer, J.K., Johnston, J.J., Finn, E.M., Peters, K., Turner, J., Cannons, J.L., Bick, D., Blakemore, L., Blumhorst, C., Brockmann, K., Calder, P., Cherman, N., Deardorff, M.A., Everman, D.B., Golas, G., Greenstein, R.M., Kato, B.M., Keppler-Noreuil, K.M., Kuznetsov, S.A., Miyamoto, R.T., Newman, K., Ng, D., O'Brien, K., Rothenberg, S., Schwartzentruber, D.J., Singhal, V., Tirabosco, R., Upton, J., Wientroub, S., Zackai, E.H., Hoag, K., Whitewood-Neal, T., Robey, P.G., Schwartzberg, P.L., Darling, T.N., Tosi, L.L., Mullikin, J.C., Biesecker, L.G., 2011. A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N. Engl. J. Med.* 365, 611–619. doi:10.1056/NEJMoa1104017
- Lindor, N.M., Rabe, K., Petersen, G.M., Haile, R., Casey, G., Baron, J., Gallinger, S., Bapat, B., Aronson, M., Hopper, J., Jass, J., LeMarchand, L., Grove, J., Potter, J., Newcomb, P., Terdiman, J.P., Conrad, P., Moslein, G., Goldberg, R., Ziogas, A., Anton-Culver, H., de Andrade, M., Siegmund, K., Thibodeau, S.N., Boardman, L.A., Seminara, D., 2005. Lower Cancer Incidence in Amsterdam-I Criteria Families Without Mismatch Repair Deficiency: Familial Colorectal Cancer Type X. *JAMA* 293, 1979. doi:10.1001/jama.293.16.1979
- Livak, K.J., Schmittgen, T.D., 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262
- Lucito, R., Suresh, S., Walter, K., Pandey, A., Lakshmi, B., Krasnitz, A., Sebat, J., Wigler, M., Klein, A.P., Brune, K., Palmisano, E., Maitra, A., Goggins, M., Hruban, R.H., 2007. Copy-number variants in patients with a strong family history of pancreatic cancer. *Cancer Biol. Ther.* 6, 1592–1599. doi:10.4161/cbt.6.10.4725
- Lupski, J.R., Stankiewicz, P., 2005. Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes. *PLoS Genet.* 1, e49. doi:10.1371/journal.pgen.0010049
- Lynch, H., Lynch, P., Lanspa, S., Snyder, C., Lynch, J., Boland, C., 2009. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.* 76, 1–18. doi:10.1111/j.1399-0004.2009.01230.x
- Lynch, H.T., de la Chapelle, A., 2003. Hereditary colorectal cancer. *N. Engl. J. Med.* 348, 919–932. doi:10.1056/NEJMra012242
- Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D., Hitchins, M.P., 2015. Milestones of Lynch syndrome: 1895–2015. *Nat. Rev. Cancer* 15, 181–194.

- MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 42, D986–D992. doi:10.1093/nar/gkt958
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. *Nature* 461, 747–753. doi:10.1038/nature08494
- Mardis, E.R., 2008. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* 9, 387–402. doi:10.1146/annurev.genom.9.081307.164359
- Marenne, G., Rodríguez-Santiago, B., Closas, M.G., Pérez-Jurado, L., Rothman, N., Rico, D., Pita, G., Pisano, D.G., Kogevinas, M., Silverman, D.T., Valencia, A., Real, F.X., Chanock, S.J., Génin, E., Malats, N., 2011. Assessment of Copy Number Variation Using the Illumina Infinium 1M SNP-Array: A Comparison of Methodological Approaches in the Spanish Bladder Cancer/EPICURO Study. *Hum. Mutat.* 32, 240–248. doi:10.1002/humu.21398
- Masson, A., Talseth-Palmer, B., Evans, T.-J., Grice, D., Duesing, K., Hannan, G., Scott, R., 2013. Copy Number Variation in Hereditary Non-Polyposis Colorectal Cancer. *Genes* 4, 536–555. doi:10.3390/genes4040536
- McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemes, J., Wysoker, A., Shapero, M.H., de Bakker, P.I.W., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., Altshuler, D., 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* 40, 1166–1174. doi:10.1038/ng.238
- McPhillips, M., Meldrum, C.J., Creegan, R., Edkins, E., Scott, R.J., 2005. Deletion Mutations in an Australian Series of HNPCC Patients. *Hered. Cancer Clin. Pract.* 3, 43–47. doi:10.1186/1897-4287-3-1-43
- Medlock, M.M., Tester, D.J., Will, M.L., Bos, J.M., Ackerman, M.J., 2012. Repeat long QT syndrome genetic testing of phenotype-positive cases: prevalence and etiology of detection misses. *Heart Rhythm Off. J. Heart Rhythm Soc.* 9, 1977–1982. doi:10.1016/j.hrthm.2012.08.010
- Mensenkamp, A.R., Vogelaar, I.P., van Zelst-Stams, W.A.G., Goossens, M., Ouchene, H., Hendriks-Cornelissen, S.J.B., Kwint, M.P., Hoogerbrugge, N., Nagtegaal, I.D., Ligtenberg, M.J.L., 2014. Somatic Mutations in MLH1 and MSH2 Are a Frequent Cause of Mismatch-Repair Deficiency in Lynch Syndrome-Like Tumors. *Gastroenterology* 146, 643–646.e8. doi:10.1053/j.gastro.2013.12.002
- Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16, 1215.
- Morak, M., Heidenreich, B., Keller, G., Hampel, H., Laner, A., de la Chapelle, A., Holinski-Feder, E., 2014. Biallelic MUTYH mutations can mimic Lynch syndrome. *Eur. J. Hum. Genet.*
- Morak, M., Laner, A., Bacher, U., Keiling, C., Holinski-Feder, E., 2010. MUTYH-associated polyposis - variability of the clinical phenotype in patients with biallelic and monoallelic MUTYH mutations and report on novel mutations. *Clin. Genet.* 78, 353–363. doi:10.1111/j.1399-0004.2010.01478.x

- Mullis, K.B., Faloona, F.A., 1987. [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, in: Ray Wu (Ed.), *Methods in Enzymology*. Academic Press, pp. 335–350.
- Nagasaka, T., Rhees, J., Kloor, M., Gebert, J., Naomoto, Y., Boland, C.R., Goel, A., 2010. Somatic Hypermethylation of MSH2 Is a Frequent Event in Lynch Syndrome Colorectal Cancers. *Cancer Res.* 70, 3098–3108. doi:10.1158/0008-5472.CAN-09-3290
- Obermair, A., Youlden, D.R., Young, J.P., Lindor, N.M., Baron, J.A., Newcomb, P., Parry, S., Hopper, J.L., Haile, R., Jenkins, M.A., 2010. Risk of endometrial cancer for women diagnosed with HNPCC-related colorectal carcinoma. *Int. J. Cancer* 127, 2678–2684. doi:10.1002/ijc.25501
- Palles, C., Cazier, J.-B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Almeida, E.G., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L.G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M.B., Humphray, S., Thomas, H.J.W., Maher, E., Evans, G., Lucassen, A., Cummings, C., Stevens, M., Walker, L., Halliday, D., Armstrong, R., Paterson, J., Hodgson, S., Homfray, T., Side, L., Izatt, L., Donaldson, A., Tomkins, S., Morrison, P., Goodman, S., Brewer, C., Henderson, A., Davidson, R., Murday, V., Cook, J., Haites, N., Bishop, T., Sheridan, E., Green, A., Marks, C., Carpenter, S., Broughton, M., Greenhalge, L., Suri, M., Donnelly, P., Bell, J., Bentley, D., McVean, G., Ratcliffe, P., Taylor, J., Wilkie, A., Donnelly, P., Broxholme, J., Buck, D., Cazier, J.-B., Cornall, R., Gregory, L., Knight, J., Lunter, G., McVean, G., Taylor, J., Tomlinson, I., Wilkie, A., Buck, D., Gregory, L., Humphray, S., Kingsbury, Z., McVean, G., Donnelly, P., Cazier, J.-B., Broxholme, J., Grocock, R., Hatton, E., Holmes, C.C., Hughes, L., Humburg, P., Kanapin, A., Lunter, G., Murray, L., Rimmer, A., Lucassen, A., Holmes, C.C., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylance, R., Sawyer, E.J., Kerr, D.J., Clark, S., Grimes, J., Kearsley, S.E., Thomas, H.J.W., McVean, G., Houlston, R.S., Tomlinson, I., 2012. Germline mutations affecting the proofreading domains of PLEKHA7 and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* 45, 136–144. doi:10.1038/ng.2503
- Palles, C., Cazier, J.-B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Guarino Almeida, E., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L.G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M.B., Humphray, S., CORGI Consortium, WGS500 Consortium, Lucassen, A., Holmes, C.C., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylance, R., Sawyer, E.J., Kerr, D.J., Clark, S., Grimes, J., Kearsley, S.E., Thomas, H.J.W., McVean, G., Houlston, R.S., Tomlinson, I., 2013. Germline mutations affecting the proofreading domains of PLEKHA7 and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* 45, 136–144. doi:10.1038/ng.2503
- Pang, A.W.C., MacDonald, J.R., Yuen, R.K.C., Hayes, V.M., Scherer, S.W., 2014. Performance of High-Throughput Sequencing for the Discovery of Genetic Variation Across the Complete Size Spectrum. *G3* 4, 63–65. doi:10.1534/g3.113.008797
- Peltomäki, P., Lothe, R.A., Aaltonen, L.A., Pylkkänen, L., Nyström-Lahti, M., Seruca, R., David, L., Holm, R., Ryberg, D., Haugen, A., 1993. Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. *Cancer Res.* 53, 5853–5855.

- Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., Goldstein, D.B., 2013. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *PLoS Genet.* 9, e1003709. doi:10.1371/journal.pgen.1003709
- Pickard, B.S., Malloy, M.P., Christoforou, A., Thomson, P.A., Evans, K.L., Morris, S.W., Hampson, M., Porteous, D.J., Blackwood, D.H.R., Muir, W.J., 2006. Cytogenetic and genetic evidence supports a role for the kainate-type glutamate receptor gene, GRIK4, in schizophrenia and bipolar disorder. *Mol. Psychiatry* 11, 847–857. doi:10.1038/sj.mp.4001867
- Pino, M.S., Mino-Kenudson, M., Wildemore, B.M., Ganguly, A., Batten, J., Sperduti, I., Iafrate, A.J., Chung, D.C., 2009. Deficient DNA Mismatch Repair Is Common in Lynch Syndrome-Associated Colorectal Adenomas. *J. Mol. Diagn.* 11, 238–247. doi:10.2353/jmoldx.2009.080142
- Plazzer, J.P., Sijmons, R.H., Woods, M.O., Peltomäki, P., Thompson, B., Den Dunnen, J.T., Macrae, F., 2013. The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. *Fam. Cancer* 12, 175–180. doi:10.1007/s10689-013-9616-0
- Plon, S.E., Eccles, D.M., Easton, D., Foulkes, W.D., Genuardi, M., Greenblatt, M.S., Hogervorst, F.B.L., Hoogerbrugge, N., Spurdle, A.B., Tavtigian, S.V., for the IARC Unclassified Genetic Variants Working Group, 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum. Mutat.* 29, 1282–1291. doi:10.1002/humu.20880
- Prenzel, N., Fischer, O.M., Streit, S., Hart, S., Ullrich, A., 2001. The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. *Endocr. Relat. Cancer* 8, 11–31.
- Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35, D61–D65. doi:10.1093/nar/gkl842
- Raguz, S., De Bella, M.T., Slade, M.J., Higgins, C.F., Coombes, R.C., Yagüe, E., 2005. Expression of RPIP9 (Rap2 interacting protein 9) is activated in breast carcinoma and correlates with a poor prognosis. *Int. J. Cancer J. Int. Cancer* 117, 934–941. doi:10.1002/ijc.21252
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurler, M.E., 2006. Global variation in copy number in the human genome. *Nature* 444, 444–454. doi:10.1038/nature05329
- Rhees, J., Arnold, M., Boland, C.R., 2014. Inversion of exons 1–7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. *Fam. Cancer* 13, 219–225. doi:10.1007/s10689-013-9688-x
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative Genomics Viewer. *Nat. Biotechnol.* 29, 24–26. doi:10.1038/nbt.1754
- Rodríguez-Soler, M., Pérez-Carbonell, L., Guarinos, C., Zapater, P., Castillejo, A., Barberá, V.M., Juárez, M., Bessa, X., Xicola, R.M., Clofent, J., Bujanda, L., Balaguer, F., Reñé, J., de-Castro, L., Marín-Gabriel, J.C., Lanás, A., Cubiella, J., Nicolás-Pérez, D., Brea-Fernández, A., Castellví-Bel, S., Alenda, C., Ruiz-Ponte, C., Carracedo, A., Castells, A., Andreu, M.,

- Llor, X., Soto, J.L., Payá, A., Jover, R., 2013. Risk of Cancer in Cases of Suspected Lynch Syndrome Without Germline Mutation. *Gastroenterology* 144, 926–932.e1. doi:10.1053/j.gastro.2013.01.044
- Rolland, D., Basrur, V., Conlon, K., Wolfe, T., Fermin, D., Nesvizhskii, A.I., Lim, M.S., Elenitoba-Johnson, K.S.J., 2014. Global phosphoproteomic profiling reveals distinct signatures in B-cell non-Hodgkin lymphomas. *Am. J. Pathol.* 184, 1331–1342. doi:10.1016/j.ajpath.2014.01.036
- Rozen, S., Skaletsky, H., 1999. Primer3 on the WWW for general users and for biologist programmers, in: *Bioinformatics Methods and Protocols*. Springer, pp. 365–386.
- Ruffalo, M., LaFramboise, T., Koyutürk, M., 2011. Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics* 27, 2790–2796. doi:10.1093/bioinformatics/btr477
- Rustgi, A.K., 2007. The genetics of hereditary colon cancer. *Genes Dev.* 21, 2525–2538. doi:10.1101/gad.1593107
- Saglam, O., Xiong, Y., Marchion, D.C., Strosberg, C., Wenham, R.M., Johnson, J.J., Saeed-Vafa, D., Cubitt, C., Hakam, A., Magliocco, A.M., 2017. ERBB4 Expression in Ovarian Serous Carcinoma Resistant to Platinum-Based Therapy. *Cancer Control J. Moffitt Cancer Cent.* 24, 89–95.
- Sampson, J.R., Jones, N., 2009. MUTYH-associated polyposis. *Best Pract. Res. Clin. Gastroenterol.* 23, 209–218. doi:10.1016/j.bpg.2009.03.006
- Sanger, F., Nicklen, S., Coulson, A.R., 1992. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnol. Read. Mass* 24, 104–108.
- Schmermund, A., Möhlenkamp, S., Stang, A., Grönemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J., Jöckel, K.-H., Erbel, R., 2002. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: Rationale and design of the Heinz Nixdorf RECALL Study. *Am. Heart J.* 144, 212–218. doi:10.1067/mhj.2002.123579
- Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwiijnenburg, D., Diepvens, F., Pals, G., 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30, e57.
- Sebat, J., 2004. Large-Scale Copy Number Polymorphism in the Human Genome. *Science* 305, 525–528. doi:10.1126/science.1098918
- Shima, N., Alcaraz, A., Liachko, I., Buske, T.R., Andrews, C.A., Munroe, R.J., Hartford, S.A., Tye, B.K., Schimenti, J.C., 2007. A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat. Genet.* 39, 93–98. doi:10.1038/ng1936
- Shlien, A., Malkin, D., 2010. Copy number variations and cancer susceptibility. *Curr. Opin. Oncol.* 22, 55–63. doi:10.1097/CCO.0b013e3283333dca4
- Sieber, O.M., Lipton, L., Crabtree, M., Heinemann, K., Fidalgo, P., Phillips, R.K., Bisgaard, M.-L., Orntoft, T.F., Aaltonen, L.A., Hodgson, S.V., others, 2003. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N. Engl. J. Med.* 348, 791–799.
- Simbolo, M., Mafficini, A., Agostini, M., Pedrazzani, C., Bedin, C., Urso, E.D., Nitti, D., Turri, G., Scardoni, M., Fassan, M., Scarpa, A., 2015. Next-generation sequencing for genetic testing of familial colorectal cancer syndromes. *Hered. Cancer Clin. Pract.* 13. doi:10.1186/s13053-015-0039-9

- Spier, I., Holzapfel, S., Altmüller, J., Zhao, B., Horpaopan, S., Vogt, S., Chen, S., Morak, M., Raeder, S., Kayser, K., Stienen, D., Adam, R., Nürnberg, P., Plotz, G., Holinski-Feder, E., Lifton, R.P., Thiele, H., Hoffmann, P., Steinke, V., Aretz, S., 2015. Frequency and phenotypic spectrum of germline mutations in *POLE* and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas: *POLE* Mutations in Polyposis. *Int. J. Cancer* 137, 320–331. doi:10.1002/ijc.29396
- Stankiewicz, P., Lupski, J.R., 2010. Structural Variation in the Human Genome and its Role in Disease. *Annu. Rev. Med.* 61, 437–455. doi:10.1146/annurev-med-100708-204735
- Stankiewicz, P., Lupski, J.R., 2002. Genome architecture, rearrangements and genomic disorders. *TRENDS Genet.* 18, 74–82.
- Steemers, F.J., Chang, W., Lee, G., Barker, D.L., Shen, R., Gunderson, K.L., 2006. Whole-genome genotyping with the single-base extension assay. *Nat. Methods* 3, 31–33. doi:10.1038/nmeth842
- Steemers, F.J., Gunderson, K.L., 2007. Whole genome genotyping technologies on the BeadArray platform. *Biotechnol. J.* 2, 41–49. doi:10.1002/biot.200600213
- Steinke, V., Engel, C., Büttner, R., Schackert, H.K., Schmiegel, W.H., Propping, P., 2013. Hereditary nonpolyposis colorectal cancer (HNPCC)/Lynch syndrome. *Dtsch. Ärztebl. Int.* 110, 32.
- Steinke, V., Vogt, S., Aretz, S., 2010. Klinik und Genetik des familiären Darmkrebses. *Med. Genet.* 22, 265–281. doi:10.1007/s11825-010-0226-z
- Subauste, M.C., Ventura-Holman, T., Du, L., Subauste, J.S., Chan, S.-L., Yu, V.C., Maher, J.F., 2009. RACK1 downregulates levels of the pro-apoptotic protein Fem1b in apoptosis-resistant colon cancer cells. *Cancer Biol. Ther.* 8, 2297–2305.
- Suzuki, K., Oneyama, C., Kimura, H., Tajima, S., Okada, M., 2011. Down-regulation of the tumor suppressor C-terminal Src kinase (Csk)-binding protein (Cbp)/PAG1 is mediated by epigenetic histone modifications via the mitogen-activated protein kinase (MAPK)/phosphatidylinositol 3-kinase (PI3K) pathway. *J. Biol. Chem.* 286, 15698–15706. doi:10.1074/jbc.M110.195362
- Svaasand, E.K., 2015. A Novel Deep Intronic Mutation Introducing a Cryptic Exon Causing Neurofibromatosis Type 1 in a Family with Highly Variable Phenotypes: A Case Study. *Hered. Genet.* 04. doi:10.4172/2161-1041.1000152
- Takei, Y., Assenberg, M., Tsujimoto, G., Laskey, R., 2002. The MCM3 Acetylase MCM3AP Inhibits Initiation, but Not Elongation, of DNA Replication via Interaction with MCM3. *J. Biol. Chem.* 277, 43121–43125. doi:10.1074/jbc.C200442200
- Talseth-Palmer, B.A., Holliday, E.G., Evans, T.-J., McEvoy, M., Attia, J., Grice, D.M., Masson, A.L., Meldrum, C., Spigelman, A., Scott, R.J., 2013. Continuing difficulties in interpreting CNV data: lessons from a genome-wide CNV association study of Australian HNPCC/lynch syndrome patients. *BMC Med. Genomics* 6, 10. doi:10.1186/1755-8794-6-10
- Tautz, D., Schlötterer, null, 1994. Simple sequences. *Curr. Opin. Genet. Dev.* 4, 832–837.
- Thompson, B.A., Spurdle, A.B., Plazzer, J.-P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capellá, G., den Dunnen, J.T., du Sart, D., Fabre, A., Farrell, M.P., Farrington, S.M., Frayling, I.M., Frebourg, T., Goldgar, D.E., Heinen, C.D., Holinski-Feder, E., Kohonen-Corish, M., Robinson, K.L., Leung, S.Y., Martins, A., Moller, P., Morak, M., Nystrom, M., Peltomaki, P., Pineda, M., Qi, M., Ramesar, R., Rasmussen, L.J., Royer-Pokora, B., Scott, R.J., Sijmons, R., Tavtigian, S.V., Tops, C.M., Weber, T., Wijnen, J., Woods, M.O., Macrae, F., Genuardi, M., Castillejo, A., Sexton, A., Chan, A.K.W., Viel, A., Blanco, A., French, A., Laner, A., Wagner, A., van den Ouweland, A., Mensenkamp, A.,

- Payá, A., Betz, B., Redeker, B., Smith, B., Espenschied, C., Cummings, C., Engel, C., Fornes, C., Valenzuela, C., Alenda, C., Buchanan, D., Barana, D., Konstantinova, D., Cairns, D., Glaser, E., Silva, F., Laloo, F., Crucianelli, F., Hogervorst, F., Casey, G., Tomlinson, I., Blanco, I., Villar, I.L., Garcia-Planells, J., Bigler, J., Shia, J., Martinez-Lopez, J., Gille, J.J.P., Hopper, J., Potter, J., Soto, J.L., Kantelinen, J., Ellis, K., Mann, K., Varesco, L., Zhang, L., Le Marchand, L., Marafie, M.J., Nordling, M., Tibiletti, M.G., Kahan, M.A., Ligtenberg, M., Clendenning, M., Jenkins, M., Speevak, M., Digweed, M., Kloor, M., Hitchins, M., Myers, M., Aronson, M., Valentin, M.D., Kutsche, M., Parsons, M., Walsh, M., Kansikas, M., Zahary, M.N., Pedroni, M., Heider, N., Poplawski, N., Rahner, N., Lindor, N.M., Sala, P., Nan, P., Propping, P., Newcomb, P., Sarin, R., Haile, R., Hofstra, R., Ward, R., Tricarico, R., Bacares, R., Young, S., Chialina, S., Kovalenko, S., Gunawardena, S.R., Moreno, S., Ho, S.L., Yuen, S.T., Thibodeau, S.N., Gallinger, S., Burnett, T., Teitsch, T., Chan, T.L., Smyrk, T., Cranston, T., Psofaki, V., Steinke-Lange, V., Barbera, V.-M., 2013. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.* 46, 107–115. doi:10.1038/ng.2854
- Toribara, N.W., Sleisenger, M.H., 1995. Screening for colorectal cancer. *N. Engl. J. Med.* 332, 861–867.
- Umar, A., Boland, C.R., Terdiman, J.P., Syngal, S., de la Chapelle, A., Rüschoff, J., Fishel, R., Lindor, N.M., Burgart, L.J., Hamelin, R., Hamilton, S.R., Hiatt, R.A., Jass, J., Lindblom, A., Lynch, H.T., Peltomaki, P., Ramsey, S.D., Rodriguez-Bigas, M.A., Vasen, H.F.A., Hawk, E.T., Barrett, J.C., Freedman, A.N., Srivastava, S., 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J. Natl. Cancer Inst.* 96, 261–268.
- Valle, L., Hernandez-Illan, E., Bellido, F., Aiza, G., Castillejo, A., Castillejo, M.-I., Navarro, M., Segui, N., Vargas, G., Guarinos, C., Juarez, M., Sanjuan, X., Iglesias, S., Alenda, C., Egoavil, C., Segura, A., Juan, M.-J., Rodriguez-Soler, M., Brunet, J., Gonzalez, S., Jover, R., Lazaro, C., Capella, G., Pineda, M., Soto, J.L., Blanco, I., 2014. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. *Hum. Mol. Genet.* 23, 3506–3512. doi:10.1093/hmg/ddu058
- Valle, L., Perea, J., Carbonell, P., Fernandez, V., Dotor, A.M., Benitez, J., Urioste, M., 2007. Clinicopathologic and Pedigree Differences in Amsterdam I-Positive Hereditary Nonpolyposis Colorectal Cancer Families According to Tumor Microsatellite Instability Status. *J. Clin. Oncol.* 25, 781–786. doi:10.1200/JCO.2006.06.9781
- Vasen, H.F., Watson, P., Mecklin, J.-P., Lynch, H.T., others, 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* 116, 1453–1456.
- Venkatachalam, R., Verwiel, E.T.P., Kamping, E.J., Hoenselaar, E., Görgens, H., Schackert, H.K., van Krieken, J.H.J.M., Ligtenberg, M.J.L., Hoogerbrugge, N., van Kessel, A.G., Kuiper, R.P., 2011. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. *Int. J. Cancer* 129, 1635–1642. doi:10.1002/ijc.25821
- Vilar, E., Gruber, S.B., 2010. Microsatellite instability in colorectal cancer—the stable evidence. *Nat. Rev. Clin. Oncol.* 7, 153–162. doi:10.1038/nrclinonc.2009.237
- Villacis, R.A.R., Miranda, P.M., Gomy, I., Santos, E.M.M., Carraro, D.M., Achatz, M.I., Rossi, B.M., Rogatto, S.R., 2015. Contribution of rare germline copy number variations and common

- susceptibility loci in Lynch Syndrome patients negative for mutations in the mismatch repair genes. *Int. J. Cancer* *J. Int. Cancer*. doi:10.1002/ijc.29948
- Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., Bos, J.L., 1988. Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* 319, 525–532. doi:10.1056/NEJM198809013190901
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Kinzler, K.W., 2013. Cancer Genome Landscapes. *Science* 339, 1546–1558. doi:10.1126/science.1235122
- von Eyss, B., Maaskola, J., Memczak, S., Möllmann, K., Schuetz, A., Loddenkemper, C., Tanh, M.-D., Otto, A., Muegge, K., Heinemann, U., Rajewsky, N., Ziebold, U., 2012. The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation. *EMBO J.* 31, 972–985. doi:10.1038/emboj.2011.451
- Wagner, A., van der Klift, H., Franken, P., Wijnen, J., Breukel, C., Bezrookove, V., Smits, R., Kinarsky, Y., Barrows, A., Franklin, B., Lynch, J., Lynch, H., Fodde, R., 2002. A 10-Mb paracentric inversion of chromosome arm 2p inactivates MSH2 and is responsible for hereditary nonpolyposis colorectal cancer in a North-American kindred. *Genes Chromosomes Cancer* 35, 49–57. doi:10.1002/gcc.10094
- Walsh, T., Casadei, S., Coats, K.H., Swisher, E., Stray, S.M., Higgins, J., Roach, K.C., Mandell, J., Lee, M.K., Ciernikova, S., Foretova, L., Soucek, P., King, M.-C., 2006. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA* 295, 1379–1388. doi:10.1001/jama.295.12.1379
- Watson, C.T., Marques-Bonet, T., Sharp, A.J., Mefford, H.C., 2014. The genetics of microdeletion and microduplication syndromes: an update. *Annu. Rev. Genomics Hum. Genet.* 15, 215–244. doi:10.1146/annurev-genom-091212-153408
- Watson, P., Vasen, H.F.A., Mecklin, J.-P., Bernstein, I., Aarnio, M., Järvinen, H.J., Myrholm, T., Sundé, L., Wijnen, J.T., Lynch, H.T., 2008. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *Int. J. Cancer* 123, 444–449. doi:10.1002/ijc.23508
- Weterings, E., van Gent, D.C., 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair* 3, 1425–1435. doi:10.1016/j.dnarep.2004.06.003
- Willis, J.A., Mukherjee, S., Orlow, I., Viale, A., Offit, K., Kurtz, R.C., Olson, S.H., Klein, R.J., 2014. Genome-wide analysis of the role of copy-number variation in pancreatic cancer risk. *Front. Genet.* 5. doi:10.3389/fgene.2014.00029
- Winchester, L., Yau, C., Ragoussis, J., 2009. Comparing CNV detection methods for SNP arrays. *Brief. Funct. Genomic. Proteomic.* 8, 353–366. doi:10.1093/bfgp/elp017
- Yang, R., Chen, B., Pfitze, K., Buch, S., Steinke, V., Holinski-Feder, E., Stocker, S., von Schonfels, W., Becker, T., Schackert, H.K., Royer-Pokora, B., Kloor, M., Schmiegel, W.H., Buttner, R., Engel, C., Lascorz Puertolas, J., Forsti, A., Kunkel, N., Bugert, P., Schreiber, S., Krawczak, M., Schafmayer, C., Propping, P., Hampe, J., Hemminki, K., Burwinkel, B., 2014. Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3. *Carcinogenesis* 35, 315–323. doi:10.1093/carcin/bgt344
- Yang, X.R., Brown, K., Landi, M.T., Ghiorzo, P., Badenas, C., Xu, M., Hayward, N.K., Calista, D., Landi, G., Bruno, W., Bianchi-Scarrà, G., Aguilera, P., Puig, S., Goldstein, A.M., Tucker, M.A., 2012. Duplication of CXC chemokine genes on chromosome 4q13 in a melanoma-prone family: 4q13 duplication in a melanoma family. *Pigment Cell Melanoma Res.* 25, 243–247. doi:10.1111/j.1755-148X.2012.00969.x
- Yu, W., Wang, Y., Xie, Z., You, J., Wang, J., Cui, X., Pei, F., Zheng, J., 2010. [Screening of phosphoprotein associated with glycosphingolipid microdomains 1 (PAG1) by cDNA

- microarray and influence of overexpression of PAG1 on biologic behavior of human metastatic prostatic cancer cell line in vitro]. *Zhonghua Bing Li Xue Za Zhi* 39, 88–94.
- Zarrei, M., MacDonald, J.R., Merico, D., Scherer, S.W., 2015. A copy number variation map of the human genome. *Nat. Rev. Genet.* 16, 172–183. doi:10.1038/nrg3871
- Zerres, K., Schüler, H., Gembruch, U., Bald, R., Hansmann, M., Schwanitz, G., 1992. Chromosomal findings in fetuses with prenatally diagnosed cysts of the choroid plexus. *Hum. Genet.* 89, 301–304.
- Zozulya, S., Echeverri, F., Nguyen, T., 2001. The human olfactory receptor repertoire. *Genome Biol.* 2, RESEARCH0018.

9 Eigene Publikationen

Kayser K., Degenhardt F., Holzapfel S., Horpaopan S., Peters S., Spier I., Morak M., Vangala D., Rahner N., von Knebel-Doerberitz M., Schakert H. K., Engel C., Büttner R., Wijnen J., Doerks T., Moebus S., Herms S., Hofmann A., Fischer S., Hoffmann P., Aretz S., Steinke-Lange V.: ***Copy number variation analysis and targeted NGS in suspected Lynch syndrome families reveals novel potential causative candidate genes.*** submitted 2017. I. J. Cancer

Adam R., Spier I., Zhao B., Kloth M., Marquez J., Hinrichsen I., Tafazzoli A., Horpaopan S., Uhlhaas S., Kirfel J., Stienen D., Friedrichs N., Altmüller J., Laner A., Holzapfel S., Peters S., Kayser K., Thiele H., Holinski-Feder E., Kristiansen G., Nöthen M.M., Büttner R., Möslein G., Betz R., Brieger A., Lifton R.P., Aretz S.: ***Exome sequencing identifies biallelic MSH3 germline mutations as a novel recessive subtype of colorectal adenomatous polyposis.*** Am. J. Hum. Genet. 2016;99:337-51

Horpaopan S., Spier I., Zink A.M., Altmüller J., Holzapfel S., Laner A., Vogt S., Uhlhaas S., Heilmann S., Stienen D., Pasternack S., Keppler K., Adam R., Kayser K., Moebus S., Draaken M., Degenhardt F., Engels H., Hofmann A., Nöthen M.M., Steinke-Lange V., Herms S., Holinski-Feder E., Fröhlich H., Thiele H., Hoffmann P., Aretz S.: ***Genome-wide CNV analysis in 221 unrelated patients and targeted high-throughput sequencing reveal novel causative candidate genes for colorectal adenomatous polyposis.*** Int. J. Cancer 136; 2015 E578-E589. doi:10.1002/ijc.29215

Spier I., Holzapfel S., Altmüller J., Zhao B., Horpaopan S., Vogt S., Chen S., Morak M., Raeder S., Kayser K., Stienen D., Adam R., Nürnberg P., Plotz G., Holinski-Feder E., Lifton R.P., Thiele H., Hoffmann P., Steinke V., Aretz S.: ***Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas: POLE Mutations in Polyposis.*** Int. J. Cancer 137, 2015, 320-331. doi:10.1002/ijc.29396

Anhang

A) Primer zur Validierung der CNVs mittels qPCR

Tab. A.1 Deletionen

Primer-Name	Primer Sequence 5'-3'	Primer Tm [°C]	Länge des Amplikons (bp)
AGMO_1F	TGCCAACATCAGAGCAAAC	58.01	100
AGMO_1R	CTTGTCTCCCTAGGTCACCG	59.18	
AGMO_2F	ACCTGCACTTTGATTGGTTCA	58.33	135
AGMO_2R	GCTCCTTACTTTACAGGAAGCC	59.31	
AGMO_3F	GGAGATGAGGCTGTACCAAAG	58.73	181
AGMO_3R	CATAGGGTTCATCATGGTAAGGA	57.43	
CLDN23_1F	CCAGTGGACGTGGAGTTGTA	59.6	146
CLDN23_1R	AGCGAGGTGACCATGAGTG	59.8	
CLDN23_2F	TGGATGGAAATCTGCCTTTC	60	142
CLDN23_2R	TATGAAAAGCCCGTCACTCC	60.1	
CSMD1_1F	CAAGCCTCACCAGAGAGACC	59.99	103
CSMD1_1R	AGAGTGAGGCATGAGGCTGT	60.02	
CSMD1_2F	GCGGCAGTTTCTAAGGATCA	60.35	123
CSMD1_2R	TTCAACATCTCACCAATGC	59.5	
CSMD1_3F	GTTGAGGCAGACTGGAGGAG	59.99	167
CSMD1_3R	CTCCATTCCCAGTCTTGAA	60.04	
DLG2_1F	AGCACCATCTGTTATGCTC	58.8	137
DLG2_1R	TTCAGGGAAGCTGTTTTGCT	60	
DLG2_2F	TTGGGTAAAATGGTCCAGAGA	59.4	149
DLG2_2R	CAAGTTTGCTGTCTTCAGGGTA	59.4	
DLG2_3F	TCCTAACGAGTCTGAAAAAGTGC	59.9	129
DLG2_3R	ATGGGAGGGAGTAAGGGAGA	59.9	
FBXO36_1F	CCAATCACATGAGGCCAGGA	59.74	127
FBXO36_1R	CCTCCCTGGTAGCTGGGATT	60.7	
FBXO36_2F	GATGGTGGAAGATCTCTAAGGA	60.1	114
FBXO36_2R	TGTTCCGTTCTTACCTTGAA	59.6	
FBXO36_3F	GACTTCCTGAACGGCTCTCA	60	115

FBXO36_3R	TGACCGTTACCTTTGCAAATCTG	59.75	
KLHDC3_1F	CCAGGACTGTGACCCAAGTT	60	147
KLHDC3_1R	CACCAAGACCTGGGAACAAG	60.5	
KLHDC3_2F	TGGCTGCCTCTACACTTCT	60	126
KLHDC3_2R	AGTCCAGCTCCATGGTTGTC	60.1	
KLHDC3_3F	GCACCCTGAAAGGGTACTCA	60.1	135
KLHDC3_3R	CCATGCACTGCAGATGATTT	59.1	
KLHL32_1F	AAGCAGTGTGAGCACAGCATA	59.7	115
KLHL32_1R	TGACTCATAAGCTCATCCCTTTC	59.7	
KLHL32_2F	GGATTTACTGGCCATGCTGT	60	111
KLHL32_2R	CTCGCATGCATTTTTCCAGT	61.2	
KLHL32_3F	TCTGTTAGCTGGGCTCTTGC	60.7	150
KLHL32_3R	ATAGATGCCCAGTCCAGGTG	60	
LINCO1473_1F	CTTCCTCTCACCACCCTCAG	59.8	110
LINCO1473_1R	ACAAGAACCACGTGTGGAAA	59	
LINCO1473_2F	TTTTCTAACCTTGCTGAAGATG	59.8	120
LINCO1473_2R	TTGAGCTTCATGGTGATGGA	60.2	
LINCO1473_3F	GCATCCACACTGGACATCAG	60.1	142
LINCO1473_3R	AAGTGCATGAAATGCACAG	60	
NMNAT3_1F	TATCCACCCATTGGCATTTC	60	137
NMNAT3_1R	ATACATGAAGGGCACCAAGG	59.8	
NMNAT3_2F	CTGAAGTGCACCCTGGGAAA	60.18	186
NMNAT3_2R	CAGTCTGTTTGTAAGGCATGCA	60	
NOSIP_PRRG2_4F	CGCTGTAGGAGCACTGTTTG	59.7	123
NOSIP_PRRG2_4R	GTGCGGGAGACCAAATAAAC	59.7	
NOSIP_PRRG2_5F	TTAGCTAGGGTGGTGGCTGT	59.8	130
NOSIP_PRRG2_5R	CTTGGCCTCAGTCAGAGCTT	59.7	
NOSIP_PRRG2_6F	CTCCGAGACCCAGAAGTTTG	59.8	136
NOSIP_PRRG2_6R	ACATCCACCCACAGACAACA	59.8	
NRG3_1F	CAAAGGAATACATTCTCCCTCA	58.2	125

Anhang

NRG3_1R	CTGCATTACCACAGCCTGAA	59.9	
NRG3_2F	GTGTGAGAAGTGGGCATGTG	60.2	112
NRG3_2R	ACACAGAAAGGGCAAAGCAG	60.4	
NRG3_3F	TTGTTTTGAAGAAGCTTGATGTGC	59.3	110
NRG3_3R	CCTTTCCTGGGGTTAGAAT	59.3	
PAG1_1F	AAAGGCCAGCCTCTTCTTTC	60	128
PAG1_1R	AGTGCTGTTGGGGTATCCTG	60	
PAG1_2F	TGTGCTGTACCACTGCCATT	60.2	127
PAG1_2R	CAAAGGGAATACATTGCCAGA	59.9	
PARK2_1F	TCGCATAGTGGAAAATGAACA	59.2	138
PARK2_1R	TTTTGCCATTCTTCCTTCTGA	59.8	
PARK2_2F	TGAGCGTATGGTGAGAAGGA	59.4	110
PARK2_2R	GAGTGGGGTAAGCACGTTTC	59.6	
PARK2_3F	CTTGTGCCAAAGTGGTCAAC	59.2	117
PARK2_3R	GGGCATGGCTTAAAGAAGGT	60.5	
PITPNC1_1F	GCAACTTGGTGTGAGGAACA	59.7	128
PITPNC1_1R	TACCCCCAGGATGAAAATCA	60.1	
PITPNC1_2F	AAGGGTTTGGACGTTGAGTG	60	122
PITPNC1_2R	CAAGGCCACAGAAACCAGTC	60.7	
PRKCA_1F	ATTTCCACCTCATCCACCAG	59.8	141
PRKCA_1R	ACCTCCCATCAGAATGTGGT	59.2	
PRKCA_2F	TAAGCAACCAGGCAAGATCC	60.2	146
PRKCA_2R	AGCAGCAGAAAGAAGCCTGA	60.4	
PRKCA_3F	CAAGCCACTTGCTGATGAGA	60.1	150
PRKCA_3R	AGAAACACACCCCACTGACC	59.9	
PRKDC_MCM4_1F	GCCAGCACTGCCATTTTAGT	60.3	121
PRKDC_MCM4_1R	AGCGAGCAACCAGAAACAC	60.4	
PRKDC_MCM4_2F	GAACACCAGGCTCTGTCCAT	60.1	144
PRKDC_MCM4_2R	AATCTTGACCCAGGTCATC	60.2	
PRKDC_MCM4_3F	TGCCTGTTCCAAATGCTAT	60.5	155

PRKDC_MCM4_3R	TTGATTACCTCCCAAGTCG	59.9	
RAP1GDS1_TSPAN5_1F	TTACAGATGAGGGCAGAAGTG	58	102
RAP1GDS1_TSPAN5_1R	TCTCATTGGCGGGATCTGTT	62.9	
RAP1GDS1_TSPAN5_2F	TGGTTGAAGCAGGCCTAGTA	58.5	160
RAP1GDS1_TSPAN5_2R	CCACATTCTTTGGCTGGCAT	63.3	
RAP1GDS1_TSPAN5_3F	GGCTCTCCAGGGCATTATT	62.5	134
RAP1GDS1_TSPAN5_3R	GTGTGCCCCAGTTTGAGAAG	60.7	
RUNDC3B_ABCB1_4F	GATTTCCTCAAAGTTGCAGATG	59.6	113
RUNDC3B_ABCB1_4R	CCGAAAGTCAATTCCATGCT	60.1	
RUNDC3B_ABCB1_5F	TCTGCAGCATTGAAAATATGG	58.8	119
RUNDC3B_ABCB1_5R	CACACCTTCTGCTATTCCAA	60.1	
RUNDC3B_ABCB1_6F	AGGGTAGAGCCTGGATCAGAG	59.8	155
RUNDC3B_ABCB1_6R	TTCTGATTTCTGAGTCTTCAGGTG	59.9	
SRBD1_1F	CAAACATTGTGCCAAGAGA	59.7	133
SRBD1_1R	ACCCCTCTGGTACCTCAGTTT	59	
SRBD1_2F	CCTGTTGTTACCAAAACACA	59.5	131
SRBD1_2R	TTCTTGGTGAAGTTAGGAACCA	58.7	
STARD13_1F	AAGTCACAGCATGCCCTTCT	59.9	140
STARD13_1R	GGTCAAAGCATTCTGCTTCA	59	
STARD13_2F	CACATGCCCATCTCAGTAATTT	59	132
STARD13_2R	TGTCTAATTGAAATCCCTCACAC	58.1	
TP63_1F	CTGCCTTCTCCATCCATC	60.6	150
TP63_1R	AAGAAGCAACATTTCCATTG	59.2	
TP63_2F	TCCATTGTAACATGCCCTCA	59.9	123
TP63_2R	CATCCAGGTGTTCTCAGTGG	59.1	
VWDE_1F	TTGGACACCAATGCAGTCAT	60	111
VWDE_1R	TGGGTGCTTTGTTTTGATG	59.6	
VWDE_2F	GGCGGTCAAATCAAAGAAG	59.7	147
VWDE_2R	TTTGCAGGAGAAATGAAGGAA	59.8	
VWDE_3F	TCGGTGAAAGTGTCTGGA	59.3	142

VWDE_3R	CCACCAGCTTATTTCTCATGC	59.7	
WFDC9-WFDC11_1F	GCAGAAGCATCACAACCTCCA	60	114
WFDC9-WFDC11_1R	TTGGGAAAAGACTTGCCTTC	59.3	
WFDC9-WFDC11_2F	TCGGCAAGAACACAAGACTG	60	131
WFDC9-WFDC11_2R	AAGTCGGAGCAGAGCAAGAG	59.9	
WFDC9-WFDC11_3F	GCCAGTGAATTGCTTTCCAA	61.1	150
WFDC9-WFDC11_3R	GGAGATGTTTGCTTGGGTGT	60	
ZNF415_1F	CCCTTCAGCTTGAACTGCT	59.6	140
ZNF415_1R	CTACGGCCCCAAAAGAAAAT	60.3	
ZNF415_2F	CCCACCTTTTGGTTTTCTT	60.2	165
ZNF415_2R	GCAAAGGGCTGAGATTCTTG	60	
ZNF415_3F	TGCCAGGAGTCTTTGGAAGT	59.8	115
ZNF415_3R	CGAGCTCCTCCCTGTGTATG	60.8	
ZNF551_ZNF776_1F	TACCTTGTGGCCAGTTTCT	59.6	121
ZNF551_ZNF776_1R	TGTAAAAAGAGCTCCATAGTTCACA	59.4	
ZNF551_ZNF776_2F	AGGCAATAAGAGGCAGTTGG	59.3	120
ZNF551_ZNF776_2R	GCTGGTGACCTGAAAAGAGG	59.8	
ZNF551_ZNF776_3F	GACTCCGGCGTCTCTACT	60.8	111
ZNF551_ZNF776_3R	TCCACCTTCTGGGTTTCAGTC	60.1	

Tab. A.2 Duplikationen

Primer-Name	Primer Sequence 5'-3'	Primer Tm [°C]	Länge des Amplikons (bp)
ENOX1_1F	TCTGAACATCAATCACCACCA	60	113
ENOX1_1R	TCCGTGAGGACAGAGACCTT	59.8	
ENOX1_2F	TCTTTGCCCTTGGTCACTCT	59.8	147
ENOX1_2R	TCACTGGGTTACAAGGAGAAGTC	59.7	
ENOX1_3F	TCCTGGTGTTGGCAAATAGA	59.1	149
ENOX1_3R	TGTGGAGCTGAGTCGTGTGT	60.5	
KATNAL1_1F	ATCCTTGATCCCAGCCAAC	59.9	116
KATNAL1_1R	CACCCCTTCTCCTTACC	59.9	

KATNAL1_2F	GGCCTGAATGTGATGACAGA	59.6	110
KATNAL1_2R	TTTCCTGTTCCACCGTGAT	60.4	
LCLAT1_CAPN13_1F	GGTCTGGGATCTTGATAATGGA	60.2	130
LCLAT1_CAPN13_1R	CAAGGTAAGGAAAGAAGTGCAGA	59.9	
LCLAT1_CAPN13_2F	TCACTGTGGCGTATCCTCAC	59.7	148
LCLAT1_CAPN13_2R	TTGTGGCACCAGAGTTGAAG	59.9	
LCLAT1_CAPN13_3F	CCCATTCCAAAGATGCTGTT	59.9	119
LCLAT1_CAPN13_3R	TGGGGTAGGGTCATCCATTA	60	
MYH6_MYH7_1F	TCTCCAGCTCATGCACATTC	59.9	150
MYH6_MYH7_1R	GAGGCCAGAGATGAGTTGGT	59.3	
MYH6_MYH7_2F	AACATGCACTCCTCCTCCAG	60.3	115
MYH6_MYH7_2R	GTCAGGGTATGGGACTGTGG	60.2	
MYH6_MYH7_3F	AACTCCAAAGGGGCTGAACT	60.1	127
MYH6_MYH7_3R	GCTGGGGAAGGTTATGTGTG	60.4	
OCLN_GTF2H2C_1F	GCTAGGACAACCTGCTGCACC	59.9	136
OCLN_GTF2H2C_1R	CCAATGGCCACCTAAATGAG	60.3	
OCLN_GTF2H2C_2F	TTTTAGGCCATTGCCATTGT	60.3	133
OCLN_GTF2H2C_2R	TGTGTTCTTGTCCACAAA	60	
PDZRN1_1F	GACCTGCGGACATGTGATAA	59.5	126
PDZRN1_1R	GGCAGCACAGACTTCAACAA	60	
PDZRN4_2F	TCTTTTGACCTCCGGGAGTA	59.7	148
PDZRN4_2R	GCTGGGTGCTAAGGAGGATT	60.6	
PDZRN4_3F	GAAAAGGGTCAGAAGAGCAAAA	59.9	130
PDZRN4_3R	GCTCTTCTTGCAGCAACAT	60.5	
RBM19_1F	ATTTTCCTTTTGTGGCATCG	59.9	144
RBM19_1R	CAGTGCCTCAGTTTCCATT	60.1	
RBM19_2F	ATGGGGAGGTGACTCTGTGT	59.4	139
RBM19_2R	GCTTTGTGGACTTCTCACC	59.7	
RBM19_3F	GGTCTTGGCCAAAAACACAT	59.8	131
RBM19_3R	GTTTTTAGGAAGGGGGCAAG	59.9	

SLC04C1_1F	CAAACCAGTGAGGTGTGGAA	59.6	149
SLC04C1_1R	TTGGGGTCTCTTTTGAAGG	59.1	
SLC04C1_2F	GACAGTGACCGGAGCTTCTC	60	141
SLC04C1_2R	CTCTGCCTTGCCTCTGACC	60	
TPRKB_DUSP11_1F	GGACTGGCAAACCTTTTGA	60.1	127
TPRKB_DUSP11_1R	CATTGAAGAGGGAGAAAAACAAA	59.6	
TPRKB_DUSP11_2F	CCAACAGGCTGGATAGGAGA	60.2	130
TPRKB_DUSP11_2R	TCCCCTGGAGAGGACTATT	59.9	
TPRKB_DUSP11_3F	TGAGGAGCGTCCTGAAAAGT	60	147
TPRKB_DUSP11_3R	GGGTCTTATCCTGGCATTGA	59.9	
TRIM41_GNB2L1_TRIM52_1F	AAGCTGGGGCAGGATTTAGT	60.1	123
TRIM41_GNB2L1_TRIM52_1R	GGCGAGCAATAATGGAAAAA	60	
TRIM41_GNB2L1_TRIM52_2F	GCTTGCACTTAGCCAGGTTC	60	110
TRIM41_GNB2L1_TRIM52_2R	AGAAGTGCCAAGCCCTTTTT	60.2	
TRIM41_GNB2L1_TRIM52_3F	AAGCACATGCCCTGATCATT	60.5	119
TRIM41_GNB2L1_TRIM52_3R	AGCTTTACACGTCGCAGCTT	60.2	
TSPAN9_1F	GGAGGGATCTCAGTTGTCA	60	113
TSPAN9_1R	CAGCATTTTGGCATTCTGA	59.8	
TSPAN9_2F	CTTCCTTCCAGCTCTGTGG	60	144
TSPAN9_2R	CATGACAATGGTGCCTATGG	59.8	
TSPAN9_3F	TCATCCGTCTGTTGCAGAG	60	110
TSPAN9_3R	AGACTCCGCCCAATGTAGA	59.7	

B) Primer der Routinediagnostik für MSH2 und MSH6

Tab. B.1 *MSH2*

Primer-Name	Primer Sequence 5'-3'	Primer Tm [°C]	Länge des Amplikons (bp)
MSH2_1F	TCGCGCATTTTCTTCAACC	62.2	285
MSH2_1R	GTCCCTCCCCAGCACGC	65.4	
MSH2_2F	GAAGTCCAGCTAATACAGTGC	54.3	286
MSH2_2R	CTTCACATTTTTATTTTTCTACTC	51.6	
MSH2_3F	GGTTCATAGAGTTTGGATTTTCC	59.3	427
MSH2_3R	AAAGAGCCTTTCCTAGGCCTGG	64.2	
MSH2_4F	TAATGTAGGTGAATCTGTTATCAC	52.9	346
MSH2_4R	CTTCTAAAAAGTCACTATAGTTTAC	48.4	
MSH2_5F	TTTTTAAAATCTTTAGAACTGGATC	54.7	266
MSH2_5R	CCATTCAACATTTTTAACCCCTT	57.1	
MSH2_6F	GTTTTCACTAATGAGCTTGCC	56.7	251
MSH2_6R	GTGGTATAATCATGTGGGTAAC	53.5	
MSH2_7F	GAGACTTACGTGCTTAGTTG	50.3	328
MSH2_7R	GTATATATTGTATGAGTTGAAGG	48.0	
MSH2_8F	CCTTTTGGATCAAATGATGCTTG	62.8	281
MSH2_8R	AACTTTCTTAAAGTGGCCTTTGC	60.2	
MSH2_9F	CCCATTATTTATAGGATTTTGTAC	57.2	210
MSH2_9R	GTATAGACAAAAGAATTATTCCAAC	52.7	
MSH2_10F	TAGAATTACATTGAAAAATGGTAG	52.6	278
MSH2_10R	CATGTTAGAGCATTTAGGGAATT	56.6	
MSH2_10F	TAGAATTACATTGAAAAATGGTAG	52.6	278
MSH2_10R	CATGTTAGAGCATTTAGGGAATT	56.6	
MSH2_11F	TGGATATGTTTCACGTAGTACAC	54.4	220
MSH2_11R	AGCCAGGTGACATTCAGAAC	57.7	
MSH2_12F	ATTCAGTATTCCTGTGTACATT	50.6	327
MSH2_12R	CGTTACCCCCACAAAGCC	61.3	
MSH2_13F	CGCGATTAATCATCAGTG	53.4	353

MSH2_13R	GGACAGAGACATACATTTCTATC	52.2	
MSH2_14F	CATTTTATGTGATGGGAAATTC	57.5	347
MSH2_14R	GGGTAGTAAGTTTCCCATTAC	52.2	
MSH2_15F	CTCTTCTCATGCTGTCCC	53.9	263
MSH2_15R	TAATAGAGAAGCTAAGTTAAAC	45.4	
MSH2_16F	GTCAGTGTCTAACATGACTTTTAG	52.7	293
MSH2_16R	ACTATTACAGACAATAGCTTATC	47.2	

Tab. B.2 *MSH6*

Primer-Name	Primer Sequence 5'-3'	Primer T _m [°C]	Länge des Amplikons (bp)
MSH6_1F	CCTGTTGATTGGCCACTGGG	66.0	488
MSH6_1R	AACCCCTGTGCGAGCCTC	67.1	
MSH6_2F	CTGCCTTTAAGGAACTTGACC	59.3	330
MSH6_2R	GCCTGTCTGTCTGTTTCTCTC	56.7	
MSH6_3F	CTGCACCCGGCCCTTATTGT	66.4	270
MSH6_3R	CCCTTCTTCCCCATCACC	65.0	
MSH6_4aF	GTCAAAAAATCATAAGTTGAACTG	55.0	420
MSH6_4aR	GAGGGCGTTTCCTCCTAG	58.9	
MSH6_4bF	CCTGAACAGCCCTGTCAAAG	60.8	391
MSH6_4bR	CTGAGACTTAATCTGCCACC	54.9	
MSH6_4cF	GATCACCCCGATTTTGATGC	62.5	406
MSH6_4cR	GTGTACCCTTGGTAATGATCC	55.9	
MSH6_4dF	CGAGTGGAACAGACTGAGAC	55.8	425
MSH6_4dR	GGTATCAGACCTTCTGAAG	53.8	
MSH6_4eF	CGAGATTTAGGACTCTAGTGG	53.4	424
MSH6_4eR	CTTGTAGTGCTGACTGTGTC	51.5	
MSH6_4fF	CTACCTCAAAAAATGCCTTATTG	57.2	401
MSH6_4fR	TGAGTAGCCTCTCAAGATCTG	54.8	
MSH6_4gF	CGTCTAGATGCCATAGAAGACC	58.0	415
MSH6_4gR	GAGTAATAAGTCCAGTCTTTTCG	52.7	

MSH6_4hF	CCTGAAGGTCGTTTTCTGA	60.2	590
MSH6_4h_int_F	CACTCGCAATTTGCCAGAAG	61.9	324
MSH6_4hR	GTCAAAATATGTACCATATATACC	49.5	
MSH6_5F	AAAACCCCAAACGATGAAGC	64.4	364
MSH6_5R	TCCTAATGTCACAAATGACTTTC	56.0	
MSH6_6F	CGTAAGGGTTCATAAGAAAGAC	54.9	246
MSH6_6R	CATCTAAATGACTGAATGAGAAC	53.3	
MSH6_7F	CGGCCAATAATTGCATAGTCTC	60.8	269
MSH6_7R	AGATAGTCTTCAAATGAGAAGTT	51.1	
MSH6_8F	CCTTTGAGTTACTTCCTTATGC	55.0	244
MSH6_8R	AGTGCCCTCTCAAAAACCGA	63.2	
MSH6_9F	TCGGTTTTTTGAGAGGGCAC	62.4	331
MSH6_9R	CCCCTTTTACTGTTTCTTTG	54.2	
MSH6_10F	GGGAAGGGATGATGCACTATG	61.6	262
MSH6_10R	CCTAGAAAGAAAATGGAAAATGG	59.4	

C) Primer zur Untersuchung selten beschriebener Ursachen für den MSH2-Ausfall

Tab. C.1 *MSH2*-Intron

Primer-Name	Primer Sequence 5'-3'	Primer T _m [°C]	Länge des Amplikons (bp)
MSH2_Intron_F	GCGGAAAGGACTTACCTTGCTTG	64.4	761bp
MSH2_Intron_mitte	CACTGTATTAGCTGGACTTC	49.9	
MSH2_Intron_R	CTTCACATTTTATTTTCTACTC	51.6	

Tab. C.2 *MSH2* Promotor

Primer-Name	Primer Sequence 5'-3'	Primer T _m [°C]	Länge des Amplikons (bp)
MSH2_Promotor_F	GCTCTACTAAGGATGCGCGTCTG	64.6	532bp
MSH2_Promotor_R	GTCCCTCCCAGCACGC	65.4	

D) Primer der Proteasom-Pathway-Analyse

Tab. D.1 *PRKCZ*

Primer-Name	Primer Sequence 5'-3'	Primer Tm [°C]	Länge des Amplikons (bp)
PRKCZ_1F	GAGTTCCGCGGAGTTGAC	59.3	283
PRKCZ_1R	CGACGGCTGAGCCCTC	61.8	
PRKCZ_2F	CAAATATGCCCCACGGTAAC	60.1	na
PRKCZ_2R	AGGCAGTTCAAGACCAGAGC	59.6	
PRKCZ_3F	GCTGCAGGTGTACTGACTTCC	59.9	385
PRKCZ_3R	AGACACCCACAGCCAGGA	60.3	
PRKCZ_4F	TGTCCAGGTCCCTCTGATTG	61.1	535
PRKCZ_4R	CCCTGGACTTCTCATGCTACC	61.0	
PRKCZ_5F	AGGCCGTGACACCATGAG	60.7	455
PRKCZ_5R	CCCTAGTCAAAGCCTCATCC	58.7	
PRKCZ_6F	CCCAACCTGGGATAGACC	59.2	
PRKCZ_6R	AGAGCTAGGACGGCACCAC	60.4	457
PRKCZ_7F	GAGGTTGAAGCTGCAGTGAG	58.8	
PRKCZ_7R	GCCAGGCTTGAAGACAATG	59.4	
PRKCZ_8F	TCAGAGTCGGCAGGTGTG	59.5	392
PRKCZ_8R	AGACATCCACCCGTGAAG	61.0	
PRKCZ_9F	GGATATGTGGGACGCAGAAC	60.3	509
PRKCZ_9R	TTCCAGAACACATCCTGGTC	58.5	
PRKCZ_10F	GGATGGTGACAATGGTGGTA	59.1	411
PRKCZ_10R	AAGAGAGTCTGCTGGCCTGA	60.3	
PRKCZ_11F	CACTTCCCAAGTCCACACAC	59.0	462
PRKCZ_11R	CACCCGTCATCCCTCTGT	59.4	
PRKCZ_12F	GCCTGGTCATTGAGAGGATG	60.6	404
PRKCZ_12R	AGTGCTCGTTGTGTACCAG	59.9	
PRKCZ_13F	ACAGGGACCTGAAGCTGGAC	61.6	428
PRKCZ_13R	GCTGAGGCCAGATGATGAC	59.3	
PRKCZ_14F	TGGTTCTGTTGGGAAGTGG	60.9	439

PRKCZ_14R	GCATGTGACACAGGGTCTG	58.6	
PRKCZ_15F	CAGCGCATGTAACCAGGAG	60.4	396
PRKCZ_15R	AACCCACACACCCGATACAC	60.6	
PRKCZ_16F	AAATGGAACGGAGCTTAACG	59.2	410
PRKCZ_16R	GGGTTGGGATGGATATTGG	59.8	
PRKCZ_17F	CACTGGCAGTCAAATACTAGGC	60.2	447
PRKCZ_17R	ACCGTGTGTTTCATCCCTTG	60.8	

Tab. D.2 *MTOR*

Primer-Name	Primer Sequence 5'-3'	Primer T _m [°C]	Länge des Amplikons (bp)
MTOR_2F	TTCTCAATCAGATCTCTCAACC	59.7	493
MTOR_2R	GCAACTGCTGCAAAGAACC	61.0	
MTOR_3F	AATCAGATGATTGCAAGTTCAG	57.0	488
MTOR_3R	TGGTCCGAGACTCCACTTC	60.2	
MTOR_4F	GGCTGGTCTCAAACCTCTGA	60.4	n/a
MTOR_4R	CCATCTCACAGCCTGGAATC	60.6	
MTOR_5F	TTGCTTTCCTTCTCTGTCC	59.4	524
MTOR_5R	ACAAAGCAAGACTCCATCAC	55.3	
MTOR_6F	ACTGGCACTAGGCACTGAGC	60.6	418
MTOR_6R	TCTTCATAGAACTTGCTGCTATCTG	59.3	
MTOR_7F	GCCACCCTCACTTTGGAAG	61.9	600
MTOR_7R	ACCTGTGAAGGCAGAAGGTC	59.3	
MTOR_8F	GGGTACAGCTCTCACC AAGG	59.7	486
MTOR_8R	GTCTGCTGCCTCTGACATTAG	57.7	
MTOR_9F	GAACCTCTGGCCTCAAACAC	59.7	n/a
MTOR_9R	CATATGATTTGCTATACCCTGGAG	59.0	
MTOR_10F	ATGGGTCCCTTTGAACTGTG	59.8	420
MTOR_10R	GCCTGTGTGTTTAGTCTAAGCTATC	57.9	
MTOR_11F	GGAGACCTGAGACTAAGCCAAG	59.5	619
MTOR_11R	CAAACCTCAGTATGGGACAAAGAAAC	60.3	

MTOR_12F	CCTGCAGTGAGTGGAGATTG	59.4	713
MTOR_12R	GCGTAGTGGGAAGATTCCAG	59.7	
MTOR_13F	CTGGAATCTTCCCACTACGC	59.7	440
MTOR_13R	ACAAATGCGGCTTTTACCAG	60.1	
MTOR_14F	TAAGGCTGTTCTGCCTGTTG	59.1	447
MTOR_14R	TGATATGGCCCTTGGTCTC	59.9	
MTOR_15F	CTTGGATTCATAAAGCCTAGAAGG	59.7	417
MTOR_15R	ATCTGGAGAGTTGGACATGTTTT	59.0	
MTOR_16F	ATGTTGGCTAGGCTGGTCTG	60.3	368
MTOR_16R	TGTCACACTCTGTTCTAGGCACT	59.1	
MTOR_17F	GATTGTTGCAGGGTTGTGTG	60.0	523
MTOR_17R	GTATCACGGATACTCCTCAGG	59.1	
MTOR_18F	GGCAAGAGAGAGATGGTTGC	60.0	491
MTOR_18R	TCCGGGATTCAAGAAGAAAC	59.1	
MTOR_19F	TGGTGAGAGGCTATTCATTGG	60.1	n/a
MTOR_19R	TGTGAGCCACTGCACCTG	60.6	
MTOR_20F	AGCCTCCAAACCTAGCCAAG	60.8	445
MTOR_20R	TTGTACCTGGGAACCTGGAG	60.0	
MTOR_21F	GAGGAGGGGAGTGAGGAGTT	59.7	419
MTOR_21R	TATGGAATGGGCATCAACCT	60.2	
MTOR_22F	AGTTTTCCAGTCCCCAGAG	60.5	400
MTOR_22R	TTCCAATGGGATAGGGACAG	59.7	
MTOR_23F	TCTTAACTGTCCCTATCCCATTG	59.4	515
MTOR_23R	GGCACTTAGCTCACATAGGTTG	58.9	
MTOR_24F	GGCCACCATCATAAACCAG	58.8	492
MTOR_24R	TGTGTATGGCTGCTCAGG	60.0	
MTOR_25F	GCTAGTTGCATGGTTGCTTG	59.5	632
MTOR_25R	TCCATATCCAAGGGAGCAC	59.9	
MTOR_26F	TGTGGTATGTTCCAGAGTTGC	58.6	529
MTOR_26R	GGAATTTGTTGCTACTAGCAAGG	60.4	

MTOR_27F	AATTCTCTTGACAGCACTGG	78.1	n/a
MTOR_27R	CAGCTCTCTACCCAGCAG	59.9	
MTOR_28F	ATTCAGTTCCTTGCCACAC	60.0	572
MTOR_28R	CAGCTCTCTACCCAGCAG	59.9	
MTOR_29F	GGGAGGAGAGCACAACCTC	61.2	346
MTOR_29R	TGAAGGTCAGGGCCAATAAC	59.9	
MTOR_30F	ATCCCAGGATCATTCCATCC	60.9	604
MTOR_30R	AAGGCTGTCTGTCTAGAGC	60.0	
MTOR_31F	AGGCATGGTTCCTATCCTTG	59.0	491
MTOR_31R	AGAAAGGACAGATGGAAGGTG	58.2	
MTOR_32F	ACGTCAACCTGCATAGTGTG	57.2	472
MTOR_32R	CTGACCTGCTGGCAGTCTTC	61.2	
MTOR_33F+34F	TCTCAAGCCATTCAAGTCTGG	60.4	698
MTOR_33R+34R	GCAAGCCAAGCAATACTG	59.5	
MTOR_35F	GCCAAGATTGCACCACTG	58.7	n/a
MTOR_35R	GGTCACCTGAGGGTGAAGT	60.6	
MTOR_36F	CTTATGGTGGGTCCTTG	61.4	850
MTOR_36R	CCAGGAGCCATCTATTTC	60.8	
MTOR_37F	AGCCCAGTCTCTCCAGTATC	59.7	490
MTOR_37R	TGCTGAACCTAGGAAAGCAAG	59.6	
MTOR_38F	AAGCCCATCTCCTTCTTGC	60.7	804
MTOR_38R	CCCTCCATCAATCACATCTTT	58.8	
MTOR_39F	CAGCCATTTCTGAGTGTCTCC	59.9	564
MTOR_39R	AAGCCCTATCTCAGGCAGTG	59.5	
MTOR_40F	TGATACAGTTGCCATTTATTGAGTC	59.5	454
MTOR_40R	GGCCCAGTCAGCTGTACTC	59.9	
MTOR_41F	GGAGGCCAAACACATTTGAG	60.5	424
MTOR_41R	GGTTCATTCCCTTCCCTTAG	58.9	
MTOR_42F	GGTCCCAGTGTCAAGTGCC	60.4	471
MTOR_42R	GGCTCAGTCTTCGAGGGAAC	61.3	

MTOR_43F	GAAAGCGTTCCTCGAAGAC	61.3	477
MTOR_43R	CCTGAAGCCAAGAGAAGAAG	56.9	
MTOR_44F	CATAGACCCTGGTCCCTTAGC	60.0	499
MTOR_44R	CAGTGGAAGGCCAGAGG	59.3	
MTOR_45F	GTTTGAAGCTCCAGGGTTG	61.9	507
MTOR_45R	GGATTTAGTGTCTGCCTCCAG	60.1	
MTOR_46F	GGGAATGTCAAGGACCTCAC	59.4	753
MTOR_46R	AGTCTCCAACACCTGCCTTC	59.3	
MTOR_47F	AGAACAGAGGAGGACAGAGTGG	59.9	636
MTOR_47R	ACTGCAGCCTTGAACCTCTG	60.6	
MTOR_48F	CCAAGATTGTGCCACTGC	58.7	n/a
MTOR_48R	TTTCTCCATATGGCCAGTGC	61.0	
MTOR_49F	CTGCACAGAAGGCCTATAAC	55.2	490
MTOR_49R	CGCTCTACAGCCAATCACAG	59.6	
MTOR_50F	AATAAGCTGGGCATGGTG	77.2	n/a
MTOR_50R	CCAGGATGACAGGCATGAG	60.2	
MTOR_51F	CAGTGGGACAACCTGTGATTC	60.4	644
MTOR_51R	TACTCAGAGGGCATGGGAAC	60.1	
MTOR_52F	ACCCTCCATACCCTGTTG	59.7	636
MTOR_52R	CCCTCACTTAGGAAGCAGACAC	60.3	
MTOR_53F	TCCAAATGGGATCAGGACAG	60.9	651
MTOR_53R	GCGTGAGAACAGACTAATACATGC	60.2	
MTOR_54F	TCACATCCCATGCTAATACCC	59.7	406
MTOR_54R	GATGCACCCACTGACTGAAG	59.3	
MTOR_55F	GTAGGGCAGGCGTTAAAGG	59.7	359
MTOR_55R	CCTTCCTCTATGTCCGTCTTTG	60.1	
MTOR_56F	CAAAGACGGACATAGAGGAAGG	60.1	316
MTOR_56R	GCCAAAGCTCGTCACTAACAC	59.9	
MTOR_57F	TCCTTCATCTTCATGTCTGACG	60.3	388
MTOR_57R	ATAACAATGGGCACATGCAG	59.4	

MTOR_58F	GTGGGCATTAGCCTTGACTG	60.7	558
MTOR_58R	GGAGCAACTAGGTCATTCTTCC	59.2	

Tab. D.3 *PIK3C2B*

Primer-Name	Primer Sequence 5'-3'	Primer T _m [°C]	Länge des Amplikons (bp)
PIK3C2B_1F	CCTCCATCTTCTGTACATTCC	59.8	559
PIK3C2B_1R	GCCCTTGGGAGGTAGAGTG	59.7	
PIK3C2B_2F	GAGAACAGAGCCAAGCAG	54.4	606
PIK3C2B_2R	CTCCGCATCATAGGTCGATT	60.1	
PIK3C2B_3F	ACCTCCAAGATCTCCCAGC	59.2	529
PIK3C2B_3R	GGTCCCAAAGAGGGTAGGAG	59.9	
PIK3C2B_4F	GCTTTGGCCACACTTTGTC	60.8	581
PIK3C2B_4R	TGATCAGGATACTTTGACCATT	58.1	
PIK3C2B_5F	GGGTTTCTCAGAGCTGCTACC	60.4	534
PIK3C2B_5R	AACCAGATAAAGCAGGCAAG	57.1	
PIK3C2B_6F	CTCGATTTCCACCTCATCTTG	59.7	616
PIK3C2B_6R	TCTTTCCTAATCCCAGCTC	59.4	
PIK3C2B_7F	CAATATTTATACCGCCAAGG	58.4	573
PIK3C2B_7R	GCTGATGGTCTGCTTGACAG	59.6	
PIK3C2B_8F	CCGCAAGTTTGACATTGAC	57.7	499
PIK3C2B_8R	GGAGAGAATGATGGCAAAGTTC	60.1	
PIK3C2B_9F	TCTTCCCTCTGGAACCTTGC	59.4	487
PIK3C2B_9R	AGCAGAATCCCTGACTACGC	59.5	
PIK3C2B_10F	GGAACGGCTGAAAGAATTG	58.3	602
PIK3C2B_10R	CAAGCAGTCCCAGGTAGCTC	60.0	
PIK3C2B_11F	CCGAGTAGCCTGAGGGTTG	60.8	479
PIK3C2B_11R	TGTGCTCTCCAGCTGTATCC	59.0	
PIK3C2B_12F	TCCCTGGGATTCTTATGGTG	59.7	885
PIK3C2B_12R	CCAGCACCAAGAGGGAATAAC	60.9	
PIK3C2B_13F	CCATCCCATTCTAGCCAATG	60.3	1596

PIK3C2B_13R	CAGTTAGGGTTAACAAGGAGCCT	60.1	
PIK3C2B_14F	GGCTGTATCCTCAGCCTCAG	60.0	491
PIK3C2B_14R	TCCTTAACATGGTCCTGTGC	58.6	
PIK3C2B_15F	CCAGACCAGGGTGTGGTTAG	60.4	544
PIK3C2B_15R	CAATTTAACCTCCAGGAGATGG	59.8	
PIK3C2B_16F	CATCTCCACCGAACCTCTTC	59.7	554
PIK3C2B_16R	TTCCTAGTGTGAGCTGATGGAG	59.5	
PIK3C2B_17F	AGCCATTTGGACTTAACTGG	56.8	648
PIK3C2B_17R	TGGTAGGCACATGGGTATTC	58.3	
PIK3C2B_18F	CATGGCTCAAGGCAGACTC	59.5	544
PIK3C2B_18R	AGACACAGCTCGTTTCAGGAG	59.7	
PIK3C2B_19F	TCCAAGTGAGTCTGCATTC	58.8	1978
PIK3C2B_19R	TGTGGGTAGAGGAGAGACAGG	59.3	
PIK3C2B_20F	AAGACAGAGTGAGGGCACAG	58.0	522
PIK3C2B_20R	GGCCTAGTTCCAGAGTCAGC	60.4	
PIK3C2B_21F	GCTGTTCTGCCATCTGCTC	59.7	571
PIK3C2B_21R	TTTCTGCCATGTGTGGGTAG	59.6	
PIK3C2B_22F	TGGGAGGTTAAGTGCTGGAG	60.2	569
PIK3C2B_22R	GCCTCCTAGAAAGCACTTCG	59.2	
PIK3C2B_23F	ATTGCAGGACAAGCTGGAAC	60.3	539
PIK3C2B_23R	CTGTCACGGCACCAAGTCTG	62.9	
PIK3C2B_24F	GCAGCTATGAACTGGAGAGACC	60.4	582
PIK3C2B_24R	TTTCTCAGTCCAGCATCAGG	59.0	
PIK3C2B_25F	CATGGGAGATGTTTCTGTGTG	59.0	582
PIK3C2B_25R	CCTTGGTAGGGAAGAACAGC	58.8	
PIK3C2B_26F	TGATGAGGGATGAGGGAAAC	59.9	499
PIK3C2B_26R	CAGATGCTGTGGGCAGTG	60.0	
PIK3C2B_27F	CCTGCCTGCCCTAGATAAAG	58.9	563
PIK3C2B_27R	ACACAGATGGGAAACCAAGG	59.8	
PIK3C2B_28F	TGTGCCAGTCATTTGCTTATG	61.0	584

PIK3C2B_28R	TGCCAGGATCCTTGACTCAG	61.4	
PIK3C2B_29F	GATCTGCCTAGAATTGGGTAGC	59.2	548
PIK3C2B_29R	GAGCTTGCCAGCAGG	59.1	
PIK3C2B_30F	GGGCCCTTGAAATGTAGTC	60.7	505
PIK3C2B_30R	ACCCTCCTCTGTCTTCTGC	60.8	
PIK3C2B_31F	AAGAACTGCATGTTGTTGGTTG	60.1	445
PIK3C2B_31R	GAGTTAATTGGGTTGGGATGG	60.4	
PIK3C2B_32F	CACAGGGTGGATATTCAGTCG	60.4	533
PIK3C2B_32R	TGAGGCTCAGAATGATTAGG	55.5	
PIK3C2B_33F	AGATGCCAGATGGCAGATG	59.8	400
PIK3C2B_33R	CTGCCATTGCAACATGAAAG	60.3	
PIK3C2B_34F	CAATACAAAGCCCTCACCTG	58.2	672
PIK3C2B_34R	ACTTTGCTGCAACCTCACAG	59.1	

Tab. D.4 *HERC1*

Primer-Name	Primer Sequence 5'-3'	Primer T _m [°C]	Länge des Amplikons (bp)
HERC1_2F	CACCCGGCTGCATTCA	63	1270
HERC1_2R	CAGTAGTCTTAACACTGTACCAACCA	59	
HERC1_3F	TGTGATGCAGTAGCCAGCTC	60	454
HERC1_3R	CCCACCTAGATACCCAGTTC	60	
HERC1_4F	ATTGTGTGTCTCCGTTTCC	59.8	
HERC1_4R	CTGACCTCGTGATCCACCTG	61.7	
HERC1_5F	AAGGCCATTTCGTCTCTTAG	59.8	597
HERC1_5R	TTAAGCCAAAGGGTTGAAGG	59.2	
HERC1_6F	TCATCCTTGGCCTGAACAC	59.6	346
HERC1_6R	TTGGCCTAGTAAATCTGCTCTC	59.1	
HERC1_7F	CCCAGTTTAGGAGCACCTT	56.8	439
HERC1_7R	TGCTCATCTGTATTTGAGTCATTC	58.4	
HERC1_8F	TGCCAGTTAGTGATGGAAAC	60	434
HERC1_8R	CTGGGTCACATATGCCAGGT	60.8	

HERC1_9F	AAAGGCAGCATGATAAATCGC	61.4	448
HERC1_9R	CCCAGGCATAAACTTCATTATCTA	58.7	
HERC1_10F	TTTGAGACCCAAGCTTATTGAAGA	61.4	569
HERC1_10R	TCGGAAACATGTCTAACCTCAAC	60.4	
HERC1_11F	AGGTAATCTGCATTCTTGG	56.8	400
HERC1_11R	GGAAACCAAACAAGAAAGAACAG	59.2	
HERC1_12F	AGAGGCGGGGTTTTGC	59.7	484
HERC1_12R	CAACATCCAATTGAAATTGTCC	59.7	
HERC1_13F	GATTAGCTATTTAGTGGCAAATCG	58.5	457
HERC1_13R	GCACAACACATTAACCCATCA	59.3	
HERC1_14F	TCCTAAATGGGATTAATTGGGTAG	59.5	494
HERC1_14R	GGTCTAGAGTTTGATTTCTATCATAGGTG	60.2	
HERC1_15F	CTCCCAAAGTGCTGGGATTA	60.1	
HERC1_15R	TGCCACACTGCTTGAGTTCT	59.6	
HERC1_16F	TGAACTGCTATGTTCACTACAGAAAC	59	523
HERC1_16R	CTGTCATGCCAAGTAGACTCTTTC	59.5	
HERC1_17F	TGTTTGTTGGTGGTACTTTCTGA	59.6	569
HERC1_17R	CTCCAGCCCAGGATTCTTTC	61.1	
HERC1_18F	TGCAGTACGGTAGGAGACTATGAGA	61.5	411
HERC1_18R	CAATTGGTGCCTGTCAGTTC	59.1	
HERC1_19F	TTGGAGTGTACAATTGGTAGAG	59.2	613
HERC1_19R	TTTCACACAGTAGGTTTAAGTGG	56.1	
HERC1_20F	GCATTTGTTAAGTAACAGTTTAGAAAGATG	60.3	379
HERC1_20R	CTTTGTATACACAACGGTACACTTC	57	
HERC1_21F	CAGTGCAACTTTAAGTAATGAGTC	55.2	622
HERC1_21R	TTCAAATGTGGCCCAATC	57.8	
HERC1_22F	TGGGTTTTTGTAGTAAAGTTCATGTG	60.5	447
HERC1_22R	CAAATTGAGATCACCAGCAGA	58.9	
HERC1_23F	TCCCTCTTTCAGTAAGTTTTTGG	58.9	618
HERC1_23R	TCAGAAATTTATAGCCCATGCTG	60.5	

HERC1_24F	GCTCAGACAAGAGTCAGGATTC	59.5	356
HERC1_24R	AACAAACATATTACTTACTCTAACCAAAG	59.5	
HERC1_25F	TGGGTTTTTAGGGACATTTG	58.8	356
HERC1_25R	TTTTGCAGTTACATTTCAAAGACAA	60.1	
HERC1_26F	TTTGCTCAGATTCACCTTAATGC	57.2	
HERC1_26R	AAAGCTCAGGTACAGGAAAGG	57.6	
HERC1_27F	AAAGCTCAGGTACAGGAAAGG	57.6	
HERC1_27R	CCATGTCTCCTACATCCAAAG	57.1	
HERC1_28F	GGCTCTAGCATAGACTAGAAATAAAA	56.2	476
HERC1_28R	AAACTGTAACCAGAAGCAGACG	58.6	
HERC1_29F	AGTAAACGGTTCCTGCAAATG	59.1	582
HERC1_29R	AAACACACAGAAGGGCAAGC	60.3	
HERC1_30F	TTGAAAGTGGCCAGTACAAGG	60.2	630
HERC1_30R	CAGAATGGGCAGTATGGTCAC	60.4	
HERC1_31F	GCTGACAGCCTAAGATTTTGC	59	374
HERC1_31R	AAGGTTACACATGGAGTGTGCTA	60	
HERC1_32F	TACGCATTCATGGGTACAGG	59.4	514
HERC1_32R	TTTCTTTCCTTCTGACCTTGC	58.6	
HERC1_33F	ACCGTGCCTGGCTTCTAAT	59.7	459
HERC1_33R	GGGAGGGAATCAGTGGAAG	60.8	
HERC1_34F	GTTCTGTTGCCCTGCA	59.3	485
HERC1_34R	GAGTTCCTGTATATCCTTCACTTGG	59.5	
HERC1_35F	TGCCCTGCAAGATAAAGAATG	60.2	514
HERC1_35R	GGCACTGAATAACCGAGTTG	59.6	
HERC1_36F	CCAATAAGATTTTTGAAACAGAT	57.8	531
HERC1_36R	TCTTGTGACTGCACGTTTGA	59	
HERC1_37F	AACAGGGCAGCTTATTCCA	58.3	934
HERC1_37R	AAACCCACTAGTATTTCAAAGAACA	57.5	
HERC1_37F	CTTACTCTTTGGGGGCCAGT	60.5	730
HERC1_37R	GCAATGGCCTATTTACCAT	59.8	

HERC1_38F	CACATCTCCCCATTTTCTCA	58.5	981
HERC1_38R	AAAAATCAGATAACATACCTTGCT	55.5	
HERC1_39F	TCATTACAATACTGTCTGTAGAGTCTGG	59.3	642
HERC1_39R	TGGGATCACCACCTAACTGG	56.8	
HERC1_40F	AACCTGCTTTGTGCTGG	57.4	579
HERC1_40R	GGCCATAATTCATCCACAG	57.9	
HERC1_41F	GCCCAGCCACATATTCTGAG	60.6	444
HERC1_41R	AACAGTCTGATGCTGACTTTGG	59.4	
HERC1_42F	TTTAAGAGGCTATGAACTTGG	55.6	445
HERC1_42R	TTTATGACGGGATATATAATAAATGG	56.8	
HERC1_43F	GAACTGTTCAAGGGATAAAGATGT	57.8	363
HERC1_43R	TTTGAATCTTCTTTAATGTGCCTTT	59.6	
HERC1_44F	GATCCAATAAGACAACCTGAGATAGG	57.5	427
HERC1_44R	AACTCTAGGCAAGGCTGCTG	59.8	
HERC1_45F	AGCATTAAATGAACATGTTGGAAA	58.6	538
HERC1_45R	CCCTTCTGCCCAAAGTGTA	60.1	
HERC1_46F	TTTGTGCTGCTTTATTAATGACCT	59.3	259
HERC1_46R	CACACACACACGAAGGTG	59.6	
HERC1_47F	AGAACAACCATTGCCAGCAG	61.2	523
HERC1_47R	AAGCGACCTTCTTTCTTCAGC	60.1	
HERC1_48F	TGTATAACCAAGTTCAGATGAGC	59.6	476
HERC1_48R	TGAATGACCGATTGAGTGAATAAC	60.2	
HERC1_49F	CGTTTAGATGCCAGTTTCCA	58.8	609
HERC1_49R	CACCTGTTGCAGCAGAAATC	59.4	
HERC1_50F	ATTCTTGAGCCTTTTGAATGA	58	462
HERC1_50R	ATCTTTGTTTCCATACACAGTGAA	57.7	
HERC1_51F	TAACCCATTTGCTTGTTCACC	59.9	489
HERC1_51R	TTGGTAACTGCCAGGAACATC	60	
HERC1_52F	AGCTGTTTGGAGGCTGAGG	60.5	590
HERC1_52R	TTGATTAATTCTAAGTCCAGATACAG	56.6	

HERC1_53F	CTGCTAGCCAAGGTTTTGAA	58.2	356
HERC1_53R	TGAAGAGAAACCACTGTTTGGGA	59.8	
HERC1_54F	CCTGCCTTCATGGTCTTATTC	58.7	550
HERC1_54R	TTTGACAGAGATGTGTGGTG	55.4	
HERC1_55F	AGCCCTTGTCACAGATAAGC	57	529
HERC1_55R	CAAGGGAGACTTGGGAAGGTG	59.7	
HERC1_56F	AACATTTGTTGAGCTTTAACTGC	57.7	675
HERC1_56R	CAACCGAGGCCAATATTCTAAC	59.9	
HERC1_57F	TGTCCAGTGTGAAACTTGGTATG	59.9	443
HERC1_57R	AATGGATCTTGCCCTTGG	59.9	
HERC1_58F	AAGCTTCAACAAAGGGAATCAG	59.8	581
HERC1_58R	TTAGCTAAATTCAGCAACGTC	55.1	
HERC1_59F	TTTCAGAATCCTCTGCCCTTAC	59.7	429
HERC1_59R	CCCTCTGGACATCAGGTTTC	59.5	
HERC1_60F	GGAATTCGCAGATTAATGG	58.6	439
HERC1_60R	CAAGAATAGGCTTAGGGTTGC	58	
HERC1_61F	TGATAAATGTGCCAGGGTTG	59.4	753
HERC1_61R	AGGTTTACAAGAGTTCAACAGTGCT	59.8	
HERC1_62F	CTAGCTTTGAATCACCAATGC	57.5	543
HERC1_62R	CAGCTGTCCATGCCTACCA	60.8	
HERC1_63F	TAAATGGATTAACGGCATGG	58.4	447
HERC1_63R	GGACAAGCTTTATCCACAATAAG	60.2	
HERC1_64F	GGATGAATCTCATACATCTCTTGG	59	511
HERC1_64R	TGGAAATTAAGGCTGCACAC	58.8	
HERC1_65F	TCAGTAGCAGCTGCAAGTAAAG	57.7	417
HERC1_65R	TGACAAGCTCAAGAATGAATCTG	59.5	
HERC1_66F	TCCAGCCTAGGTGACAGAGC	60.6	452
HERC1_66R	CCTGATCTTCCTGAGGGTGT	59.1	
HERC1_67F	TGGTTTAGCCACAGACATTTG	58.7	594
HERC1_67R	GGAGACAGAAGCTTATCCTTGC	59.5	

HERC1_68F	TTTGGATTTGGGTGGAACAG	60.9	318
HERC1_68R	GAGAGCACATTCTGTAACACTGG	58.9	
HERC1_69F	CAGAACGTGGTGTACCTTTCC	59.5	523
HERC1_69R	ATACTTCTTTGGCCGAGTC	59.3	
HERC1_70F	TATTGATGCTGATGCATGG	56.5	451
HERC1_70R	GCAGAACAAATGCCTGCTATG	60.8	
HERC1_71F	CCAACACAAATGCACACACC	60.9	550
HERC1_71R	CACAGAAGGAAGACCACTCTGTC	60.3	
HERC1_72F	CACCACAGCCTCCCAAAG	60.2	467
HERC1_72R	TGCTAGGCCCACTGTAACCA	61.6	
HERC1_73F	GTGGTTACAGTGGGCCTAGC	59.6	420
HERC1_73R	GAGGCTCACAAATGCATCC	59.2	
HERC1_74F	TACTCCTGCCCTGGATTGAC	60.1	527
HERC1_74R	GGCGTAAGCCACCATGC	61.2	
HERC1_75F	TCAACGTGTGATTAGGGAAGC	60.1	603
HERC1_75R	TAAACTGCTGCGGTTGAGC	60.1	
HERC1_76F	TGAGTCATCACATTCTGCTG	59.8	538
HERC1_76R	TGAGACTGAACCACTGCACTC	59	
HERC1_77F	CAGCAGGGAGGCACTCTAAC	60	709
HERC1_77R	GGTGGCACGCACCTGTAG	61.3	
HERC1_78F	CTGGAGGAGGTGAGGACAAG	59.8	634
HERC1_78R	AACCTGACTGAGAGCACTTCG	59.7	

E) Alle CNVs

Tab. E.1 CNVs aller mutationsnegativer Patienten (hg19)

CNV Chromosomenposition	Patienten-ID	Länge CNV (bp)	Anzahl SNP-Marker	CNV	max. LBF	Gen / Gene
chr2:115453185-115689106	HNPCC_72	235.922	59	3	66	<i>DPP10</i>
chr2:230822296-230875311	HNPCC_24	53.016	8	1	22	<i>FBXO36</i>
chr2:30767529-30967074	HNPCC_32	199.546	79	3	189	<i>LCLAT1, CAPN13</i>
chr2:45752140-45765157	HNPCC_31	13.018	5	1	10	<i>SRBD1</i>
chr2:73937901-74007136	HNPCC_45	69.236	13	3	54	<i>TPRKB, DUSP11</i>
chr3:139306828-139336381	HNPCC_18	29.554	8	1	24	<i>NMNAT3</i>
chr3:142800051-143588201	HNPCC_97	788.151	40+237+12	3	119+86 0+30	<i>CHST2, SLC9A9</i>
chr3:182669780-183421149	HNPCC_73	751.370	177	3	304	<i>DCUN1D1, MCCC1, LAMP3, MCF2L2, B3GNT5, KLHL6, KLHL24, YEATS2</i>
chr4:8476118-8570973	HNPCC_43	94.856	39	1	89	<i>TRMT44</i>
chr4:99298166-99428163	HNPCC_38	129.998	27	1	97	<i>RAP1GDS1, TSPAN5</i>
chr5:101625154-101641550	HNPCC_31	16.397	7	3	21	<i>SLCO4C1</i>
chr5:143093377-143372963	HNPCC_40	279.587	63	3	246	<i>HMHB1</i>
chr5:180653039-180693127	HNPCC_38	40.089	13	3	30	<i>TRIM41, GNB2L1, TRIM52</i>
chr6:151118271-151412754	HNPCC_50	294.484	143	3	283	<i>PLEKHG1, MTHFD1L</i>
chr6:97440636-97476678	HNPCC_91	36.043	12	1	49	<i>KLHL32</i>
chr7:12335168-12487380	HNPCC_32	152.213	59	1	255	<i>VWDE</i>
chr7:15413574-15456251	HNPCC_29	42.678	10	1	34	<i>AGMO</i>
chr7:158401375-158651087	HNPCC_92	249.713	65	3	86	<i>NCAPG2, ESYT2, WDR60</i>
chr7:87307509-87341709	HNPCC_34	34.201	27	1	120	<i>ABC1, RUNDC3B</i>
chr8:48792199-48882899	HNPCC_41	90.701	13	1	43	<i>PRKDC, MCM4</i>
chr8:81940988-81958305	HNPCC_47	17.318	11	1	18	<i>PAG1</i>
chr10:129321875-129350641	HNPCC_45	28.767	22	1	70	<i>NPS</i>
chr10:2711041-3212425	HNPCC_28	501.385	285	3	881	<i>PFKP, PITRM1</i>
chr10:7460252-7767196	HNPCC_89	306.945	142	3	369	<i>ITIH5, ITIH2</i>
chr10:84303712-84336472	HNPCC_38	32.761	10	1	40	<i>NRG3</i>
chr11:120487899-120875383	HNPCC_94	387.485	155	1	258	<i>GRIK4</i>
chr11:84596023-84627035	HNPCC_28	31.013	11	1	30	<i>DLG2</i>

chr12:114285113-114324609	HNPCC_38	39.497	13	3	30	<i>RBM19</i>
chr12:191619-216234	HNPCC_26	24.616	6	3	24	<i>IQSEC3</i>
chr12:3374920-3411438	HNPCC_18	36.519	20	3	38	<i>TSPAN9</i>
chr12:41846679-41891527	HNPCC_41	44.849	12	3	48	<i>PDZRN4</i>
chr13:25284183-25334331	HNPCC_97	50.149	20	3	38	<i>ATP12A</i>
chr13:30787214-30830634	HNPCC_90	43.421	14	3	43	<i>KATNAL1</i>
chr13:44030966-44172328	HNPCC_24	141.363	33	3	105	<i>ENOX1</i>
chr14:23856026-23883184	HNPCC_18	27.159	21	3	44	<i>MYH6, MYH7</i>
chr14:94195177-94557404	HNPCC_01	362.228	117 + 62	3	249 + 139	<i>PRIMA1, FAM181A, ASB2, OTUB2, DDX24, IFI27L1</i>
chr16:87781726-87802776	HNPCC_33	21.051	10	1	24	<i>KLHDC4</i>
chr17:64513878-64550045	HNPCC_42	36.168	19	1	42	<i>PRKCA</i>
chr17:65400254-65431947	HNPCC_21	31.694	9	1	14	<i>PITPNC1</i>
chr19:50056123-50086806	HNPCC_18	30.684	9	1	32	<i>NOSIP, PRRG2</i>
chr19:58194388-58259946	HNPCC_79	65.559	24	1	28	<i>ZNF551, ZNF154, ZNF671, ZNF776</i>
chr20:44238741-44279034	HNPCC_55	40.294	10	1	20	<i>WFDC9, WFDC10A, WFDC11</i>

F) CRC-Kandidatengene und Interaktionspartnergene von MSH2

Tab. F.1 CRC-Kandidatengene und Interaktionspartnergene von *MSH2* +/- 50 kb (hg19)

Gen	Chr	Start	Ende	Größe	Start - 50kb	Ende + 50kb
<i>MUTYH</i>	1	45794913	45805787	10.874	45744913	45855787
<i>JUN</i>	1	59246462	59249785	3.323	59196462	59299785
<i>USP1</i>	1	62901974	62917475	23.398	62851974	62967475
<i>EXO1</i>	1	242011492	242053241	41.749	241961492	242103241
<i>EPCAM</i>	2	47596286	47614167	17.881	47546286	47664167
<i>KCNK12</i>	2	47747914	47797470	49.556	47697914	47847470
<i>MSH6</i>	2	48010220	48034092	23.872	47960220	48084092
<i>ERCC3</i>	2	128014865	128051752	36.887	127964865	128101752
<i>PMS1</i>	2	190648810	190742355	93.545	190598810	190792355
<i>BARD1</i>	2	215593274	215674428	81.154	215543274	215724428
<i>XRCC5</i>	2	216974019	217071016	96.997	216924019	217121016

<i>OGG1</i>	3	9791627	9808353	16.726	9741627	9858353
<i>XPC</i>	3	14186647	14220172	33.525	14136647	14270172
<i>TGFBR2</i>	3	30647993	30735633	87.64	30597993	30785633
<i>MLH1</i>	3	37034840	37092337	57.497	36984840	37142337
<i>CDC25A</i>	3	48198667	48229801	31.134	48148667	48279801
<i>ATRIP</i>	3	48488113	48507054	18.941	48438113	48557054
<i>TREX1</i>	3	48506918	48509044	2.126	48456918	48559044
<i>FHIT</i>	3	59735035	61237133	1.502.098	59685035	61287133
<i>MCM2</i>	3	127317252	127341278	24.026	127267252	127391278
<i>MBD4</i>	3	129149792	129158852	9.06	129099792	129208852
<i>ATR</i>	3	142168076	142297668	129.592	142118076	142347668
<i>RFC4</i>	3	186507681	186524484	16.803	186457681	186574484
<i>MSX1</i>	4	4861391	4865660	4.269	4811391	4915660
<i>DHX15</i>	4	24529087	24586184	57.097	24479087	24636184
<i>RFC1</i>	4	39289068	39368001	78.933	39239068	39418001
<i>N4BP2</i>	4	40058523	40159872	101.349	40008523	40209872
<i>AASDH</i>	4	57204456	57253638	49.182	57154456	57303638
<i>MAD2L1</i>	4	120980578	120988013	7.435	120930578	121038013
<i>MSH3</i>	5	79950466	80172634	222.168	79900466	80222634
<i>XRCC4</i>	5	82373316	82649579	276.263	82323316	82699579
<i>APC</i>	5	112043201	112181936	138.735	111993201	112231936
<i>REEP5</i>	5	112212080	112258031	45.951	112162080	112308031
<i>RAD50</i>	5	131892615	131980313	87.698	131842615	132030313
<i>SIL1</i>	5	138282409	138534065	251.656	138232409	138584065
<i>DNAH8</i>	6	38683116	38998574	315.458	38633116	39048574
<i>POLH</i>	6	43543877	43588260	44.383	43493877	43638260
<i>MCM3</i>	6	52128811	52149679	20.868	52078811	52199679
<i>ESR1</i>	6	152011630	152424408	412.778	151961630	152474408
<i>NUDT1</i>	7	2281856	2290780	8.924	2231856	2340780
<i>PMS2</i>	7	6012869	6048737	35.868	5962869	6098737

<i>RFC2</i>	7	73645831	73668738	22.907	73595831	73718738
<i>BRAF</i>	7	140433812	140624564	190.752	140383812	140674564
<i>SCARA3</i>	8	27491576	27534286	42.71	27441576	27584286
<i>WRN</i>	8	30890777	31031277	140.5	30840777	31081277
<i>MCM4</i>	8	48872762	48890719	17.957	48822762	48940719
<i>NBN</i>	8	90945563	90996899	51.336	90895563	91046899
<i>MYC</i>	8	128748314	128753680	5.366	128698314	128803680
<i>CDKN2A</i>	9	21967750	21994490	26.74	21917750	22044490
<i>XPA</i>	9	100437190	100459691	22.501	100387190	100509691
<i>POLE3</i>	9	116169517	116173029	3.512	116119517	116223029
<i>PTEN</i>	10	89623194	89728532	105.338	89573194	89778532
<i>MGMT</i>	10	131265453	131565783	300.33	131215453	131615783
<i>HRAS</i>	11	532241	535550	3.309	482241	585550
<i>FEN1</i>	11	61560108	61564714	4.606	61510108	61614714
<i>OTUB1</i>	11	63753324	63765892	12.568	63703324	63815892
<i>MRE11A</i>	11	94150468	94227040	76.572	94100468	94277040
<i>CHEK1</i>	11	125495030	125546150	51.12	125445030	125596150
<i>RAD52</i>	12	1021254	1058863	37.609	971254	1108863
<i>DPPA3</i>	12	7864088	7870152	6.064	7814088	7920152
<i>AICDA</i>	12	8754761	8765442	10.681	8704761	8815442
<i>RECQL</i>	12	21621843	21654603	32.76	21571843	21704603
<i>KRAS</i>	12	25358179	25403854	45.675	25308179	25453854
<i>UNG</i>	12	109535398	109548798	13.4	109485398	109598798
<i>RFC5</i>	12	118454505	118470042	15.537	118404505	118520042
<i>UBC</i>	12	125396191	125399587	3.396	125346191	125449587
<i>POLE</i>	12	133200347	133263945	63.598	133150347	133313945
<i>BRCA2</i>	13	32889616	32973809	84.193	32839616	33023809
<i>RFC3</i>	13	34392205	34540695	148.49	34342205	34590695
<i>ERCC5</i>	13	103498190	103528351	30.161	103448190	103578351
<i>LIG4</i>	13	108859791	108870716	10.925	108809791	108920716

<i>F10</i>	13	113777112	113803843	26.731	113727112	113853843
<i>MAX</i>	14	65472818	65569262	96.444	65422818	65619262
<i>MLH3</i>	14	75480466	75518235	37.769	75430466	75568235
<i>RAD51</i>	15	40987326	41024356	37.03	40937326	41074356
<i>BLM</i>	15	91260578	91358686	98.108	91210578	91408686
<i>NTHL1</i>	16	2089815	2097867	8.052	2039815	2147867
<i>CREBBP</i>	16	3775055	3930121	155.066	3725055	3980121
<i>ERCC4</i>	16	14014013	14046205	32.192	13964013	14096205
<i>E2F4</i>	16	67226067	67232821	6.754	67176067	67282821
<i>CDH1</i>	16	68771194	68869444	98.25	68721194	68919444
<i>RPA1</i>	17	1733272	1802848	69.576	1683272	1852848
<i>TP53</i>	17	7571719	7590868	19.149	7521719	7640868
<i>NF1</i>	17	29421944	29704695	282.751	29371944	29754695
<i>KAT2A</i>	17	40265128	40273382	8.254	40215128	40323382
<i>BRCA1</i>	17	41196311	41277500	81.189	41146311	41327500
<i>SMAD4</i>	18	48556582	48611411	54.829	48506582	48661411
<i>STK11</i>	19	1205797	1228434	22.637	1155797	1278434
<i>ERCC2</i>	19	45854648	45873845	19.197	45804648	45923845
<i>ERCC1</i>	19	45910590	45927177	16.587	45860590	45977177
<i>BAX</i>	19	49458116	49465055	6.939	49408116	49515055
<i>NR1H2</i>	19	50879679	50886285	6.606	50829679	50936285
<i>POLD1</i>	19	50887579	50921275	33.696	50837579	50971275
<i>PCNA</i>	20	5095598	5107268	11.67	5045598	5157268
<i>SPO11</i>	20	55904830	55919049	14.219	55854830	55969049
<i>CHEK2</i>	22	29083730	29137822	54.092	29033730	29187822
<i>XRCC6</i>	22	42017294	42060052	42.758	41967294	42110052

G) Target-Panel NGS

Tab. G.1 Genliste NGS

Genname	Ziel-Gen Herkunft	Chromosomenposition (hg19)	Genlänge (bp)
<i>ABC1</i>	CNV-Analyse	chr7:87133178-87342639	209.461

Anhang

<i>AGMO</i>	CNV-Analyse	chr7:15239942-15601640	361.698
<i>ASB2</i>	CNV-Analyse	chr14:94400498-94443076	42.578
<i>B3GNT5</i>	CNV-Analyse	chr3:182971031-182991179	20.148
<i>C9orf84</i>	CNV-Analyse	chr9:114448900-114545779	96.879
<i>CAPN13</i>	CNV-Analyse	chr2:30945637-31030311	84.674
<i>CHST2</i>	CNV-Analyse	chr3:142838617-142842856	4.239
<i>DCUN1D1</i>	CNV-Analyse	chr3:182660558-182698326	37.768
<i>DDX24</i>	CNV-Analyse	chr14:94517267-94547558	30.291
<i>DLG2</i>	CNV-Analyse	chr11:83166055-85338314	2.172.259
<i>DPP10</i>	CNV-Analyse	chr2:115199898-116602326	1.402.428
<i>DUSP11</i>	CNV-Analyse	chr2:73989324-74007284	17.960
<i>ENOX1</i>	CNV-Analyse	chr13:43787665-44203613	415.948
<i>ESYT2</i>	CNV-Analyse	chr7:158523688-158622319	98.631
<i>FAM181A</i>	CNV-Analyse	chr14:94385239-94395954	10.715
<i>FBXO36</i>	CNV-Analyse	chr2:230787206-230948016	160.810
<i>GNB2L1</i>	CNV-Analyse	chr5:180663927-180670906	6.979
<i>GNG10</i>	CNV-Analyse	chr9:114423850-114432526	8.676
<i>GRIK4</i>	CNV-Analyse	chr11:120382467-120856969	474.502
<i>IFI27L1</i>	CNV-Analyse	chr14:94547638-94569060	21.422
<i>ITIH5</i>	CNV-Analyse	chr10:7601231-7708961	107.730
<i>KATNAL1</i>	CNV-Analyse	chr13:30776766-30881191	104.425
<i>KLHDC4</i>	CNV-Analyse	chr16:87741417-87799598	58.181
<i>KLHL24</i>	CNV-Analyse	chr3:183353410-183402304	48.894
<i>KLHL32</i>	CNV-Analyse	chr6:97372495-97588630	216.135
<i>KLHL6</i>	CNV-Analyse	chr3:183205318-183273499	68.181
<i>LAMP3</i>	CNV-Analyse	chr3:182840002-182880667	40.665
<i>LCLAT1</i>	CNV-Analyse	chr2:30670122-30867091	196.969
<i>MCCC1</i>	CNV-Analyse	chr3:182733005-182817365	84.360
<i>MCF2L2</i>	CNV-Analyse	chr3:182895830-183145855	250.025
<i>MCM4</i>	CNV-Analyse	chr8:48872762-48890719	17.957
<i>MTHFD1L</i>	CNV-Analyse	chr6:151186814-151423023	236.209
<i>NCAPG2</i>	CNV-Analyse	chr7:158424002-158497520	73.518
<i>NMNAT3</i>	CNV-Analyse	chr3:139279022-139346606	67.584
<i>NOSIP</i>	CNV-Analyse	chr19:50058724-50083829	25.105
<i>NRG3</i>	CNV-Analyse	chr10:83635069-84746935	1.111.866
<i>OTUB2</i>	CNV-Analyse	chr14:94492723-94515276	22.553
<i>PAG1</i>	CNV-Analyse	chr8:81880045-82024303	144.258
<i>PDZRN4</i>	CNV-Analyse	chr12:41582249-41968392	386.143
<i>PITPNC1</i>	CNV-Analyse	chr17:65373396-65693379	319.983
<i>PLEKHG1</i>	CNV-Analyse	chr6:150920998-151153341	232.343
<i>PRIMA1</i>	CNV-Analyse	chr14:94184643-94254766	70.123
<i>PRKCA</i>	CNV-Analyse	chr17:64298925-64806862	507.937
<i>PRKDC</i>	CNV-Analyse	chr8:48685668-48872743	187.075
<i>PRRG2</i>	CNV-Analyse	chr19:50084586-50094265	9.679
<i>RAP1GDS1</i>	CNV-Analyse	chr4:99182526-99365012	182.486
<i>RBM19</i>	CNV-Analyse	chr12:114259858-114404176	144.318

<i>RUNDC3B</i>	CNV-Analyse	chr7:87257728-87461613	203.885
<i>SLC9A9</i>	CNV-Analyse	chr3:142984063-143567373	583.310
<i>SLCO4C1</i>	CNV-Analyse	chr5:101569691-101632253	62.562
<i>SRBD1</i>	CNV-Analyse	chr2:45615818-45838433	222.615
<i>TPRKB</i>	CNV-Analyse	chr2:73956956-73961718	4.762
<i>TRIM41</i>	CNV-Analyse	chr5:180650262-180662808	12.546
<i>TRIM52</i>	CNV-Analyse	chr5:180683385-180688119	4.734
<i>TRMT44</i>	CNV-Analyse	chr4:8442531-8481260	38.729
<i>TSPAN5</i>	CNV-Analyse	chr4:99391517-99579812	188.295
<i>TSPAN9</i>	CNV-Analyse	chr12:3186520-3395730	209.210
<i>VWDE</i>	CNV-Analyse	chr7:12370508-12443852	73.344
<i>WDR60</i>	CNV-Analyse	chr7:158649268-158738883	89.615
<i>WFDC10A</i>	CNV-Analyse	chr20:44258384-44259831	1.447
<i>WFDC11</i>	CNV-Analyse	chr20:44277201-44298878	21.677
<i>WFDC9</i>	CNV-Analyse	chr20:44236577-44259907	23.330
<i>YEATS2</i>	CNV-Analyse	chr3:183415605-183530413	114.808
<i>ZNF154</i>	CNV-Analyse	chr19:58211809-58220579	8.770
<i>ZNF671</i>	CNV-Analyse	chr19:58231118-58238995	7.877
<i>ZNF776</i>	CNV-Analyse	chr19:58193336-58269527	76.191
<i>APC</i>	CRC-Gen	chr5:111993201-112231936	238.735
<i>APC2</i>	CRC-Gen	chr19:1400148-1523243	123.095
<i>AXIN1</i>	CRC-Gen	chr16:287440-452676	165.236
<i>AXIN2</i>	CRC-Gen	chr17:63474683-63607740	133.057
<i>BMPR1A</i>	CRC-Gen	chr10:88466396-88734945	268.549
<i>BUB1B</i>	CRC-Gen	chr15:40403210-40563337	160.127
<i>CDH1</i>	CRC-Gen	chr16:68721195-68919444	198.249
<i>CHEK2</i>	CRC-Gen	chr22:29033730-29187822	154.092
<i>FBXW7</i>	CRC-Gen	chr4:153192410-153506185	313.775
<i>GREM1</i>	CRC-Gen	chr15:32960205-33076870	116.665
<i>GSK3B</i>	CRC-Gen	chr3:119490802-119863264	372.462
<i>MAP2K4</i>	CRC-Gen	chr17:11874135-12097148	223.013
<i>MLH1</i>	CRC-Gen	chr3:36984840-37142337	157.497
<i>MSH2</i>	CRC-Gen	chr2:47580206-47760367	180.161
<i>MSH6</i>	CRC-Gen	chr2:47960220-48084092	123.872
<i>MUTYH</i>	CRC-Gen	chr1:45744913-45855787	110.874
<i>PDGFRA</i>	CRC-Gen	chr4:55045264-55214412	169.148
<i>PMS2</i>	CRC-Gen	chr7:5962869-6098737	135.868
<i>POLD1</i>	CRC-Gen	chr19:50837579-50971275	133.696
<i>POLE</i>	CRC-Gen	chr12:133150347-133313945	163.598
<i>PPP2R1B</i>	CRC-Gen	chr11:111558605-111687169	128.564
<i>PTEN</i>	CRC-Gen	chr10:89573194-89778532	205.338
<i>PTPRJ</i>	CRC-Gen	chr11:47952110-48242394	290.284
<i>SFRP1</i>	CRC-Gen	chr8:41069476-41216990	147.514
<i>SMAD2</i>	CRC-Gen	chr18:45309466-45506970	197.504
<i>SMAD4</i>	CRC-Gen	chr18:48506582-48661411	154.829
<i>TCF7L2</i>	CRC-Gen	chr10:114660009-114977436	317.427

<i>TP53</i>	CRC-Gen	chr17:7521719-7640868	119.149
<i>WIF1</i>	CRC-Gen	chr12:65394404-65565346	170.942
<i>AKT1</i>	Proteasom-Pathway	chr14:105235636-105262130	26.494
<i>HERC1</i>	Proteasom-Pathway	chr15:63850817-64176147	325.330
<i>MTOR</i>	Proteasom-Pathway	chr1:11116588-11372608	256.020
<i>PIK3C2B</i>	Proteasom-Pathway	chr1:204341758-204513852	172.094
<i>PPP2CA</i>	Proteasom-Pathway	chr5:133,532,148-133,561,950	29.802
<i>PRKCZ</i>	Proteasom-Pathway	chr1:1931909-2166834	234.925
<i>TSC1</i>	Proteasom-Pathway	chr9:135766684-135820070	53.386
<i>TSC2</i>	Proteasom-Pathway	chr16:2097939-2138763	40.824

H) Pathway-Analyse

Tab. H.1 Involvierte Signalwege der CRC-Kandidatengene

kanonische Signalwege	Gene	p-Wert
Colorectal Cancer Metastasis Signaling	<i>TP53, MAP2K4, PIK3C2B, SMAD2, AXIN1, GNB2L1, MLH1, APC, CDH1, AKT1, MSH2, MSH6, SMAD4, GSK3B, TCF7L2</i>	1.00E-13
Factors Promoting Cardiogenesis in Vertebrates	<i>SMAD2, AXIN1, BMPR1A, SMAD4, MYH7, GSK3B, TCF7L2, APC, PRKCZ, PRKCA</i>	5.01E-11
Molecular Mechanisms of Cancer	<i>TP53, MAP2K4, PIK3C2B, PRKDC, SMAD2, AXIN1, PRKCZ, APC, CDH1, AKT1, BMPR1A, SMAD4, GSK3B, CHEK2, PRKCA</i>	7.76E-10
ILK Signaling	<i>MAP2K4, PIK3C2B, MYH6, CDH1, MTOR, AKT1, PPP2CA, MYH7, GSK3B, PPP2R1B, PTEN</i>	3.63E-09
HER-2 Signaling in Breast Cancer	<i>TP53, TSC1, PIK3C2B, AKT1, TSC2, GSK3B, PRKCZ, PRKCA</i>	6.03E-09
UVB-Induced MAPK Signaling	<i>TP53, MAP2K4, PIK3C2B, MTOR, AKT1, PRKCZ, PRKCA</i>	1.12E-08
ErbB Signaling	<i>MAP2K4, PIK3C2B, MTOR, AKT1, NRG3, GSK3B, PRKCZ, PRKCA</i>	1.62E-08
Glioma Signaling	<i>TP53, PIK3C2B, MTOR, AKT1, PDGFRA, PRKCZ, PRKCA, PTEN</i>	4.57E-08
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	<i>MAP2K4, PIK3C2B, WIF1, AKT1, AXIN1, APC2, GSK3B, SFRP1, TCF7L2, APC, PRKCZ, PRKCA</i>	5.01E-08
14-3-3-mediated Signaling	<i>MAP2K4, TSC1, PIK3C2B, AKT1, TSC2, GSK3B, PRKCZ, PRKCA</i>	1.82E-07
IL-8 Signaling	<i>MAP2K4, PIK3C2B, CDH1, MTOR, AKT1, GNB2L1, PRKCZ, PRKCA</i>	5.37E-07
mTOR Signaling	<i>TSC1, PIK3C2B, MTOR, AKT1, PPP2CA, TSC2, PPP2R1B, PRKCZ, PRKCA</i>	6.17E-07
p53 Signaling	<i>TP53, PIK3C2B, PRKDC, AKT1, GSK3B, CHEK2, PTEN</i>	8.32E-07
p70S6K Signaling	<i>PIK3C2B, MTOR, AKT1, PPP2CA, PPP2R1B, PRKCZ, PRKCA</i>	3.02E-06
Huntington's Disease Signaling	<i>TP53, MAP2K4, PIK3C2B, MTOR, AKT1, GNB2L1, PRKCZ, PRKCA</i>	3.31E-06
AMPK Signaling	<i>TSC1, PIK3C2B, MTOR, AKT1, PPP2CA, TSC2, PFKF, PPP2R1B</i>	4.37E-06
Role of NFAT in Cardiac Hypertrophy	<i>MAP2K4, PIK3C2B, AKT1, GNB2L1, GSK3B, PRKCZ, PRKCA</i>	4.57E-06
HIPPO signaling	<i>SMAD2, PPP2CA, DLG2, SMAD4, PPP2R1B, PRKCZ</i>	6.03E-06
Neuregulin Signaling	<i>MTOR, AKT1, NRG3, PRKCZ, PRKCA, PTEN</i>	6.76E-06
Breast Cancer Regulation by Stathmin1	<i>TP53, PIK3C2B, PPP2CA, GNB2L1, PPP2R1B, PRKCZ, PRKCA</i>	7.41E-06
EGF Signaling	<i>MAP2K4, PIK3C2B, MTOR, AKT1, PRKCA</i>	1.10E-05
Epithelial Adherens Junction Signaling	<i>MYH6, CDH1, AKT1, MYH7, TCF7L2, APC, PTEN</i>	1.17E-05
ErbB2-ErbB3 Signaling	<i>PIK3C2B, AKT1, NRG3, GSK3B, PTEN</i>	1.20E-05
Gαq Signaling	<i>PIK3C2B, AKT1, GNB2L1, GSK3B, PRKCZ, PRKCA</i>	1.23E-05
CXCR4 Signaling	<i>MAP2K4, PIK3C2B, AKT1, GNB2L1, PRKCZ, PRKCA</i>	1.51E-05
ErbB4 Signaling	<i>PIK3C2B, AKT1, NRG3, PRKCZ, PRKCA</i>	1.55E-05
Role of Tissue Factor in Cancer	<i>TP53, PIK3C2B, MTOR, AKT1, PRKCA, PTEN</i>	2.45E-05

Tight Junction Signaling	<i>MYH6, AKT1, PPP2CA, MYH7, PPP2R1B, PRKCZ, PTEN</i>	2.82E-05
CCR5 Signaling in Macrophages	<i>MAP2K4, GNB2L1, PRKCZ, PRKCA</i>	3.09E-05
CREB Signaling in Neurons	<i>PIK3C2B, AKT1, GRIK4, GNB2L1, PRKCZ, PRKCA</i>	3.24E-05
Type II Diabetes Mellitus Signaling	<i>MAP2K4, PIK3C2B, MTOR, AKT1, PRKCZ, PRKCA</i>	3.31E-05
Protein Kinase A Signaling	<i>PTPRJ, GNB2L1, SMAD4, DUSP11, GSK3B, TCF7L2, PRKCZ, PRKCA, PTEN</i>	3.55E-05
B Cell Receptor Signaling	<i>MAP2K4, PIK3C2B, MTOR, AKT1, PAG1, GSK3B, PTEN</i>	3.63E-05
P2Y Purigenic Receptor Signaling Pathway	<i>PIK3C2B, AKT1, GNB2L1, PRKCZ, PRKCA</i>	3.80E-05
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	<i>MAP2K4, PIK3C2B, AKT1, PPP2CA, PPP2R1B, PRKCZ, PRKCA</i>	4.57E-05
RAR Activation	<i>MAP2K4, SMAD2, AKT1, SMAD4, PRKCZ, PRKCA, PTEN</i>	6.31E-05
UVC-Induced MAPK Signaling	<i>TP53, MAP2K4, PRKCZ, PRKCA</i>	6.76E-05
Xenobiotic Metabolism Signaling	<i>CHST2, MAP2K4, PIK3C2B, ABCB1, PPP2CA, PPP2R1B, PRKCZ, PRKCA</i>	9.55E-05
G Beta Gamma Signaling	<i>AKT1, GNB2L1, PRKCZ, PRKCA</i>	1.00E-04
UVA-Induced MAPK Signaling	<i>TP53, MAP2K4, PIK3C2B, MTOR, PRKCA</i>	1.00E-04
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	<i>TP53, PRKDC, PRKCZ, CHEK2</i>	1.26E-04
Cardiomyocyte Differentiation via BMP Receptors	<i>BMPR1A, SMAD4, MYH7</i>	1.48E-04
Cardiac Hypertrophy Signaling	<i>MAP2K4, PIK3C2B, MTOR, AKT1, GNB2L1, GSK3B</i>	1.74E-04
Tec Kinase Signaling	<i>MAP2K4, PIK3C2B, GNB2L1, PRKCZ, PRKCA</i>	1.78E-04
HGF Signaling	<i>MAP2K4, PIK3C2B, AKT1, PRKCZ, PRKCA</i>	2.29E-04
Fc Epsilon RI Signaling	<i>MAP2K4, PIK3C2B, AKT1, PRKCZ, PRKCA</i>	2.57E-04
fMLP Signaling in Neutrophils	<i>PIK3C2B, GNB2L1, PRKCZ, PRKCA</i>	2.57E-04
CCR3 Signaling in Eosinophils	<i>PIK3C2B, GNB2L1, PRKCZ, PRKCA</i>	3.72E-04
Non-Small Cell Lung Cancer Signaling	<i>TP53, PIK3C2B, AKT1, PRKCA</i>	3.72E-04
NRF2-mediated Oxidative Stress Response	<i>MAP2K4, PIK3C2B, AKT1, GSK3B, PRKCZ, PRKCA</i>	3.72E-04
Erythropoietin Signaling	<i>PIK3C2B, AKT1, PRKCZ, PRKCA</i>	4.17E-04
Thrombin Signaling	<i>PIK3C2B, AKT1, GNB2L1, PRKCZ, PRKCA</i>	4.90E-04
IL-3 Signaling	<i>PIK3C2B, AKT1, PRKCZ, PRKCA</i>	5.25E-04
LPS-stimulated MAPK Signaling	<i>MAP2K4, PIK3C2B, PRKCZ, PRKCA</i>	5.89E-04
NF-κB Activation by Viruses	<i>PIK3C2B, AKT1, PRKCZ, PRKCA</i>	5.89E-04
VEGF Family Ligand-Receptor Interactions	<i>PIK3C2B, AKT1, PRKCZ, PRKCA</i>	6.76E-04
IL-12 Signaling and Production in Macrophages	<i>MAP2K4, PIK3C2B, AKT1, PRKCZ, PRKCA</i>	6.76E-04
PDGF Signaling	<i>MAP2K4, PIK3C2B, PDGFRA, PRKCA</i>	7.08E-04
Relaxin Signaling	<i>PIK3C2B, AKT1, GNB2L1, PRKCZ</i>	7.24E-04
eNOS Signaling	<i>PIK3C2B, NOSIP, AKT1, PRKCZ, PRKCA</i>	9.12E-04
TR/RXR Activation	<i>PIK3C2B, MTOR, AKT1, PFKP</i>	1.05E-03
α-Adrenergic Signaling	<i>GNB2L1, PRKCZ, PRKCA</i>	1.12E-03
Signaling by Rho Family GTPases	<i>MAP2K4, PIK3C2B, CDH1, GNB2L1, PRKCZ</i>	1.45E-03
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	<i>AKT1, PRKCZ, PRKCA, PTEN</i>	1.45E-03
nNOS Signaling in Neurons	<i>DLG2, PRKCZ, PRKCA</i>	1.91E-03
Nitric Oxide Signaling in the Cardiovascular System	<i>PIK3C2B, AKT1, PRKCZ, PRKCA</i>	1.91E-03
Role of NFAT in Regulation of the Immune Response	<i>PIK3C2B, AKT1, GNB2L1, GSK3B</i>	2.04E-03
Renin-Angiotensin Signaling	<i>MAP2K4, PIK3C2B, PRKCZ, PRKCA</i>	2.57E-03
Natural Killer Cell Signaling	<i>PIK3C2B, AKT1, PRKCZ, PRKCA</i>	2.69E-03

Anhang

Hepatic Fibrosis / Hepatic Stellate Cell Activation	<i>SMAD2, MYH6, PDGFRA, SMAD4, MYH7</i>	2.75E-03
Androgen Signaling	<i>GNB2L1, PRKCZ, PRKCA</i>	2.75E-03
Thrombopoietin Signaling	<i>PIK3C2B, PRKCZ, PRKCA</i>	3.02E-03
Wnt/Ca+ pathway	<i>AXIN1, GSK3B, PRKCA</i>	3.16E-03
IL-6 Signaling	<i>MAP2K4, PIK3C2B, ABCB1, AKT1</i>	3.24E-03
GM-CSF Signaling	<i>PIK3C2B, AKT1, GNB2L1</i>	4.17E-03
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	<i>MAP2K4, PIK3C2B, PRKCZ, PRKCA</i>	4.27E-03
Antiproliferative Role of Somatostatin Receptor 2	<i>PIK3C2B, GNB2L1</i>	4.37E-03
Cardiac β -adrenergic Signaling	<i>PPP2CA, GNB2L1, PPP2R1B</i>	5.25E-03
Macropinocytosis Signaling	<i>PIK3C2B, PRKCZ, PRKCA</i>	5.50E-03
Growth Hormone Signaling	<i>PIK3C2B, PRKCZ, PRKCA</i>	5.62E-03
Role of MAPK Signaling in the Pathogenesis of Influenza	<i>MAP2K4, AKT1, PRKCA</i>	5.62E-03
Actin Cytoskeleton Signaling	<i>PIK3C2B, MYH6, APC2, MYH7, APC</i>	5.62E-03
Melatonin Signaling	<i>MAP2K4, PRKCZ, PRKCA</i>	5.89E-03
Ephrin B Signaling	<i>AXIN1, GNB2L1</i>	6.61E-03
Prolactin Signaling	<i>PIK3C2B, PRKCZ, PRKCA</i>	6.61E-03
Synaptic Long Term Depression	<i>PPP2CA, PPP2R1B, PRKCZ, PRKCA</i>	6.61E-03
Axonal Guidance Signaling	<i>PIK3C2B, AKT1, GNB2L1, GSK3B, PRKCZ, PRKCA</i>	7.76E-03
Cell Cycle Control of Chromosomal Replication	<i>MCM4, CHEK2</i>	8.71E-03
Gap Junction Signaling	<i>PIK3C2B, AKT1, PRKCZ, PRKCA</i>	8.91E-03
3-phosphoinositide Biosynthesis	<i>PIK3C2B, PTPRJ, DUSP11, PTEN</i>	9.12E-03
GPCR-Mediated Nutrient Sensing in Enteroendocrine Cells	<i>PRKCZ, PRKCA</i>	9.77E-03
FGF Signaling	<i>PIK3C2B, AKT1, PRKCA</i>	1.00E-02
Dopamine-DARPP32 Feedback in cAMP Signaling	<i>PPP2CA, PPP2R1B, PRKCZ, PRKCA</i>	1.02E-02
Hepatic Cholestasis	<i>MAP2K4, ABCB1, PRKCZ, PRKCA</i>	1.02E-02
Apoptosis Signaling	<i>TP53, MAP2K4, PRKCA</i>	1.15E-02
Virus Entry via Endocytic Pathways	<i>PIK3C2B, PRKCZ, PRKCA</i>	1.15E-02
G Protein Signaling Mediated by Tubby	<i>GNB2L1</i>	1.20E-02
IL-1 Signaling	<i>MAP2K4, GNB2L1</i>	1.20E-02
VEGF Signaling	<i>PIK3C2B, AKT1, PRKCA</i>	1.26E-02
RhoGDI Signaling	<i>CDH1, GNB2L1, PRKCA</i>	1.29E-02
Ephrin Receptor Signaling	<i>AKT1, AXIN1, GNB2L1</i>	1.32E-02
PPARα/RXRα Activation	<i>MAP2K4, SMAD2, SMAD4, PRKCA</i>	1.41E-02
T Cell Receptor Signaling	<i>MAP2K4, PIK3C2B, PAG1</i>	1.45E-02
Amyotrophic Lateral Sclerosis Signaling	<i>TP53, PIK3C2B, GRIK4</i>	1.48E-02
Neuropathic Pain Signaling In Dorsal Horn Neurons	<i>PIK3C2B, PRKCZ, PRKCA</i>	1.55E-02
Cholecystokinin/Gastrin-mediated Signaling	<i>MAP2K4, PRKCZ, PRKCA</i>	1.58E-02
ERK/MAPK Signaling	<i>PIK3C2B, PPP2CA, PPP2R1B, PRKCA</i>	1.70E-02
Superpathway of Inositol Phosphate Compounds	<i>PIK3C2B, PTPRJ, DUSP11, PTEN</i>	1.86E-02
phagosome formation	<i>PIK3C2B, PRKCZ, PRKCA</i>	1.95E-02
Mechanisms of Viral Exit from Host Cells	<i>PRKCZ, PRKCA</i>	1.95E-02
Leukocyte Extravasation Signaling	<i>MAP2K4, PIK3C2B, PRKCZ, PRKCA</i>	2.04E-02
D-myo-inositol (1,4,5,6)-Tetrakisphosphate Biosynthesis	<i>PTPRJ, DUSP11, PTEN</i>	3.02E-02

D-myo-inositol (3,4,5,6)-tetrakisphosphate Biosynthesis	<i>PTPRJ, DUSP11, PTEN</i>	3.02E-02
GNRH Signaling	<i>MAP2K4, PRKCZ, PRKCA</i>	3.02E-02
NAD Biosynthesis from 2-amino-3-carboxymuconate Semialdehyde	<i>NMNAT3,</i>	3.63E-02
NAD Salvage Pathway III	<i>NMNAT3,</i>	3.63E-02
Phospholipase C Signaling	<i>GNB2L1, PRKCZ, PRKCA</i>	3.63E-02
D-myo-inositol-5-phosphate Metabolism	<i>PTPRJ, DUSP11, PTEN</i>	4.07E-02
3-phosphoinositide Degradation	<i>PTPRJ, DUSP11, PTEN</i>	4.17E-02
Calcium-induced T Lymphocyte Apoptosis	<i>PRKCZ, PRKCA</i>	4.47E-02
Aldosterone Signaling in Epithelial Cells	<i>PIK3C2B, PRKCZ, PRKCA</i>	4.57E-02
PXR/RXR Activation	<i>ABCB1, AKT1</i>	4.68E-02

Danksagung

Ich danke Prof. Stefan Aretz für die Übernahme des Doktorvaters im Verlauf der Arbeit und für die damit verbundene Betreuung.

Ich bedanke mich bei allen weiteren Mitgliedern der Prüfungskommission für die Begutachtung dieser Arbeit.

Ich möchte Per Hoffmann danken, der mir mit Rat und Tat in kritischen Momenten zur Seite stand.

Frau Steinke danke ich für die Bereitstellung dieses spannenden Themas und für die Zurverfügungstellung der Drittmittel.

Für die Einarbeitung im Labor möchte ich mich bei Marlies Sengteller, Helena Bär, Dietlinde Stienen und Sascha Fischer bedanken.

Ich danke der AG-Aretz, besonders Ning, Chris und Sophia, für den wissenschaftlichen Austausch und die schöne Zeit.

Mein besonderer Dank gilt meinen Freunden und meiner Familie.

Nicht zuletzt möchte ich meinem Freund danken, der mich über die Jahre hinweg immer wieder unterstützt und begleitet hat.

Ich danke der Deutschen Krebsforschung, allen Kooperationspartnern und den partizipierenden Patienten und ihren Familien.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass

- die vorgelegte Arbeit – abgesehen von den ausdrücklich bezeichneten Hilfsmitteln – persönlich, selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde.
- die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte unter Angabe der Quelle kenntlich gemacht sind.
- die vorgelegte Arbeit nicht bereits anderweitig als Dissertation eingereicht worden ist.
- für die Erstellung der vorgelegten Arbeit und/oder die Gelegenheit zur Promotion keine fremde Hilfe, insbesondere keine entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen wurde.

Bonn, März 2017

Katrin Kayser