



TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Tierzucht

**Etablierung und Anwendung der Exom-Sequenzierung
und eines NGS-LIMS zur Identifikation seltener
pathogener genetischer Varianten bei monogenen
Erkrankungen**

Elisabeth Graf

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende:

- Prof. Angelika Schnieke, Ph.D.

Prüfer der Dissertation:

1. Prof. Dr. Hans-Rudolf Fries
2. Prof. Dr. Thomas Meitinger

Die Dissertation wurde am 12.09.2018 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 07.05.2019 angenommen.

**Etablierung und Anwendung der
Exom-Sequenzierung und eines
NGS-LIMS zur Identifikation seltener
pathogener genetischer Varianten bei
monogenen Erkrankungen**

Elisabeth Graf

Für Max, Niels und Lina Charlotte.

Für Daniel.

Zusammenfassung

Die Identifizierung genetischer Veränderungen als Ursache humaner Erkrankungen hat eine herausragende Stellung in der Medizin. Sie ist die Grundlage zum Verständnis der Entstehung von Krankheiten und bildet die Basis zur Entwicklung neuer therapeutischer Ansätze.

Die bisherigen molekulargenetischen Untersuchungsmethoden, wie Kandidatengenanalysen und Kopplungsanalysen sind häufig zeit- und kostenintensiv. Die im Jahr 2005 eingeführten Sequenziertechniken der zweiten Generation hingegen bieten die Möglichkeit, gesamte Genome, Exome und Transkriptome in vergleichsweise kurzer Zeit und bei stetig fallenden Kosten zu analysieren. Die Hochdurchsatzsequenzierung erlangte in der letzten Dekade enorme Popularität und ist eine bahnbrechende Methodik zur direkten Identifizierung pathogener Genvarianten. Der Fokus der derzeitigen Forschung liegt dabei auf der Untersuchung der kodierenden genetischen Sequenz, dem Exom, als günstige und gleichzeitig erfolgversprechendste Anwendung.

Die vorliegende Arbeit beschreibt die Einführung einer automatisierten Exom-*Library*-Präparation in zwei Schritten, sowie die Einführung von Qualitätssicherung und Standardisierungen innerhalb der Laborprozesse. Mit der Entwicklung von Protokollen für eine Semi-Automation im ersten Schritt, konnte der Probandendurchsatz einer Exom-*Library*-Präparation von 12 auf 48 Proben erhöht und in einem weiteren Schritt, durch Optimierungen der Laborgeräte, auf bis zu 96 Proben aufskaliert werden. Durch die Einführung diverser Qualitätssicherungsschritte vor, während und nach der *Library*-Präparation und Sequenzierung, wird die Möglichkeit von Probenverwechslungen oder -kontaminationen minimiert. Eine Optimierung und Standardisierung der Qualitätskontrolle eingehender DNA-Proben führte zudem zur Herstellung von komplexen *Libraries* und vergleichbaren sowie reproduzierbaren Sequenzierdaten.

Der steigende Probandendurchsatz machte die Handhabung von Probeneingängen, Probenbearbeitungen sowie der Ergebnisdarstellung sehr komplex. Die klassische Nutzung von Excel-Tabellen zur Dokumentation und Informationsweitergabe war dabei nicht mehr möglich. Als Anwendung zur Unterstützung administrativer Aufgaben der Probenbearbeitung beschreibt die vorliegende Arbeit die Entwicklung eines Labor-Informations- und Management-Systems (LIMS), welches zum Beginn der Exom-Sequenzierung am Institut für Humangenetik für die

Next-Generation-Sequenzierung nicht zur Verfügung stand. Die web-basierte Anwendung unterstützt die Laborarbeit auf den Ebenen des Probeneingangs, der Probenbearbeitung und der Qualitätskontrolle im Hochdurchsatz. Das LIMS wurde weiterhin durch die Einführung eines *sample manifest* für den Probeneingang, eines *sample tracking sheet* für die Probenprozessierung, sowie einem 2D-Barcode System und der Kontaminationsanalyse in seiner Funktionsweise erweitert und verfeinert.

Die Evolution der Laborfertigkeiten und des LIMS führte zu einer erfolgreichen Anwendung der NG-Sequenzierung am Institut für Humangenetik. Die Automation wurde genutzt, um ein Protokoll für die *Library*-Präparation von FFPE-DNA zu entwickeln und daraus abgeleitet, Kriterien an das Ausgangsmaterial für eine erfolgreiche Sequenzierung zu definieren. Die Semi-Automation und das LIMS wurde erstmals im Jahr 2011 zur Erkennung von kausativen Varianten bei monogenen Erkrankungen eingesetzt. Seither folgten weitere Studien am Institut für Humangenetik und die Exom- und Genom-Sequenzierung entwickelte sich zu einer Routineanwendung. In der vorliegenden Arbeit werden exemplarisch zwei Studien vorgestellt. In der ersten Studie aus dem Jahr 2011 wurde durch eine Exom-Sequenzierung die ursächliche genetische Variante c.1858G>A, p.Asp620Asn im Gen *VPS35* bei drei Familien mit Morbus Parkinson identifiziert. Alle Patienten leiden unter einer sich spät manifestierenden Form des Morbus Parkinson mit einem mittleren Erkrankungsalter von 53 Jahren. Die Betroffenen tragen einen gemeinsamen Haplotyp von 65 kb, der das *VPS35*-Gen überspannt. Computergestützte sequenzbasierte Analysen, sowie Untersuchungen zur Moleküldynamik, prognostizieren die familiäre Variante p.Asp620Asn als schädlich. In einer zweiten Studie aus dem Jahr 2014 werden exemplarisch die Grenzen der Exom-Sequenzierung aufgezeigt. Genetische Varianten, die außerhalb der kodierenden Bereiche liegen, können mit der Exom-Sequenzierung nicht erkannt werden. So konnten nur durch eine Kombination aus Exom- und Genom-Sequenzierung bei Patienten mit Burn-McKeown-Syndrom biallele Mutationen im Gen *TXNL4A* als ursächlich für die Erkrankung identifiziert werden. In neun von elf untersuchten Familien fanden sich Veränderungen in *TXNL4A*. Wobei in acht Familien seltene *loss-of-function*-Mutationen (*nonsense*, *frameshift* oder Mikrodeletionen) auf dem einen Allel, in Kombination mit einer 34 bp-Deletion in der Kern-Promotorregion auf dem zweiten Allel, auftraten. In einer konsan-

guinen Familie sind alle vier Betroffene homozygot für eine abweichende 34 bp-Promotordeletion in *TXNL4A*. Die Erkennung der Promotordeletion wurde erst durch eine Genom-Sequenzierung möglich, da diese Bereiche durch die Exom-Sequenzierung nicht repräsentiert wurden. Funktionelle Studien durch die Projektpartner konnten zeigen, dass die Promotordeletionen zu einer verminderten *TXNL4*-Expression führen. *TXNL4A* ist eine essentielle Untereinheit des U5-Spleißkomplex. Untersuchungen an Hefe nach Depletion von *TXNL4A* (*Dib1*) resultierten in einer reduzierten Assemblierung des tri-snRNP Komplexes. Die Ergebnisse zeigen, dass das Burn-McKeown-Syndrom eine autosomal-rezessive Erkrankung ist, die primär durch *compound*-heterozygote-Promotordeletionen in Kombination mit sehr seltenen *loss-of-function*-Mutationen hervorgerufen wird.

Inhaltsverzeichnis

Zusammenfassung	6
1 Einleitung	14
1.1 Genetische Erkrankungen	14
1.2 Identifizierung genetischer Varianten	15
1.3 <i>Next-Generation</i> -Sequenzierung	17
1.3.1 Illuminas <i>Sequencing-by-Synthesis</i>	18
1.3.2 Exom-Sequenzierung	21
1.3.2.1 Exom-Anreicherungstechniken	21
1.4 Labor-Informations- und Management-System	23
1.5 Bioinformatik	24
1.6 Strategien zur Identifizierung krankheitsverursachender genetischer Varianten aus Exom-Daten	28
1.6.1 Priorisierungsstrategie bei familiären Erkrankungen	28
1.6.2 <i>de novo</i> -Strategie bei Eltern-Kind-Trios	29
1.6.3 Überlappungsstrategie in Patientengruppen mit gleichem Phänotyp	30
1.6.4 Homozygotie-Strategie bei konsanguinen Eltern	30
1.6.5 Priorisierungsstrategie bei rezessiven Erkrankungen	31
1.6.6 Kandidatengen-Strategie bei individuellen Patienten	31
1.7 Zielsetzung der Arbeit	33
2 Material und Methoden	34
2.1 Material	34
2.1.1 Probenmaterial	34
2.1.2 Geräte	34
2.1.3 Verbrauchsmaterialien	35
2.1.4 Software	35
2.1.5 Internetquellen - URLs	36
2.2 Molekularbiologische Methoden	36
2.2.1 DNA-Isolation und Konzentrationsbestimmung	36
2.2.2 Oligonukleotide	36
2.2.3 Amplifikation von DNA-Fragmenten mittels PCR	36

2.2.4	Agarose-Gelelektrophorese	37
2.2.5	Aufreinigung der PCR-Produkte für eine anschließende Sanger-Sequenzierung	37
2.2.6	Sanger-Sequenzierung	38
2.3	Mutationsscreening	40
2.3.1	IDAHO-Screening	40
2.4	<i>Next-Generation</i> -Sequenzierung	41
2.4.1	Genom-Sequenzierung	41
2.4.2	Exom-Sequenzierung	42
2.4.3	Hochdurchsatz-Sequenzierplattform Illumina, Inc.	43
2.4.4	Bioinformatik	43
3	Ergebnisse	45
3.1	Entwicklung eines Labor-Informations- und Management-Systems für die <i>Next-Generation</i> -Sequenzierung	46
3.1.1	Die Anforderungen an ein LIMS	46
3.1.2	Die Probenebene	48
3.1.3	Die Prozessebene	51
3.1.4	Die Qualitätssicherung innerhalb der Analyseebene	56
3.1.5	Erweiterung und Entwicklung des LIMS durch zusätzliche Anwendungen	57
3.1.5.1	Einführung des <i>Sample Manifest</i>	58
3.1.5.2	Einführung des <i>Sample Tracking Sheets</i>	59
3.1.5.3	Einführung von 2D-Barcode Röhrchen	59
3.1.5.4	Rekalkulation von genomischen <i>Library-Pools</i>	63
3.1.5.5	Einführung der Kontaminationsbestimmung bei humane Proben	65
3.2	Etablierung der automatisierten Exom- <i>Library</i> -Präparation	70
3.2.1	Einführung der manuellen Exom- <i>Library</i> -Präparation	70
3.2.2	Einführung der semi-automatisierten Exom- <i>Library</i> -Präparation	71
3.2.2.1	Automatisierte AMPure XP <i>Bead</i> -basierte Aufreinigung	75
3.2.2.1.1	Prozessschritte der AMPure XP <i>Bead</i> -Aufreinigung auf dem Bravo	79

3.2.2.1.2	Vergleich der manuellen und automatisierten AMPure XP <i>Bead</i> -Aufreinigung	81
3.2.2.2	Automation der Exom-Anreicherung und der Elution	83
3.2.2.2.1	Automatisierte Prozessschritte der Exom-Anreicherung und Elution auf dem Bravo	84
3.2.2.3	Anwendung der automatisierten AMPure XP <i>Bead</i> -Aufreinigung zur Herstellung von genomischen <i>Libraries</i>	88
3.2.2.4	Kontrollpunkte während der Automation	88
3.2.2.5	Automationsprotokolle und <i>Lab Tracking</i> -Formulare	89
3.2.3	Einführung einer weitgehenden Automation zur Herstellung von exomischen <i>Libraries</i>	92
3.2.4	Qualitätskontrolle und Prozessoptimierung der <i>Library</i> -Präparation	93
3.2.4.1	DNA-Qualitätskontrolle	93
3.2.4.1.1	Bestimmung der DNA-Konzentration	93
3.2.4.1.2	Bestimmung der Reinheit von DNA-Proben	95
3.2.4.1.3	Integritätskontrolle von DNA-Proben	97
3.2.4.2	Ergebnis der Exom-Sequenzierung von eingesandter DNA versus intern-präparierter DNA	99
3.2.4.2.1	Ableitung der Anforderungen an eingesandtes Material für eine Exom-Sequenzierung	102
3.2.4.3	Optimierung des Laborablaufs zur Steigerung des Durchsatzes	103
3.2.4.4	Entwicklungen von Methodiken zum Ausschluss von Probenverwechslungen und Kontaminationen	105
3.2.5	Austestung einer Exom- <i>Library</i> -Präparation mit weniger Einsatzmenge und größerer <i>Insert</i> -Größe	108
3.2.5.1	Hintergrund	108
3.2.5.2	Methodik und Ergebnisse	109

3.2.5.3	Effekt der gesteigerten <i>Insert</i> -Größe und DNA-Einsatzmenge auf die Qualitätsparameter der Sequenzierung	112
3.2.6	Evolution der <i>NGS</i> -Plattform am Institut für Humangenetik	114
3.3	Anwendung der <i>Next-Generation</i> -Sequenzierung	118
3.3.1	Anwendung der etablierten Automation zur Herstellung von <i>Exom-Libraries</i> aus FFPE-DNA	119
3.3.1.1	Hintergrund	119
3.3.1.2	Methodik und Ergebnisse	120
3.3.1.2.1	Einfluss der DNA-Qualität auf die <i>Library</i> -Präparation	120
3.3.1.2.2	Sequenzierung von Exomen aus FFPE-DNA	125
3.3.2	Exom-Sequenzierung zur Identifizierung pathogener genetischer Varianten bei Patienten mit spät beginnendem Morbus Parkinson	127
3.3.2.1	Zusammenfassung	128
3.3.2.2	Hintergrund	128
3.3.2.3	Patienten	130
3.3.2.4	Methodik und Ergebnisse	130
3.3.3	Exom- und Genom-Sequenzierung zur Identifizierung pathogener Mutationen bei Patienten mit Burn-McKeown-Syndrom	137
3.3.3.1	Zusammenfassung	137
3.3.3.2	Hintergrund	138
3.3.3.3	Probandenauswahl	140
3.3.3.4	Exom-Sequenzierung	140
3.3.3.5	Genom-Sequenzierung	142
4	Diskussion	148
4.1	Entwicklung eines LIMS für die <i>Next-Generation</i> -Sequenzierung	149
4.1.1	Weiterentwicklung des <i>NGS</i> -LIMS	150
4.2	Einführung einer automatisierten <i>Exom-Library</i> -Präparation	154
4.2.1	Qualifikationsablauf und Ableitung der Anforderungen an DNA-Material für die <i>NGS</i>	158

4.2.2	Methodiken zum Ausschluss von Probenverwechslungen und Kontaminationen bei der <i>NG</i> -Sequenzierung	160
4.3	Anwendung des LIMS und der Automation	161
4.3.1	Entwicklung neuer Automationsprotokolle für spezielle Anwendungen	161
4.3.2	Exom-Sequenzierung bei Patienten mit Morbus Parkinson	166
4.3.3	Exom- und Genom-Sequenzierung bei Patienten mit Burn-McKeown-Syndrom	169
4.4	Fazit und Ausblick	172
Literaturverzeichnis		175
Abbildungsverzeichnis		208
Tabellenverzeichnis		211
Abkürzungsverzeichnis		212
5	Anhang	214
5.1	Protokoll zur automatisierten Exom- <i>Library</i> -Präparation auf der <i>Bravo Liquid Handling Station</i>	214
5.2	Protokoll zur automatisierten Genom- <i>Library</i> -Präparation auf der <i>Bravo Liquid Handling Station</i>	241
5.3	<i>Lab Tracking</i> -Formulare für die automatisierten Exom- <i>Library</i> -Präparation auf der <i>Bravo Liquid Handling Station</i>	259
5.4	<i>Lab Tracking</i> -Formulare für die automatisierten Genom- <i>Library</i> -Präparation auf der <i>Bravo Liquid Handling Station</i>	267
5.5	Primer	273
5.5.1	Primer für <i>VPS35</i>	273
Danksagung		274

1 Einleitung

Das diploide humane Genom besteht aus 6×10^9 Nukleotiden. Deren Zusammensetzung bildet die Grundlage für die Individualität und die Vielfalt der menschlichen Spezies. Wenigstens sechs Prozent der genetischen Sequenz stand in den vergangenen 100 Millionen Jahren unter hohem evolutionärem Selektionsdruck und wird als biologisch funktional erachtet. Tatsächlich proteinkodierend sind hingegen nur circa 1,5 % der genetischen Sequenz, während die verbleibenden 4,5 % nicht-kodierende funktionelle Elemente darstellen [1]. Das Ziel der humangenetischen Forschung ist die Untersuchung dieser Sequenzen und die Erlangung des Verständnisses darüber, ob eine Veränderung dieser Abschnitte zu Gesundheit oder Krankheit führt. Die Identifizierung krankheitsverursachender Gendefekte hat eine herausragende Stellung in der humangenetischen Diagnostik. Erst durch das Wissen um eine pathogene Genvariante kann der Patient aufgeklärt, therapiert und eine Prognose gestellt werden. Die genetische Diagnose einer Erkrankung unterstützt die Beratung des Patienten und dessen Familienangehörigen, gibt Einblicke in die physiologische Rolle der betroffenen Gene und Genprodukte und gibt Hinweise auf die zugrunde liegenden pathogenen Signal- und Stoffwechselwege [2].

1.1 Genetische Erkrankungen

Genetische Erkrankungen entstehen durch Veränderungen im Genom, die ererbt oder durch Neumutationen erworben sein können. Die Alterationen betreffen die DNA-Sequenz *per se* oder die chromosomalen Strukturen in Form von numerischen Änderung der Chromosomenzahl oder Chromosomenbrüchen. Genetische Erkrankungen können anhand der Anzahl der ursächlichen Gene unterschieden werden: Monogene Erkrankungen sind gekennzeichnet durch Veränderungen in nur einem Gen. Sie folgen den klassischen mendelschen Vererbungsregeln im autosomal-dominanten (zum Beispiel Chorea Huntington, Marfan-Syndrom oder Neurofibromatose Typ 1 und 2.), autosomal-rezessiven (zum Beispiel Sichelzellanämie, Cystische Fibrose und Spinale Muskelatrophie) oder X-chromosomalem (zum Beispiel Hämophilie A und B, Mukopolysaccharidose Typ II, Rot-Grün-Blindheit oder Retinitis pigmentosa) Vererbungsmodus. Die meisten monogenen Erkrankungen sind eher selten und betreffen eine Person unter mehreren 1.000

gesunden Menschen.

Das Risiko von häufigen Erkrankungen, wie Morbus Crohn, Adipositas oder Diabetes Mellitus, kann durch mehrere genetische Varianten beeinflusst sein [3]. Hier spricht man von multigener oder komplexer Vererbung [4].

Der einfachste Weg zum Verständnis einer Genfunktion ist die Untersuchung von vererbbaaren, pathogenen Varianten bei monogenen Erkrankungen [5]. In der online Datenbank OMIM (*Online Mendelian Inheritance in Men*) [6] sind derzeit 6.801 monogene Erkrankungen beschrieben (Stand April 2018). Die zurückliegenden Strategien der letzten Jahrzehnte, zur Identifizierung von pathogenen Varianten, wie Kandidatengenanalysen und Kopplungsanalysen, führten zur Identifikation der genetischen Ursache bei fast der Hälfte aller mendelschen Erkrankungen [3, 7]. Gegenwärtig ist bei 5.218 der beschriebenen monogenen Erkrankungen die molekulare Basis bekannt. Die verbleibenden 1.583 Phänotypen unterliegen der mendelschen Vererbung mit bisher unbekannter molekularer Basis.

Zur Aufklärung genetischer Veränderungen stand seit vielen Jahren die Sanger-Sequenzierung humaner DNA-Abschnitte als Methodik an oberster Stelle. Hat man früher für die Identifizierung einer pathogenen Veränderung zum Teil einige Jahre gebraucht, ist es nun möglich, mit der Einführung der Hochdurchsatzsequenzierung, ein gesamtes Genom innerhalb weniger Tage zu sequenzieren und zu analysieren [2]. Die Techniken der Hochdurchsatzsequenzierung werden unter dem Begriff *Next-Generation-Sequenzierung* (*NGS*) zusammengefasst [8].

1.2 Identifizierung genetischer Varianten

Vor dem Beginn einer genetischen Analyse steht die traditionelle medizinische Untersuchung. Mediziner erkennen die klassischen Symptome einer Erkrankung und nutzen die gezielte genetische Untersuchung zur Bestätigung der Eingangsd Diagnose. Konventionelle molekulargenetische Untersuchungen beinhalten Kandidatengenanalysen und Kopplungsanalysen.

Kandidatengenanalyse mittels Sanger-Sequenzierung. Die Kandidatengenanalyse basiert auf dem Wissen über die biologische Funktion eines Gens in Zusammenhang mit dem Phänotyp eines Patienten. Entsprechend der Erkrankung ist das Genprodukt und dessen Funktion im beteiligten Stoffwechselweg lokalisiert und/oder funktionell integriert. Um pathogene Alterationen zu identifizieren, werden die entsprechenden Gene mittels Sanger-Sequenzierung unter-

sucht.

Kopplungsanalysen. Werden genetische Loci oder Allele gemeinsam mit einem Phänotyp vererbt, spricht man von genetischer Kopplung. Je näher dabei zwei Merkmale auf einem Chromosom zusammen liegen, desto höher ist die Wahrscheinlichkeit, dass sie gemeinsam vererbt werden. Durch chromosomale Rekombination während der Meiose können gekoppelte Allele voneinander getrennt werden. In Familien mit monogenen Erkrankungen kann mittels Kopplungsanalyse die genomische Region, welche eine mögliche pathogene Variante enthält, eingegrenzt werden. Eine Kopplungskarte gibt dabei die Information über die gemeinsamen genetischen Bereiche in den Betroffenen. Das statistische Kriterium zur Auswertung ist der *LOD score* [9]. Er gibt eine Aussage über die Wahrscheinlichkeit, dass eine genetische Region tatsächlich gekoppelt vererbt wird, oder nur durch Zufall zusammen auftritt. Ist eine gekoppelte Region identifiziert, welche ausschließlich in den Erkrankten einer Familie kosegregiert, kann basierend darauf eine Kandidatengenanalyse für die betreffenden Gene in dieser Region durchgeführt werden.

Die beiden klassischen Strategien zur Identifizierung genetischer Varianten limitieren die Möglichkeiten der Suche nach krankheitsassoziierten Genen [5]. Mit der zeit- und kostenintensiven Kandidatengenanalyse können nur kleine genomische Bereiche abgedeckt und wenige Gene sequenziert werden. Die Kopplungsanalyse ist eine indirekte Methode, für die möglichst große Familien mit mehreren Betroffenen für eine Untersuchung zur Verfügung stehen sollten. Es konnte Monate bis hin zu Jahren dauern, bis eine genetische Diagnose gestellt wurde [10, 11], wobei selbst nach eingehender, voll ausgeschöpfter Methodologie, in lediglich circa 50 % der medizinischen Fälle eine genetische Aufklärung gelang [12]. Bei Erkrankten ohne Diagnose kann es zum Versäumnis von Therapieansätzen kommen und das klinische Management schlechter ausgeführt werden. Für Familien ist eine gesicherte Diagnose erleichternd und unterstützt den behandelnden Arzt bei der Beratung, Prognose und Therapie [13, 7].

Die Entwicklung der Hochdurchsatzsequenzierung vermag die Defizite der klassischen genetischen Diagnostik zu schließen. Mit der *NGS*-Technik ist es möglich, bei weiterhin fallenden Kosten, ein gesamtes Genom innerhalb weniger Tage zu sequenzieren und zu analysieren. Dabei erlaubt sie die direkte Identifizierung einer Variante im Vergleich zu Kopplungsstudien. Vor allem bei heterogenen

Erkrankungen oder der Suche nach *de novo*-Mutationen, wie zum Beispiel bei Entwicklungsstörungen des Nervensystems, sowie bei Krebserkrankungen ist die *NG*-Sequenzierung erfolgversprechend.

Die Hochdurchsatzsequenzierung erlangte in der letzten Dekade enorme Popularität und ist eine bahnbrechende Methodik zur Identifizierung pathogener Genvarianten.

1.3 *Next-Generation-Sequenzierung*

Die Sanger-Sequenzierung mittels Kettenabbruchmethodik [14] war der bisherige Goldstandard zur Bestimmung der Basenabfolge eines DNA-Abschnitts [15]. Der Durchsatz bei der Sanger-Sequenzierung bleibt gering. Für jedes zu untersuchende DNA-Fragment wird eine separate PCR und Sequenzreaktion benötigt [15]. Der Parallelisierung sind dadurch Grenzen gesetzt. Die im Jahr 2005 vorgestellte Technik der Hochdurchsatzsequenzierung - *Next-Generation-Sequenzierung* (*NGS*), hat die Besonderheit, durch eine hohe Parallelisierung, gesamt-exomische, gesamt-transkriptomische und gesamt-genomische Analysen, in vergleichsweise kurzer Zeit und mit weiterhin sinkendem finanziellen Aufwand, für individuelle Patienten durchzuführen. Der Begriff *Next-Generation-Sequenzierung* umfasst verschiedenste Sequenziermethoden der zweiten und dritten Generation mit den unterschiedlichsten Sequenzierplattformen, von denen eine ganze Reihe kommerziell zur Verfügung stehen oder noch in der Entwicklung befindlich sind [8]. Die zugrunde liegenden Techniken und Chemikalien unterscheiden sich dabei zum Teil erheblich voneinander. Daraus resultieren unterschiedliche Leselängen mit unterschiedlichen Fehlerraten sowie abweichende Beladungskapazitäten. Der Durchsatz ist jedoch bei allen *NGS*-Anwendungen um ein vielfaches größer als bei der Sanger-Sequenzierung [2]. Die Technik ermöglicht die Detektion von Punktmutationen (*single nucleotide variants*, *SNVs*) über Insertionen und Deletionen bis hin zu größeren strukturellen Veränderungen (genomische *rearrangements*) in einer Analyse.

Von allen *NGS*-Anbietern ist gegenwärtig die gebräuchlichste Plattform, die der Firma Illumina (Illumina Inc.). 90 % aller bisher publizierten *NGS*-Daten wurden auf den Sequenzierplattformen der Firma Illumina generiert [16]. Die *Sequencing-by-Synthesis*-Sequenzier Technologie (*SBS*) [17, 18] bietet einen bisher nicht übertroffenen Durchsatz, Genauigkeit, Geschwindigkeit und Skalierbarkeit,

was zu einer signifikanten Reduktion der Kosten führt [19]. Die *SBS*-Technologie und die ersten Sequenziermaschinen wurde von der Firma Solexa entwickelt und 2007 von Illumina gekauft. Das erste Modell der Firma Illumina war der *Genome Analyzer* (GA) mit dem in einem Sequenzier-Lauf 1 Gb Daten erzeugt wurden. Die produzierte Datenmenge steigerte sich mit jedem neu eingeführten Modell bis hin zu 6 TB pro Lauf auf dem aktuellen NovaSeq. Mit dem HiSeqX kann ein humanes Genom für circa 1.000 US-Dollar sequenziert werden [15].

1.3.1 Illuminas *Sequencing-by-Synthesis*

Von der Nukleinsäure eines Patienten bis zur Basenabfolge sind drei Schritte notwendig: die Herstellung der DNA-Bibliothek (*Library-Präparation*) (Abb.: 1), die Generierung von klonalen *Clustern* auf der *Flow Cell* (*cluster generation*) (Abb.: 2) und die eigentlichen Sequenzierung (Abb.: 3). Im ersten Schritt, der Präparation der *Library*, wird die DNA mechanisch fragmentiert und die Enden der entstandenen Fragmente so modifiziert, dass plattformsspezifische Oligonukleotid-Adapter an beide Fragmentenden ligiert werden können. Die Adapter sind doppelsträngige Oligonukleotide, welche die Basenabfolge der Sequenzierprimer, sowie komplementäre Bereiche zu den Oligonukleotiden auf der *Flow Cell* enthalten. Gleichzeitig sind in den Adapter-Sequenzen individuelle Indizes enthalten, welche aus 6 oder mehr Nukleotiden bestehen und das Multiplexen von mehreren Proben erlauben. Die modifizierten DNA-Moleküle werden zusammen als *DNA-Library* bezeichnet [15]. Vor der eigentlichen Sequenzierung wird die *Library* amplifiziert um eine möglichst große Anzahl identischer Moleküle für die photographische Detektion zu generieren. Diese sogenannte *Cluster Generation* findet mittels einer *Bridge-Amplification* auf der *Flow Cell* statt. Dabei wird aus einem einzelnen DNA-Fragment, durch eine isothermale PCR, eine DNA-Kolonie (*Cluster*) erzeugt (Abb.: 2) [20]. Diese dienen als Template für das eigentliche *Sequencing-by-Synthesis*, welches im Anschluss an die *Cluster Generation* erfolgt (Abb.: 3) [18].

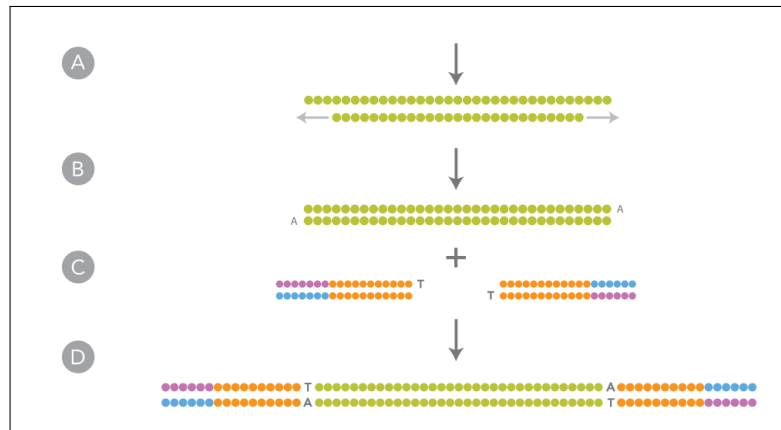


Abbildung 1: Schematische Darstellung der *Library*-Präparation [21]. (A) Die kohäsiven Enden des fragmentierten genomische DNA-Doppelstrangs, in grün dargestellt, werden in der *End-Repair* aufgefüllt. (B) Ligation eines A-Nukleotids als 3'-Überhang an die fragmentierte genomische DNA als Vorbereitung zur Adapter-Ligation. (C) Ligation der plattformspezifischen Adapter. Die orangefarbenen Sequenzen enthalten die Bindestellen für die Sequenzier-Primer. Die in blau und lila dargestellten Abschnitte sind komplementär zu den Oligonukleotiden auf der *Flow Cell*. (D) Struktur eines *Library*-Moleküls.

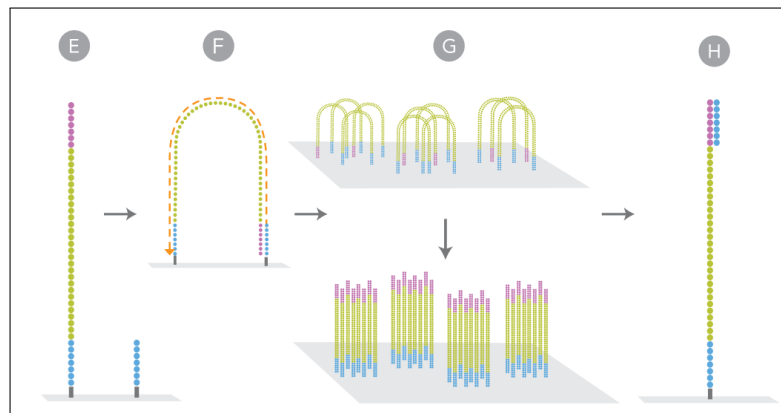


Abbildung 2: *Cluster Generation* [21]. In grün ist das DNA-Insert dargestellt. Die Adapter-Sequenzen sind blau und lila markiert. Zusammen bilden das DNA-Insert und die beiden Adapter ein einzelsträngiges *Library*-Molekül. Die Anreicherung der *Library* erfolgt als *Bridge-Amplifikation* auf einer Durchflusszelle (*Flow Cell*). (E) Die *Flow Cell* enthält auf ihrer Oberfläche immobilisierte Oligonukleotide, die komplementär zu den an die DNA gebundenen Adaptern sind. (FG) Nach der Hybridisierung der DNA-Fragmente an die Oligonukleotide findet die klonale Amplifikation zu einem Cluster statt. (H) In der folgenden Denaturierung werden die reversen Stränge entfernt und an die verbleibenden Einzelstränge Sequenzierprimer hybridisiert.

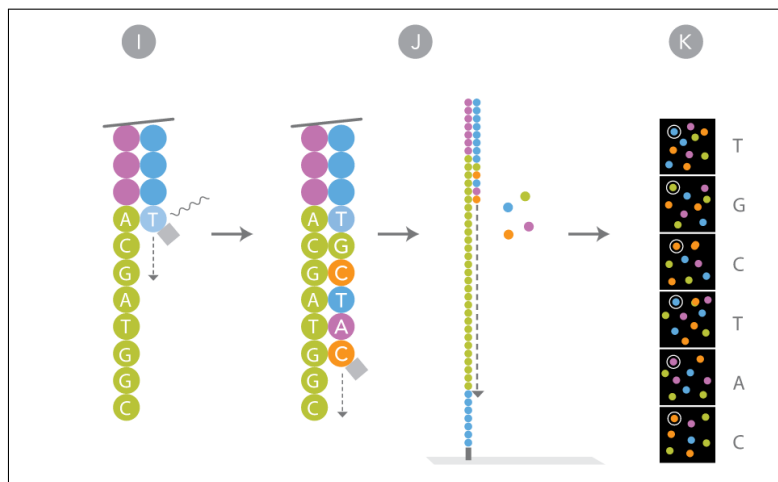


Abbildung 3: *Sequencing-by-Synthesis* [21]. In grün ist die Nukleotidsequenz des zu sequenzierenden DNA-Inserts dargestellt. Der einzelsträngige Adapter in lila wird durch den Sequenzier-Primer in blau gebunden. Bei der eigentlichen Sequenzieretechnologie handelt es sich um ein *Sequencing-by-Synthesis*-Verfahren (SBS). (I) Dabei werden parallel alle vier mit Farbstoffen markierte dNTPs über die Cluster gespült (in blau, grün, orange und lila dargestellt). Die Besonderheit ist hierbei, dass die dNTPs einen terminalen 3'-Blocker besitzen, der den Einbau von nur einem Nukleotid erlaubt. Nicht inkorporierte Nukleotide werden weg gewaschen und das zum DNA-Fragment komplementäre Nukleotid wird über den gekoppelten Farbstoff photographisch detektiert. (J) Im Anschluss werden die terminalen Blocker abgebaut und ein neuer Zyklus kann beginnen. (K) Die Nukleotidsequenz wird im anschließenden *Base Calling* erzeugt.

1.3.2 Exom-Sequenzierung

Zwar ist es in der Ära der *Next-Generation*-Sequenzierung möglich ein gesamtes Genom innerhalb weniger Tage zu sequenzieren, dennoch ist der routinemäßige Hochdurchsatz für eine Genom-Sequenzierung mit hohen Kosten verbunden. Daher fokussiert sich die aktuelle Forschung vor allem auf die proteinkodierenden Bereiche des Genoms (1,5 % [1, 3]), die in ihrer Gesamtheit als Exom bezeichnet werden.

Neben dem Kostenfaktor stellen vor allem zwei Punkte eine Exom-Sequenzierung als erfolgreich in Aussicht: a) Die bisherige Erfahrung zeigt eine hohe Erfolgsrate bei Kandidatengenanalysen, die kausative Variante zu identifizieren [5]. b) Circa 85 % der bekannten pathogenen Allele, die einer mendelschen Erkrankung zugeordnet sind, liegen in proteinkodierenden Bereichen [22, 23]. Veränderungen in kodierenden Bereichen können besser erklärt werden als Mutationen in einem wenig bis gar nicht charakterisierten inter- oder intragenischen Bereich [24]. Somit repräsentiert die Exom-Sequenzierung einen aussichtsreichen Ansatz zur Identifizierung von genetischen Veränderungen mit hoher Effektgröße. Die Exom-Sequenzierung wurde im Jahr 2010 erstmals zur Identifizierung genetischer Alterationen bei monogenen Erkrankungen angewendet [25].

1.3.2.1 Exom-Anreicherungstechniken

Kurz nach dem Einzug der *NG*-Sequenzierertechnologie, wurden Methoden entwickelt um das Exom anzureichern [26, 27, 28, 29, 30, 31, 32, 33]. Die gebräuchlichsten Anreicherungstechniken finden mittels Hybridisierungsverfahren in Lösung statt [28, 34]. Die populärsten Systeme sind: Roche NimbleGen mit dem *SeqCap EZ System* [35], Illumina mit dem *Nextera Rapid Capture Exome Kit* [36] und Agilent mit dem *SureSelect XT Human All Exon* [28] Anreicherungs-kit. Die *Library*-Präparation aller Kits ist sehr ähnlich (Abb.: 4). Grundlegend für alle Methoden, ist die Hybridisierung der zuvor präparierten *DNA-Library* an einzelsträngige, biotinylierte *DNA-* oder *RNA-Baits*. Die *Baits* sind 60 bis 120 Basen lange Oligonukleotide, die komplementär zu den exomischen Sequenzen sind [37]. Im Anschluss an die Hybridisierung erfolgt eine selektive Rückgewinnung durch die Aufreinigung mit streptavidingekoppelten magnetischen *Beads*. Alle nicht hybridisierten Fragmente werden dabei weg gewaschen, wohingegen alle angereicherten genomischen Abschnitte final eluiert und amplifiziert werden.

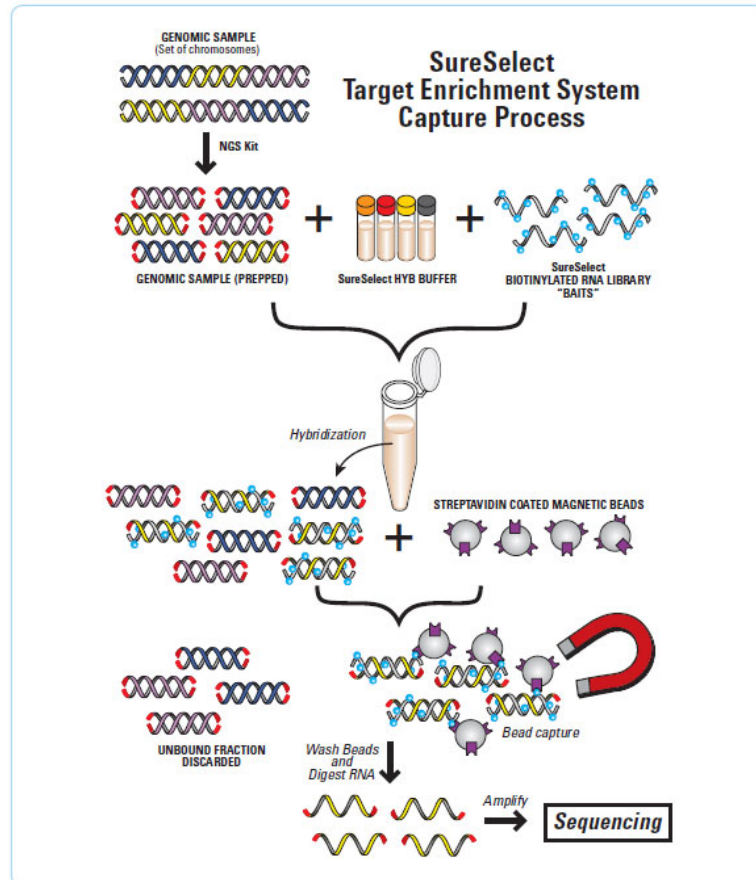


Abbildung 4: Übersicht zur Exom-Anreicherungstechnik der Firma Agilent [38]. Die hochmolekulare genomische DNA wird im ersten Schritt durch fokussierte Schallenergie auf 150 bis 600 bp fragmentiert. Im Anschluss werden die Enden der DNA-Fragmente modifiziert und Adapter ligiert. Es folgt die Hybridisierung an RNA-Baits und eine Selektion der DNA/RNA-Hybride über magnetische streptavidingekoppelte Beads. Die Angereicherten DNA-Moleküle werden amplifiziert, wobei probenspezifische Indizes eingebaut werden. Im Anschluss an die Qualitätskontrolle und das Pooling kann die angereicherte DNA-Library sequenziert werden.

Tabelle 1: Vergleich der gegenwärtig aktuellen Anreicherungskits. [42, 43]

Kit	Zielregion [Mb]	Anzahl der Baits	Bait Art	DNA Einsatzmenge [ng]	Adapter Addition durch	Bait Länge	RefSeq [Mb]	UTR [Mb]
SureSelect XT v.6	60	758.086	RNA	200-3000	Ligation	120 mer	k.A.	k.A.
SureSelect XT v.5	51	655.872	RNA	200-3000	Ligation	120 mer	30	3,6
Illumina Nextera Rapid Capture v1.2	45	>340.000	DNA	50	Transposase	95 mer	33,5	3,4
Roche Nimblegen SeqCap EZ Exome v3.0	64	>2.100.000	DNA	1000	Ligation	60-90 mer	33,5	8,7

In der Tabelle 1 sind die Parameter der gegenwärtig aktuellen Anreicherungskits vergleichend dargestellt. Sie unterscheiden sich in der Art, Länge und Dichte der zur Anreicherung genutzten *Baits*. Obwohl sich alle drei Kits an den Datenbanken RefSeq [39], Ensembl [40] und CCDS [41] zur Bestimmung der Zielsequenz orientieren, gibt es Unterschiede in der Definition der *Baits*. Zusätzlich zu den kodierenden Regionen decken alle Kits mehr oder weniger große UTR-Bereiche ab. Alle Techniken erzielen eine hohe Abdeckung in den gewünschten Regionen und zeigen eine hohe technische Reproduzierbarkeit. Letztlich obliegt es dem Anwender für welches Kit er sich entscheidet. Ausschlaggebend können hier die Qualität der Daten, die Kosten und die zu investierende Zeit für eine *Library*-Präparation sein. In dieser Studie wurde mit den *SureSelect Human all Exon* Kits als Standardanreicherungstechnik gearbeitet.

1.4 Labor-Informations- und Management-System

Ein **Labor-Informations- und Management-System** (LIMS) wird zur Unterstützung administrativer Aufgaben der Probenbearbeitung innerhalb des Laborbetriebs eingesetzt. Es dient der Erfassung, Auswertung und Präsentation von Analysedaten. Zeitgleich werden alle anfallenden Daten in standardisierter und leicht zugänglicher Form gespeichert [44, 45].

Seine Wurzel hat das LIM-System in der Labor-Automation und wurde 1973 zum ersten Mal auf der *Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopie* diskutiert. [46, 47]. Perkin-Elmer publizierte als erste Firma ihr kommerzielles *LIMS 2000* im Jahr 1983. Mit ihm war es möglich Daten verschiedener Labor-Automationen zu sammeln, daraus Ergebnisse zu berechnen, graphisch darzustellen und zu drucken. Gleichzeitig konnte die gegenwärtige Auslastung der Automationen bis hin zu deren Produktivität und Profitabilität über einen größeren Zeitraum bestimmt werden [48].

Die Probenverwaltung, beginnend mit dem Probeneingang bis hin zu Analyseergebnissen, wurde in der Vergangenheit häufig mittels Tabellendokumenten bewältigt, die auf den Computern der jeweiligen Bearbeiter gespeichert wurden [49]. Die Steigerung des Probenaufkommens, die komplexen Abläufe im Labor, sowie die weiterhin steigende Menge an Daten, macht das Arbeiten mit solchen Dokumenten und deren Weitergabe unmöglich [50, 51]. Arbeiten mehrere Benutzer an dem gleichen Tabellendokument kann es zu redundanten Einträgen kommen. Es entstehen verschiedene Versionen des Dokuments. Eine Standardisierung und Kontrolle der Daten ist schwer möglich, vor allem wenn häufig Einträge aktualisiert oder gelöscht werden [49].

Die Verfolgung und Kontrolle prozessrelevanter Daten hilft Arbeitsabläufe zu optimieren, Probleme zu identifizieren und zu beheben. Das Risiko von Fehlern während der komplexen Laborabläufe und die damit einhergehenden Kosten werden durch die Standardisierung reduziert [45].

Neben der Probenregistrierung und dem Controlling von Laborprozessen bietet eine zentrale Datenbank allen Kooperationspartnern gleichermaßen Zugang zu den Projektdaten und Analyseergebnissen [49]. Der Zugang zur Datenbank kann, im Kontrast zu Tabellendokumenten, ausschließlich autorisierten Nutzern und Nutzerkreisen über eine Nutzerkennung und ein individuelles Passwort gewährleistet werden. Tabellendokumente können zwar passwortgeschützt sein, jeder Benutzer kennt und nutzt jedoch das gleiche Passwort. Eine wirkliche Datensicherheit ist nicht gewährleistet [49]. LIM-Systeme sind kommerziell erwerbbar jedoch ist die Lizenzgebühr einer solcher Applikationen sehr hoch bei gleichzeitig geringer Komplexität und Flexibilität der Anwendungen [49, 52]. Mit der Entwicklung eines eigenen *NGS-LIMS* kann die Anwendung von Grund auf nach den individuellen Bedürfnissen gestaltet und erweitert werden. Dem Nutzer steht eine intuitive graphische Oberfläche zur Verfügung in der *NGS*-relevante Daten systematisch und standardisiert erfasst und abgerufen werden können.

1.5 Bioinformatik

Jedes Hochdurchsatz-Sequenzierexperiment ergibt eine große Menge an Daten, die erst durch eine Abfolge informatischer Schritte für den Nutzer analysierbar werden. Die Primärdaten der Sequenzierung sind die durch die Kamera aufgenommenen Bilder der Fluoreszenz der einzelnen Nukleotid-Inkorporationen. Die-

se werden in der *Image Analysis* und im *Base Calling* in eine Nukleotidsequenz umgeschrieben (Abb.: 3K). Die Einzelsequenzen werden als *Read* bezeichnet. Sie werden im Anschluss an das *Base Calling* über ihren Index mit der Software CASAVA den einzelnen Patienten zugeordnet (*Demultiplexing*), um danach im *Read Alignment* definierten Positionen auf dem Referenzgenom zugeordnet zu werden. Hierzu nutzt man zum Beispiel die Software BWA [53]. Als Referenzsequenz dienen die Genom-Definitionen der Datenbanken University of Santa Cruz (UCSC) oder Genome Reference Consortium (GRC) in den jeweils aktuellen Versionen. Die Summe aller *Reads* die eine genomischen Position abdecken, wird als Sequenziertiefe (*Coverage*) bezeichnet. Je höher die Abdeckung einer einzelnen genomischen Position ist, desto sicherer kann man homozygote von heterozygoten Varianten unterscheiden und Sequenzierfehler von echten Varianten trennen [24]. Bei einer Exom-Sequenzierungen zielt man auf eine Abdeckung von mindestens 20-fach bei mehr als 90 % (durchschnittlich 95 %) der Basen ab. Bei Studien in denen durch den Vergleich der genetischen Information von Eltern und Kindern *de novo*-Mutationen identifiziert werden sollen spielt die Sensitivität eine große Rolle. Es wird empfohlen, die Abdeckung für alle drei Proben auf mindestens 60-fach zu erhöhen, um vor allem die beiden Allele der Eltern ausreichend vorliegen zu haben [54]. Je höher die *Coverage* einer genomischen Position ist, desto sicherer gelingt auch das *Variant Calling* durch Software wie SAMtools [55] oder GATK [56]. Hierbei wird definiert an welchen Nukleotid-Positionen sich die sequenzierte DNA vom Referenzgenom unterscheidet. Zudem ist das Gelingen des *Variant Callings* abhängig von der Qualität der Sequenzierung. Für eine qualitative Evaluation der erhaltenen Sequenzierdaten werden verschiedene Parameter herangezogen die in Tabelle 2 näher definiert sind. Zeigt eine *Library* verringerte Qualitätsparameter nach der Sequenzierung kann es zu einem vermehrten Auftreten von falsch-positiven Varianten kommen. An das *Variant Calling* schließt die Annotation an. Um mögliche funktionelle Konsequenz einer Variante vorherzusagen, fließen Parameter aus *in silico* Analysen ein, wie die evolutionäre Konservierung des Nukleotids (GERP [59], PhyloP [60]) oder Werte zur biochemischen Distanz zwischen wildtyp- und varianter Aminosäure (Grantham [61]). Zusätzlich dazu wird mit Anwendungen wie SIFT (sorting intolerant from tolerant) [62] und PolyPhen2 (polymorphism phenotyping) [63] eine Vorhersage auf Grund von strukturellen Effekten der Veränderung getroffen.

Tabelle 2: Definition der Parameter zur Beurteilung der Sequenzierqualität.
Die Angaben zum Durchschnitt beziehen sich auf die Berechnungen aus den Daten von 100 internen Exomen von DNA aus peripheralem Blut.

Parameter	Definition	Durchschnitt
<i>Mapped Reads</i>	Prozentuale Anzahl aller <i>Reads</i> , die dem Referenzgenom eindeutig zugeordnet werden konnten.	>99 %
Gb	Anzahl der pro <i>Library</i> sequenzierten Gigabasen.	>8%
∅ Abdeckung	Durchschnittliche Anzahl der <i>Reads</i> , die jede Ziel-Base des Exoms abdecken	>120-fach
20-fache Abdeckung	Prozentuale Anzahl der Basen, die von wenigstens 20 <i>Reads</i> abgedeckt werden.	>98 %
<i>On-Bait</i>	Prozentsatz aller <i>Reads</i> , die auf der definierten Zielsequenz liegen +/- 50 bp.	>77%
Duplikatrate	Prozentualer Anteil der <i>Reads</i> eines Exoms, die exakt die gleiche chromosomale Position abdecken.	9%
<i>CNV Noise</i>	Statistischer Wert der angibt, wie vergleichbar die Abdeckung eines Exoms im Kontrast zu einem Datenset aus 100 zufälligen Exomen ist, um mögliche CNVs zu detektieren [57].	<2,0%
Kontamination	Prozentualer Wert der statistisch errechnet Kontamination. [58].	<2,0%

Tabelle 3: Übersicht zur Art und Anzahl von genetischen Varianten in einem humanen Exom. Die Durchschnittswerte beziehen sich auf die Analyse von 14.486 internen Exomen. Zusätzlich zu den tabellarisch dargestellten Varianten finden sich weitere Veränderungen in inter- oder intragenischen Bereichen, in miRNAs sowie *near-splice*.

Art	durchschnittliche Anzahl pro Exom
Gesamt	71.185
nicht-synonym (<i>missense</i>)	11.400
synonym	11.600
<i>frameshift</i>	302
Indel	335
<i>nonsense</i> (Stopp)	114
Spleißstelle	7.500
<i>stoploss</i>	38
UTR	8.052

Es resultiert eine Liste aus durchschnittlich 71.000 exomischen Varianten (basierend auf der Analyse von rund 14.500 internen Exomen) mit dem dazugehörigen Qualitätswert, der Art der Veränderung sowie der genomischen Position. Ein Exom beherbergt mehr als 23.000 kodierende Varianten (*frameshift*, *missense*, *nonsense*, *stoploss*, synonym) im Vergleich zur humanen Referenz. Durchschnittlich 11.400 davon sind nicht-synonyme Aminosäureveränderungen sowie 11.600 synonyme Varianten (Tab.: 3).

Mögliche ursächliche Varianten werden auf Grund der Häufigkeit ihres Vorkommens und der vorhandenen Annotation gefiltert. Zur Interpretation der verbleibenden Varianten stehen der Bioinformatik verschiedene Datenquellen zur Verfügung. Informationen aus den Datenbanken HGMD (*The Human Gene Mutation Database*) [64] oder OMIM (*Online Mendelian Inheritance in Man*) [6] geben Aufschluss über eine bereits bekannte Assoziation der Variante oder eines Gens mit einem Phänotyp. Zusätzlich fließen Frequenzinformationen aus dbSNP [65], dem Exome Variant Server (EVS) [66] sowie der ExAC Datenbank [67] ein. Für die weitere Filterung der Varianten werden verschiedene Priorisierungsstrategien angewendet.

1.6 Strategien zur Identifizierung krankheitsverursachender genetischer Varianten aus Exom-Daten

Die wohl größte Herausforderung der Hochdurchsatzsequenzierung ist die Identifikation einer pathogenen Genvariante in einer großen Anzahl von Varianten. Ein Exom beherbergt drei verschiedene Klassen von Mutationen [68]. Punktmutationen treten als Transitionen oder Transversionen auf, wobei eine Base gegen eine andere getauscht wird (Basensubstitution). Bei Insertionen und Deletionen kommt es zu einem Hinzufügen oder Entfernen von einem oder mehreren Nukleotiden, bis hin zu Änderungen der Kopiezahl von Genen (*copy number*-Varianten). Punktmutationen können in kodierenden als auch in nicht-kodierenden Bereichen des Genoms auftreten. Kodierende Varianten werden unterteilt in synonyme, *missense*, *frameshift*, *stop* oder *stop-loss*-Varianten sowie in funktionellen Domänen gelegen, wie zum Beispiel der kanonischen Spleißstelle oder der 5'-UTR (*untranslated region*). Genetische Varianten treten heterozygot (ein Allel betreffend) oder homozygot (beide Allele betreffend) auf. So werden die rund 71.000 genetischen Veränderungen in einem Exom klassifiziert und eingeteilt (Tab.: 3). Zur Priorisierung des Datensatzes, von seltenen, nicht-synonymen Varianten einer *NG*-Sequenzierung, nutzt man abhängig vom Studiendesign, verschiedene Priorisierungsstrategien [2] zur Identifikation der, für einen pathogenen Phänotyp kausativen, genetischen Veränderung (Abb.: 5).

1.6.1 Priorisierungsstrategie bei familiären Erkrankungen

Segregiert eine Erkrankung in einer Familie werden die Exome oder Genome von zwei oder mehreren, von einer Krankheit betroffene, Familienmitglieder untersucht (Abb.: 5a) [2]. Die Varianten werden auf gemeinsam auftretende Veränderungen hin priorisiert. Nah verwandte Menschen teilen sich viele gemeinsame Alterationen, während weit entfernt Verwandte nur noch wenige genetische Gemeinsamkeiten teilen. Je größer der Stammbaum einer Familie ist, aus dem zwei entfernt Verwandte analysiert werden, desto höher ist der Erfolg bei dieser Priorisierungsstrategie. Cousin und/oder Cousinen 1. Grades teilen sich 12,5 % genetische Sequenz, während Verwandte 3. Grades unter 1 % genetisches Material gemeinsam haben. Zusätzlich dienen die Exom-Daten von nicht betroffenen Mitgliedern einer Familie als Kontrolle [25]. Varianten, die in den Betroffenen, nicht

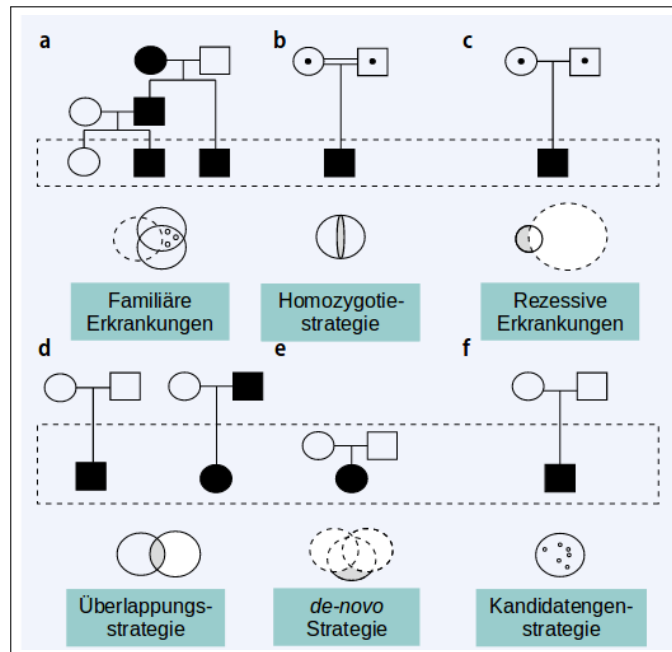


Abbildung 5: Priorisierungsmöglichkeiten von Exom-Daten (angepasst nach Neveling *et al.* [2]).

aber in den Gesunden auftreten, werden weiterführend untersucht. Die identifizierten Kandidatengene werden in der gesamten Familie und in einer krankheitsunabhängigen Kontrollpopulation analysiert. Kosegregiert eine Variante mit der Erkrankung innerhalb der Familie, gibt es nicht-verwandte Betroffene, welche die gleiche Variante oder eine Veränderung im gleichen Gen tragen und findet man diese Varianten nicht in einer Kontrollpopulation und gesunden Familienmitgliedern, so ist sie wahrscheinlich pathogen und ursächlich für den untersuchten Phänotyp [69, 70].

1.6.2 *de novo*-Strategie bei Eltern-Kind-Trios

Die Priorisierung anhand der *de novo*-Strategie [2] hat sich vor allem bei schweren Entwicklungsstörungen mit früher Letalität etabliert [71]. Essentiell für die Suche nach *de novo*-Alterationen, ist die Sequenzierung eines Eltern-Kind-Trios (Abb.: 5e), wobei das Kind von der Erkrankung betroffen ist, während die Eltern gesund sind. Pro Patient identifiziert man durchschnittlich 1 bis 2 nicht-synonyme *de novo*-Varianten im Exom. Die Trio-basierte *de novo*-Strategie hat sehr hohe Ansprüche an die Qualität und die Reproduzierbarkeit der Anreiche-

rung und Sequenzierung. Um eine Variante in den Eltern ausschließen zu können, sollte die durchschnittliche Sequenziertiefe aller drei Proben 60-fach betragen, die Qualität der Varianten hoch sein und gefundene Veränderungen mittels Sanger-Sequenzierung validiert werden [72]. Etliche Studien zeigen die erfolgreiche Anwendung dieser Strategie in heterogenen Erkrankungen wie Autismus und Schizophrenie [71, 73, 74, 72] sowie bei Patienten mit schwerer nicht-syndromaler mentaler Retardierung [75].

Eine ganz ähnliche Analyse findet auch Anwendung bei der Suche nach somatischen Mutationen in Tumoren. Anstelle eines Trios wird hier DNA aus gesundem Gewebe sowie aus Tumorgewebe sequenziert. Initial vergleicht man die Daten des Tumors mit denen des gesunden Gewebes und filtert die somatischen Varianten heraus. Diese wiederum können in einem nächsten Schritt mit anderen Tumor-Gewebe-Paaren abgeglichen werden, wobei alle gesunden Gewebe als Kontrolle dienen und gemeinsame Genvarianten, oder das gleiche betroffene Gen, in den Tumorproben gefiltert werden [76, 77].

1.6.3 Überlappungsstrategie in Patientengruppen mit gleichem Phänotyp

In Patientengruppen, die eine Erkrankung ohne genetische Heterogenität gemeinsam haben, ist die Anwendung der Überlappungsstrategie [2] erfolgversprechend. Das ist besonders in dem Fall hilfreich, wenn ein autosomal-dominanter Erbgang vermutet wird und keine großen Familien mit mehreren Betroffenen zur Verfügung stehen (Abb.: 5d). Die Untersuchungshypothese setzt dabei voraus, dass nicht-verwandte Personen mit der gleichen Erkrankung, auch die gleiche genetische Variante, oder Veränderungen im gleichen Gen gemeinsam haben. Tatsächlich ist die Anzahl der Krankheitsgene in denen mehrere nicht-verwandte Betroffene Mutationen tragen sehr gering [78]. Eine gute klinische Diagnose der Patienten und die tatsächlich vorhandene genetische Homogenität reichen aus, um in einer sehr kleinen Patientengruppe ein Kandidatengen zu identifizieren [79, 80, 81].

1.6.4 Homozygotie-Strategie bei konsanguinen Eltern

Während bei unverwandten oder entfernt verwandten Paaren ein Basisrisiko von 2-4 % besteht, ein durch einen Gendefekt krankes Kind zu bekommen, steigt dieses Risiko bei Verwandten 1. Grades auf bis zu 25 %. Bei Cousin/Cousine 1.

Grades liegt das Risiko noch bei 6,25 % [82]. Untersucht man rezessive Erkrankungen bei Kindern mit gesunden, blutsverwandten Eltern, kann auf rezessive Veränderungen in homozygoten Genomabschnitten hin priorisiert werden [2]. Die homozygoten DNA-Abschnitte können direkt durch die Exom-Analyse definiert werden. Nun sucht man nach rezessiven Varianten in den Betroffenen und schließt ihr Auftreten bei gesunden Verwandten und in einer Kontrollpopulation aus [2]. Mit dieser Strategie konnte die genetische Ursache von Epilepsie und Myoklonie in einer konsanguinen Ägyptischen Familie aufklärt werden [83].

1.6.5 Priorisierungsstrategie bei rezessiven Erkrankungen

In Fällen, in denen nur sehr wenige Patienten oder kleine Familien zur Verfügung stehen, kann keine der bisher beschriebene Priorisierungen durchgeführt werden [2]. Bei rezessiven Erkrankungen kann nach der ersten allgemeinen Filterung der Varianten nach ihrer Funktionalität und der Häufigkeit ihres Auftretens, auf homozygote oder *compound*-heterozygote Varianten hin priorisiert werden. Dies verringert die Anzahl der zu untersuchenden Alterationen um ein vielfaches. Verbliebene Varianten können anschließend in anderen Patienten analysiert oder ihre Kosegregation untersucht werden. So konnten Pierce *et al.* in einer kleinen Familie mit zwei betroffenen Schwestern mit Perrault-Syndrom eine *compound*-heterozygote Variante in *HSD17B4* identifizieren, die zu einer deutlichen Reduktion der HSD17B4-Proteinexpression führt [84].

1.6.6 Kandidatengen-Strategie bei individuellen Patienten

Es gibt Fälle, bei denen man nur zu einem individuellen Patienten Zugang hat. Dies hat zur Folge, dass die große Anzahl an genetischen Veränderungen nicht mit Familienmitgliedern oder anderen Patienten abgeglichen werden können [2]. Letztlich erfolgt die Priorisierung anhand möglicher Effekte auf die Proteinfunktion. Einen hohen Stellenwert erhalten alle Stopp- und *frameshift*-Mutationen und solche, die die kanonische Spleißstelle betreffen. *Missense*-Mutationen werden anhand ihres Effektes auf das Protein priorisiert (Grantham-Wert), sowie die evolutionäre Konservierung der betroffenen Aminosäure und der Nukleotide (phyloP oder GERP-Wert) [2]. Dem zugrunde liegt der Nachweis, dass pathogene Varianten vor allem hoch konservierte Bereiche betreffen, im Gegensatz zu neutralen Veränderungen [71]. Es gibt einige bioinformatische *in silico* Anwen-

dungen, die genetische Veränderungen als schädlich oder unschädlich, mit einer Richtigkeit von 75 bis 80 %, einstufen [54]. Klassische Prädiktionsprogramme sind PolyPhen2 [63], SIFT [85], Mutation Taster [86], MutPred und SNAP [87]. Die Anwendung der Kandidatengenstrategie, zusammen mit computergestützten Prädiktionsprogrammen zur Identifizierung eines Kandidatengens bei erblicher spastischer Paraparese, zeigten Erlich *et al.*. Sie definierten anhand dieser Methodik eine Mutation in *KIF1A* als wahrscheinlich pathogen [88].

Die beschriebenen Priorisierungsstrategien allein, oder in Kombination miteinander, haben starkes Potential. Zur Validierung einer möglichen Assoziation zwischen Mutation und Phänotyp sowie dem Verständnis zur biologischen Grundlage einer Erkrankung sind funktionelle Studien notwendig. Erste Schritten könnten die Analyse der Proteinsynthese,- faltung, und -degradation sein. Zusätzlich kann der Einfluss auf den Proteintransport und Protein-Protein-Interaktionen zwischen dem Wildtyp und der Mutante untersucht werden [54]. Mittlerweile unterstützt auch hier die Bioinformatik mit geeigneter Software zur funktionellen Beurteilung, wie zum Beispiel bei Simulationen von Moleküldynamiken [89]. So kann computergestützt gezeigt werden, wie sich die Faltung oder Stabilität eines Proteins durch den Austausch von Aminosäuren verändert, wodurch Rückschlüsse auf funktionelle Veränderungen möglich sein können [69].

1.7 Zielsetzung der Arbeit

Der erste Teil der Aufgabenstellung meiner Arbeit umfasste die Entwicklung eines Labor-Informations- und Management-Systems (LIMS) für den Bereich der *Next-Generation*-Sequenzierung (Beginn Ende 2009). Es sollten die laborrelevanten Bereiche der Probenerfassung, Probenprozessierung und Qualitätssicherung enthalten. Ein solches System war bisher für die Aufgaben innerhalb der Hochdurchsatzsequenzierung am Institut für Humangenetik noch nicht verfügbar.

Im einem zweiten Teil dieser Arbeit sollte die Exom-Sequenzierung im Hochdurchsatz eingeführt werden. Dies inkludierte die Etablierung einer automatisierten *Library*-Präparation, der Standardisierung der Laborabläufe sowie der Sicherstellung der Qualität von Laborarbeit und der erhaltenen Sequenzierdaten.

Die Einführung der automatisierten *Library*-Präparation gliederte sich in zwei Schritte: mit dem Erwerb des Pipettierroboters *Bravo Liquid Handling Station* der Firma Agilent im Jahr 2010 konnte der manuelle Durchsatz auf 48 Proben, bei geringerer Bearbeitungszeit durch den Anwender, gesteigert werden.

In einem zweiten Schritt, zwei Jahre später, wurde der *Bravo* erweitert und die parallel Prozessierung von 96 Proben ist seither möglich. Im Laufe der Evolution der Laborfertigkeiten wurden Laborprozesse optimiert und standardisiert.

Der dritte Teil der vorliegenden Arbeit hatte die Anwendung der Exom-Sequenzierung zur Identifikation pathogener Genvarianten bei Patienten mit häufigen und seltenen genetischen Erkrankungen zur Aufgabe. Es wurde exemplarisch in einer Studie aus dem Jahr 2011 bei Patienten mit Morbus Parkinson gezeigt, dass die Exom-Sequenzierung erfolgreich zur Identifikation kausativer Varianten bei monogenen Erkrankungen angewendet werden kann.

In einer zweiten Studie aus dem Jahr 2014 bei Patienten mit Burn-McKeown-Syndrom zeigte sich die Limitation der Exom-Sequenzierung. Es wird demonstriert, dass durch den Fokus auf proteinkodierende Bereiche kausative Varianten, wie in diesem Fall innerhalb des Promotorbereichs, durch die Exom-Sequenzierung allein nicht detektierbar sind. Durch eine Kombination aus Exom- und Genom-Sequenzierung konnten die ursächlichen, *compound*-heterozygoten-Varianten in Betroffenen mit Burn-McKeown-Syndrom nachgewiesen werden.

2 Material und Methoden

2.1 Material

2.1.1 Probenmaterial

Der Hauptteil des Probenmaterials wurde uns in Form von DNA von den jeweiligen Kooperationspartnern zugesandt. Dort wurde es im Rahmen der genetischen Diagnostik und Beratung gesammelt. Die Proben des Parkinson Projektes erhielten wir von der Medizinischen Universität Wien. Alle Patienten und deren Verwandten, die an dieser Studie teilnahmen, haben ein schriftliches Einverständnis zur Nutzung ihrer Daten gegeben. Als Kontrollpopulation nutzen wir DNA der Probanden der KORA-AGE und KORA-S4 Studie. Die Patientenproben für die Studie bei Patienten mit Burn-McKeown-Syndrom wurden uns durch unsere Kooperationspartner zur Verfügung gestellt. Als Kontrollpopulation wurden gesunde Probanden aus Manchester, Essen und der KORA-Studie ausgewählt.

2.1.2 Geräte

Geräte	Hersteller
Bravo Liquid Handling Platform Option B	Agilent Technologies
NanoDrop 1000 Spectrophotometer	Thermo Fisher Scientific Inc.
Qbit Fluorometer	Thermo Fisher Scientific Inc.
peqStar PCR Thermocycler	VWR International GmbH
Safire II micorplate reader	Tecan Group Ltd.
Covaris E220 series Sample Preparation System	Covaris
Covaris S2	Covaris
Agilent 2100 Bioanalyzer	Agilent Technologies
LabChip GX	PerkinElmer Inc.
cBot v 1.5	Illumina, Inc.
Savant Speed Vac	Thermo Fisher Scientific Inc.
Eppendorf Concentrator Plus	Eppendorf AG
Genome Analyzer IIx	Illumina, Inc.
HiSeq2000, HiSeq 2500, HiSeq4000	Illumina, Inc.
ABI Prism 3730 DNA Sequenzer	Thermo Fisher Scientific Inc.
7900HT real-time PCR System	Thermo Fisher Scientific Inc.
LightCycler 480	Roche Ltd
LightScanner HR I 384 Instrument	Idaho Technology Inc.
chemagig MSM I	PerkinElmer Inc.

2.1.3 Verbrauchsmaterialien

Reagenzien	Hersteller
Quant-iT Broad-Range DNA Assay Kit	Thermo Fisher Scientific Inc.
Qubit dsDNA BR Assay Kit	Thermo Fisher Scientific Inc.
Covaris micro TUBE with AFA fiber and snap cap	Covaris
Agilent DNA 1000 Kit	Agilent Technologies
Agilent High Sensivity DNA Kit	Agilent Technologies
SureSelect Human All Exon Kit v3, v5, v6	Agilent Technologies
TruSeq DNA PCR-Free Sample Preparation Kit	Illumina, Inc.
DNA 1K Reagent Kit	PerkinElmer Inc.
DNA High Sensitivity Reagent Kit	PerkinElmer Inc.
KAPA Library Quantification Kit	Kapa Biosystems
LCGreen® Plus und Melting Dye	BIOKÉ
LCGreen-Lightscanner-PCR	Thermo Fisher Scientific Inc.
TruSeq Paired end Cluster Kit	Illumina, Inc.
TruSeq SBS-Kit HS	Illumina, Inc.
PCR Master Mix	Promega Corporation
BigDye Terminator Cycle Sequencing Ready Reaction v.3.1 Kit	Thermo Fisher Scientific Inc.
NucleoFast 96 PCR clean up Kit	Macherey-Nagel
1 Kb DNA Extension Ladder	Thermo Fisher Scientific Inc.

2.1.4 Software

Software	Hersteller
CASAVA v.1.8	Illumina, Inc., San Diego, USA
Cyrillic	Cyrillic Software, Wallingford, England
Integrative Genomics Viewer	Broad Institute, Cambridge, USA
Staden Package	Sourceforge
Burrows-Wheeler-Aligner	http://bio-bwa.sourceforge.net
SAMtools	http://samtools.sourceforge.net

2.1.5 Internetquellen - URLs

Anwendung	URL
Exome Variant Server	http://evs.gs.washington.edu/EVS/
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/
Primer3	http://frodo.wi.mit.edu/primer3/input.htm
ExonPrimer	http://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html
Online Mendelian Inheritance in Man	http://www.ncbi.nlm.nih.gov/omim
HGMD Professional	http://www.biobase-international.com/product/hgmd
UCSC Genome Browser	http://genome.ucsc.edu
ClustalW2	http://www.ebi.ac.uk/Tools/msa/clustalw2
PyMOL	http://www.pymol.org
1000 Genomes	http://browser.1000genomes.org
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq

2.2 Molekularbiologische Methoden

2.2.1 DNA-Isolation und Konzentrationsbestimmung

Genomische DNA aus Gesamtblut wurde automatisiert mit den chemagig MSM I nach Angaben des Herstellers, mit dem chemagic DNA Blood3k oder dem chemagic DNA Blood7k Kit extrahiert und bei -20°C gelagert. Die DNA-Konzentration wurde mit dem NanoDrop 1000 bestimmt. Zusätzlich fand eine fluoreszenzbasierte Quantifizierung mit dem *Quant-iT Broad-Range DNA Assay Kit* in Verbindung mit dem LightScanner zum Auslesen der Platten statt.

2.2.2 Oligonukleotide

Die Oligonukleotide wurden von der Firma Metabion (Martensried, Deutschland) synthetisiert und für die PCR und die Sanger-Sequenzierung verwendet. Die Sequenzen der verwendeten Oligonukleotide können dem Anhang entnommen werden.

2.2.3 Amplifikation von DNA-Fragmenten mittels PCR

Mit Hilfe der Polymerasekettenreaktion ist es möglich gezielt spezifische DNA-Bereiche zu amplifizieren [90]. Alle genutzten Primer wurden mit den Programmen Primer3 und ExonPrimer ermittelt. Dazu wurde der Promega PCR Master Mix, nach Angaben des Herstellers in 25 μl Reaktionsvolumen und 0,4 μM finaler Primerkonzentration genutzt. Der Promega-MasterMix enthält 50 U/ml Taq

DNA Polymerase, Puffer, je 400 μ M dATP, dTTP, dGTP, dCTP und 3 mM $MgCl_2$. Pro PCR Ansatz setzten wir 50 ng DNA ein. Die PCR erfolgte im peq-Star Thermocycler (Peqlab) nach folgendem *touch-down*-Programm:

Initiale Denaturierung 95°C 5 min

8x

Denaturierung 95°C 30 sec

Annealing 65°C 30 sec -1°C pro Zyklus

Elongation 72°C 30 sec

25x

Denaturierung 95°C 30 sec

Annealing 58°C 30 sec

Elongation 72°C 30 sec

finale Elongation 72°C 5 min

Im Anschluss an die PCR wurden die amplifizierten DNA-Abschnitte im Agarosegel elektrophoretisch aufgetrennt und analysiert. Die in dieser Studie genutzten intronischen Primer für *VPS35* sind im Anhang aufgelistet.

2.2.4 Agarose-Gelelektrophorese

Die horizontale Agarose-Gelelektrophorese dient zur Auftrennung von amplifizierten DNA-Fragmenten. Zur Herstellung der Gele wurde die Agarose in 50 ml 1xTBE im Mikrowellenherd aufgeköcht und mit 3 μ l Ethidiumbromid versetzt. Die Suspension wurde anschließend in Gelschalen gegossen und abgekühlt. 3 μ l Probe wurden mit 7 μ l Ladepuffer (1:4 verdünntes Orange G) vermischt und in die Geltaschen pipettiert. Die Auftrennung erfolgte in 1x TBE Puffer bei einer Spannung von 180-200V. Zur Abschätzung der Fragmentgrößen wurden als Standard die 1 kb DNA Ladder der Firma Invitrogen verwendet.

2.2.5 Aufreinigung der PCR-Produkte für eine anschließende Sanger-Sequenzierung

Die in den jeweiligen PCR-Ansätzen erhaltenen Produkte wurden für die folgende Sanger-Sequenzierung aufgereinigt, um Puffer, Nukleotide und ungebundene Pri-

mer zu entfernen. Für die Reinigung wurden der *NucleoFast 96 PCR clean up Kit* (Macherey-Nagel) verwendet. Das Kit funktioniert nach einem Ultrafiltrations-Prinzip. Das PCR-Produkt wird mit Wasser auf ein Volumen von mindestens 100 μ l eingestellt, auf eine Ultrafiltrationsmembran geladen und unter Vakuum (600 mbar) für 15 bis 45 min gesetzt. Die Porengröße der Filtermembran ist so gewählt, dass Fragmente, die größer als 150 bp sind, auf der Membran haften bleiben, während kleinere Fragmente wie ungebundene Primer, Salze und Nukleotide durch die Membran gesaugt werden. Um das gereinigte PCR-Produkt zu erhalten, werden mindestens 25 μ l Wasser zugefügt und durch Mischen auf einem Schüttler (200 rpm, 15 min) von der Membran gelöst. Anschließend können die PCR-Produkte in eine neue PCR Platte überführt und für weitere Schritte verwendet werden.

2.2.6 Sanger-Sequenzierung

Die Sequenzreaktion nach Sanger [14] basiert auf der Kettenabbruch-Methode. Dabei werden die DNA-Fragmente mit Fluoreszenzfarbstoffen an Didesoxynukleotiden markiert, die von dem Kapillarsequenziergerät ABI Prism 3730 DNA Sequenzer (Applied Biosystems) während der Elektrophorese mittels Laser angeregt und gemessen werden. Die vier Didesoxynukleotide (ddATP, ddTTP, ddGTP, ddCTP) sind mit vier verschiedenen Farbstoffen markiert und ermöglichen so die Unterscheidung der Nukleotide. Die Sequenzreaktion wird mit dem *BigDye Terminator Cycle Sequencing Ready Reaction v.3.1 Kit* der Firma Applied Biosystems angesetzt.

Material und Methoden

Ansatz Sequenzreaktion:

Reagenz	Volumen in μl
Big Dye Terminator Puffer	1.0
Primer F bzw. R (0.2 μM)	1.0
Big Dye Terminator	0.5
PCR-Produkt	(50-100 ng/ μl) 1,0-2,5
H ₂ O	auf 5 μl auffüllen

PCR-Bedingungen der Sequenzreaktion:

25x

96 °C 10 sec

50 °C 5 sec

60 °C 4 min

Im Anschluss an die Sequenzreaktion erfolgt die Fällung der DNA-Fragmente zur Entfernung nicht gebundene Didesoxynukleotide aus dem Sequenzansatz durch Ethanol. Zum Reaktionsansatz in der Mikrotiterplatte werden 25 μl Ethanol (100 %) zugegeben, 15 min bei Raumtemperatur inkubiert und danach für 30 min bei 3000 x g zentrifugiert. Anschließend wird das Pellet mit 50 μl Ethanol (70 %) gewaschen und zentrifugiert (15 min, 2000 x g). Um das Pellet zu trocknen, wird die Mikrotiterplatte mit der Öffnung nach unten 1 min bei 600 x g zentrifugiert, 5 min bei Raumtemperatur getrocknet und schließlich in 25 bis 50 μl Wasser gelöst. Die anschließende Sanger-Sequenzierung wird auf dem ABI Prism 3730 DNA Sequenzer (Applied Biosystems) nach den vom Hersteller angegebenen Bedingungen durchgeführt. Es werden je 25 μl der aufgereinigten Sequenzreaktion in eine Sequenzierplatte übertragen. Die erhaltenen Daten werden durch das Staden Package als Elektropherogramme dargestellt und können manuell ausgewertet werden.

2.3 Mutationsscreening

2.3.1 IDAHO-Screening

Wir untersuchten genetische Alterationen in Kandidatengenomen in einer Population via *High Resolution Melting (HRM) Curve* Analyse. Dabei werden Sequenzabweichungen zu einer Wildtyp-DNA durch Veränderungen im Schmelzverhalten sichtbar. Initial muss die zu untersuchende Region mittels PCR amplifiziert werden. Genomische DNA (je 5 ng pro I) wurde dazu auf einer 384-Mikrotiterplatte als Duplikat verteilt und eingetrocknet. Zu der getrockneten DNA wurden 5 µl Mastermix gegeben. Der Mastermix enthielt 3 µl H₂O, 0.5 µl 10x Puffer, 0.5 µl dNTPs (2 mM), 0.5 µl MgCl₂ (25 mM), 0.05 µl Taq-Polymerase und je 0.1 µl des Vorwärts und Rückwärtsprimers (10 µM). In den PCR-Mastermix wurden 0.25 µl der Farbstoff LCGreen Dye Plus (Idaho Technology Inc.) beigemischt der doppelsträngige DNA bindet und bei einer Wellenlänge von 470-520 nm detektiert werden kann. Amplifiziert wurde mit einem speziellen Touchdown-PCR Programm:

Initiale Denaturierung 95°C 10 min

25x

Denaturierung 95°C 30 sec

Annealing 70°C 30 sec -1°C pro Zyklus

Elongation 72°C 30 sec

25x

Denaturierung 95°C 30 sec

Annealing 54°C 30 sec

Elongation 72°C 30 sec

finale Elongation 72°C 10 min

Die PCR-Produkte wurden anschließend mit 8 µl Öl überschichtet und im LightScanner unter folgenden Einstellungen aufgeschmolzen:

Hold Temperatur: 74°C

Start Temperatur: 77°C

End Temperatur: 96°C

Die Fluoreszenz nimmt im Verlauf der Aufschmelzung immer weiter ab. PCR-Produkte mit unterschiedlicher Nukleotidabfolge weisen dabei einen unterschiedlichen Kurvenverlauf auf. Es können nicht nur SNPs und SNVs sondern auch Insertionen und Deletionen detektiert werden [91]. Die Auswertung der Schmelzkurven der Amplifikationsprodukte erfolgte mit der LightScanner Software v.1.1.0.566 (Idaho Technology Inc.). Die Referenzsequenz wurde als größte Gruppe gleichverlaufender Schmelzkurven definiert, in der sich auch Wildtyp-Proben als Negativkontrolle befanden. Anschließend wurde ein Kurvenverlauf festgelegt. Kam es zu einer auffälligen Abweichung des Kurvenverlaufs des Duplikats einer Probe von der Referenz-Schmelzkurve wurden die betreffenden PCR-Produkte sequenziert um die genetische Alteration zu definieren. Dazu wurden 8 µl PCR-Ansatz entnommen, mit 100 µl H₂O aufgefüllt und dann nach den beschriebenen Standardprotokollen aufgereinigt und Sanger-sequenziert.

2.4 *Next-Generation-Sequenzierung*

Der Begriff *Next-Generation-Sequenzierung* (*NGS*) umfasst zum Einen die initiale Präparation der zu analysierenden DNA und zum Anderen die eigentliche Hochdurchsatzsequenzierung. Anschließend erfolgt die bioinformatische Auswertung der einzelnen Proben. Unter dem Begriff *Library-Präparation* versteht man die initiale Fragmentierung der DNA gefolgt von einer Modifizierung der Fragmentenden entsprechen der genutzten Sequenzierplattform.

2.4.1 *Genom-Sequenzierung*

Der Begriff Genom-Sequenzierung beschreibt die Sequenzierung des gesamten Genoms mit einer *NGS*-Plattform. Vor der eigentlichen Sequenzierung wurde die *Library-Präparation* mit dem *TruSeq DNA PCR-Free Sample Preparation Kit* der Firma Illumina, nach Angaben des Herstellers durchgeführt. Dabei wurden 1 µg hochmolekulare DNA mittels Untraschall (Covaris INC.) fragmentiert, die Enden der doppelsträngigen DNA-Moleküle gleichmäßig aufgefüllt und plattformspezifische Adapter ligiert. Die finale Quantifizierung der Genom-*Libraries* erfolgte mit dem *KAPA Library Quantification Kit* (Kapa Biosystems) nach Angaben des Herstellers im *ABI HT 7900 Real Time* PCR Gerät.

2.4.2 Exom-Sequenzierung

Der Begriff Exom-Sequenzierung beschreibt die Anreicherung und Sequenzierung aller exonischen Regionen des Genoms. Dabei werden die kodierenden Regionen oftmals durch weitere Bereiche wie zum Beispiel 5'-UTRs erweitert [26, 27, 28, 29, 30, 31, 32, 33]. Die derzeit verfügbaren Methodiken, zur Anreicherung gewünschter Sequenzen, beruhen auf Hybridisierungsmethoden [28, 92, 34]. Dabei werden zuvor amplifizierte Genom-*Libraries* an komplementäre, biotynilierte DNA (Illumina und NimbleGen) oder RNA (Agilent) *Baits* gebunden [35, 36, 28]. Im anschließenden Schritt, werden die DNA/RNA- bzw. DNA/DNA-Hybride an Streptavidin-gekoppelte *beads* gebunden und alle unerwünschten Fragmente weg gewaschen. Final wird die zu untersuchende DNA aus dem Hybrid gelöst, amplifiziert und kann im Anschluss sequenziert werden (Abb.: 4). Für die Herstellung der Exom-*Libraries* verwendeten wir das *SureSelect Human All Exon Kit* in den Versionen 38 Mb und v5 (Agilent Technologies [93, 38],) und arbeiteten nach Angaben des Herstellers. Das Kit in der Version v3 deckt 38 Mb des Genoms ab und wurde für die initiale Studie zur Identifizierung von genetischen Varianten bei M. Parkinson Patienten genutzt. Alle folgenden Proben wurden mit der v5 Version des Kits angereichert, welches 50 Mb des Genoms abdeckt. Zusätzlich dazu wurden die Exom-*Libraries* für das Parkinson-Projekt [69] über die in dieser Arbeit beschriebenen Semi-Automation generiert, während die Exom-*Libraries* des zweiten Projekts über die weitgehende Automation [94] mittels *Bravo Liquid Handling Plattform* der Firma Agilent hergestellt wurden.

Der initiale Schritt ist die Fragmentierung der hochmolekularen genomischen DNA. Der Prozess der *Cluster Generation* ist abhängig von der Größe der DNA-Fragmente und verläuft am effizientesten bei Fragmentlängen zwischen 200 und 500 bp [95]. Zum Fragmentieren genomischer DNA nutzen wir die *Adaptive Focused Acoustics* (AFA) Technologie der Firma Covaris [96, 97]. Dabei wird stoßweise akustische Ultraschallenergie freigesetzt, die fokussiert auf das Probengefäß trifft, in dem sich die in Wasser gelöste DNA befindet. Durch das Einwirken des Ultraschalls wird die hochmolekulare DNA geschert. Es werden zufällig doppelsträngige DNA-Fragmente mit 5'- und 3'-Überhängen (*sticky-ends*) generiert, die teilweise ein 5'-Phosphat tragen [95]. Als nächster Schritt werden mittels eines Enzym-Mixes *blunt-end*-Fragmente mit einem 5'-Phosphat generiert (*End-Repair*). Am 3'-Ende wird im Anschluss ein A-Nukleotid ligiert (*A-Tailing*), dessen Überhang

notwendig ist für die darauf folgende Ligation der Illumina-spezifischen Adapter (Adapter Ligation), die einen T-Nukleotid-Überhang tragen. In einem folgenden PCR-Schritt werden die *Libraries* amplifiziert um ausreichend Material für die Hybridisierung zu generieren. Nun wird die einzelsträngige DNA-Library an biotinylierte RNA-Moleküle hybridisiert. Die circa 120 Nukleotide langen RNA-Moleküle sind komplementär zu den exomischen Bereichen des Genoms [38]. Die Biotinylierung der RNAs erlaubt das selektive Rückgewinnen der DNA/RNA-Hybride über streptavidingekoppelte *Beads*. Im Anschluss an den Waschprozess wird das DNA/RNA-Hybrid aufgespalten und die *Library* in einer finalen PCR amplifiziert.

2.4.3 Hochdurchsatz-Sequenzierplattform Illumina, Inc.

Die in der vorliegenden Arbeit analysierten Exome und Genome wurden mit den *NGS*-Plattformen der Firma Illumina sequenziert [98]. Für die initiale Studie zur Identifizierung von genetischen Varianten bei M. Parkinson Patienten, sequenzierten wir ein Exom auf zwei *Lanes* einer *Flow Cell* als 54 bp *paired end*-Lauf. Dazu wurde der *GA IIx* genutzt. Alle Schritte wurden nach Angaben des Herstellers durchgeführt. Die Exome zweiten Studie bei Patienten mit Burn-McKeown-Syndrom wurden als 100 bp *paired-end*-Lauf auf dem HiSeq2500 System (Illumina) mit einem *Library-Pool* aus 12 Proben pro 4 *Lanes* durchgeführt. Die Genome der BMKS-Patienten wurden als 150 bp-Lauf auf dem HiSeq2500 System sequenziert. Vergleichende Kontaminationsanalysen fanden zwischen der HiSeq2500- und der HiSeq4000-Plattform statt.

2.4.4 Bioinformatik

Die Analyse der Primärdaten, die *Image* Analyse und das *Base Calling* wurden mit der *Illumina Real Time Analysis* Software durchgeführt [99]. Zum Demultiplexen der DNA-Pools nutzten wir CASAVA 1.8 (Illumina, Inc). Die daraus resultierenden Sequenzen (*Reads*) wurden mit dem *Burrows-Wheeler-Aligner* (BWA) [53] an das menschliche Referenzgenom (hg 19 GRCh37) *aligned*. Mit der Software SAMtools wurden Punktmutationen sowie kleinere Insertionen und Deletionen im Vergleich zur Referenz, spezifisch für die angereicherten Regionen, detektiert. Die Qualität aller genetischen Alterationen wurde mit dem SAMtools *varFilter* Skript berechnet [55]. Alle identifizierten Varianten wurden

anschließend über benutzerdefinierte Perl Skripte annotiert. Die Annotation enthält Informationen zu bekannten Transkripten (UCSC *Known Gene Track* [100] und RefSeq Gene [101]), bekannte Varianten aus dbSNP [102], die Art der genetischen Veränderung und wenn vorhanden, der Austausch der Aminosäure im betroffenen Protein. Die Daten der Annotation wurden anschließend in unsere interne Datenbank importiert, wo sie für eine Filterung und Priorisierung zur Verfügung stehen. Unabhängig vom Studiendesign schlossen wir alle Varianten mit einer SNV-Qualität kleiner als 30 und Veränderungen die in weniger als 15 % der Reads auftraten aus. Zusätzlich sollten mindestens 5 % der Reads von unterschiedlichen Strängen (Vorwärts- oder Rückwärtsstrang) kommen. Wir exkludierten alle Alterationen, die mit einer Allelfrequenz größer als 0.02 in der jeweils aktuellen Version von dbSNP auftraten.

Zusätzliche Aussagen zur einer genetischen Veränderung können computergestützte Prädiktionsprogramme geben. Dazu werden vorrangig die Programme PolyPhen2 [63] und Sift [62] angewendet, die anhand von evolutionärer Konservierung, Nukleotidposition und der Art der Aminosäureveränderung eine Pathogenität errechnen. Die Werte beider Programme liegen zwischen 0 und 1. Dabei sind niedrige Werte bei Sift als sicher schädlich eingestuft und hohe Werte als wahrscheinlich benign. Gegensätzlich verhält sich die Kalkulation von PolyPhen2, bei dem Wert um 1 eine Variante als pathogen und Werte um 0 eine Variante als benign deklarieren. Die Angaben stellen eine Wahrscheinlichkeit dar und geben lediglich einen Hinweis auf eine mögliche Pathogenität. Häufig sind genetische Varianten, welche schon mit einem pathogenen Phänotyp in Verbindung gebracht wurden, in öffentlichen Datenbanken gelistet. So gibt unsere interne Datenbank zusätzlich Aussage darüber, ob ein Gen oder eine Variante in den Datenbanken: OMIM [6], HGMD [64] oder ClinVar [103] gelistet ist. Gleichzeitig erhalten wir die Frequenzinformationen aus dem Datensatz des 1000 Genome Projekts [104] und des *Exom Variant Servers* [105]. All diese Informationen können neben den eigentlichen Priorisierungsstrategien zu einer verbesserten Filterung der Varianten beitragen. Zur Visualisierung der einzelnen *Reads* nutzen wir den *Integrative Genomics Viewer* (IGV) (Broad Institute) [106]. Diese manuelle Untersuchung unterstützt die Bewertung einer Mutation in ihrem genetischen Umfeld. So bekommt man Informationen über mögliche GC-reiche Sequenzen oder umgebende repetitive Elemente und kann dadurch falsch positive Varianten minimieren.

3 Ergebnisse

Die Ergebnisse der vorliegenden Arbeit gliedern sich in drei Abschnitte:

Der erste Teil behandelt die Entwicklung eines Labor-Informations- und Management-Systems für die *Next-Generation*-Sequenzierung. Er beschreibt den Aufbau und die Funktionsweise der drei Ebenen des LIMS, die Einführung von *Sample Manifest*, *Sample Tracking Sheet* und 2D-Barcodes zur Qualitätssicherung, sowie der LIMS-basierten Rekalkulation von Sequenzier-*Pools* zum Erreichen einer homogenen Menge an Sequenzierdaten innerhalb eines *Library-Pools*. Dazu beschreibt der Abschnitt die Einführung und Anwendung einer auf *NGS*-Daten basierenden Kontaminationsbestimmung.

Der zweite Teil der vorliegenden Arbeit beschreibt die Etablierung einer automatisierten *Library*-Präparation. Die Einführung erfolgte in zwei Schritten von einer Semi-Automation zur weitgehenden Automation. Es wird die Qualitätssicherung, Standardisierung und Optimierung der Laborabläufe und eingehenden Proben beschrieben. Zusätzlich finden sich die Ergebnisse der Austestung eines Anreicherungsprotokolls mit weniger DNA-Startmaterial in Verbindung mit einer erhöhten *Insert*-Größe bei exomischen *Libraries*.

Die Möglichkeiten, sowie die Grenzen der Exom-Sequenzierung, zur Identifikation pathogener Varianten in häufigen und seltenen Phänotypen, beschreibt der dritte Teil der Arbeit. Es wird gezeigt, dass neben der Sequenzierung von qualitativ hochwertiger genomischer DNA, auch DNA aus FFPE-Gewebe mit verminderter Qualität und Quantität erfolgreich in eine exomische *Library* konvertiert und sequenziert werden kann. Es wird die publizierte Studie eines frühen Exoms vorgestellt, bei der durch Exom-Sequenzierung die kausative Variante in drei Familien mit Morbus Parkinson identifiziert wurde. Eine zweite publizierte Studie beschreibt die Kombination aus Exom- und Genom-Sequenzierung zur Identifikation pathogener *compound*-heterozygoter-Varianten in *TXNL4* bei Patienten mit Burn-McKeown-Syndrom und zeigt die Grenzen der Exom-Sequenzierung auf.

3.1 Entwicklung eines Labor-Informations- und Management-Systems für die *Next-Generation*-Sequenzierung

Zu Beginn der Etablierung der *Next-Generation*-Sequenzierung wurden klassische Tabellendokumente zur Dokumentation und Verwaltung von Proben und Arbeitsabläufen genutzt. Mit steigendem Probenaufkommen wurde dieses Verfahren unübersichtlich und eignete sich nicht mehr für die Weitergabe von Daten. Um Proben- und Prozessinformationen sowie die Ergebnisse einer *NG*-Sequenzierung in ein standardisiertes System zu bringen und für die Bearbeiter zu jederzeit zugänglich zu machen, wurde ein web-basiertes **Labor-Informations- und Management-System** (LIMS) entwickelt. Tim Strom, Sebastian Eck und Thomas Wieland übernahmen dabei die Programmierarbeit. Zum Zeitpunkt des Beginns der *Next-Generation*-Sequenzierung am Institut für Humangenetik, war solch ein System mit allen notwendigen Anwendungen für den *NGS*-Bereich nicht verfügbar. Mittlerweile sind kommerzielle [107] und frei verfügbare [108, 109] Software-Anwendungen erhältlich.

3.1.1 Die Anforderungen an ein LIMS

Das Ziel eines LIMS ist es, administrative Aufgaben zu unterstützen, Arbeitsabläufe effizient zu gestalten, sowie die Daten sicher und in einfach zugänglicher Form zu speichern [44]. Die Übersichtlichkeit des vielschichtigen LIMS wird durch die Einteilung in drei verschiedene Ebenen sichergestellt. Zu den administrativen Aufgaben zählen die Proben-, Patienten- und Projektverwaltung, die innerhalb der Probenebene bearbeitet werden (Abb.: 6A). Die Prozessebene dient der Verwaltung von Laborabläufen beginnend mit der *Library*-Präparation bis hin zur Sequenzierung (Abb.: 6B). Die Hauptaufgabe der Analyseebene ist die web-basierte Auswertung der erhaltenen *NGS*-Daten. Die für das Labor relevante Anwendung der Analyseebene ist die Einsicht in die Qualitätsparameter einer sequenzierten *Library* (Abb.: 6C). Bei der Programmierung wurde berücksichtigt, dass die Benutzerschnittstelle übersichtlich und intuitiv angelegt ist. Daten sollen schnell und einfach eingetragen, gesucht und geändert werden können. Es sollte Funktionen zur Übersicht des Bearbeitungsstaus einzelner Proben geben, sowie eine Standardisierung aller notwendigen Rechenschritte innerhalb der Probenprozes-

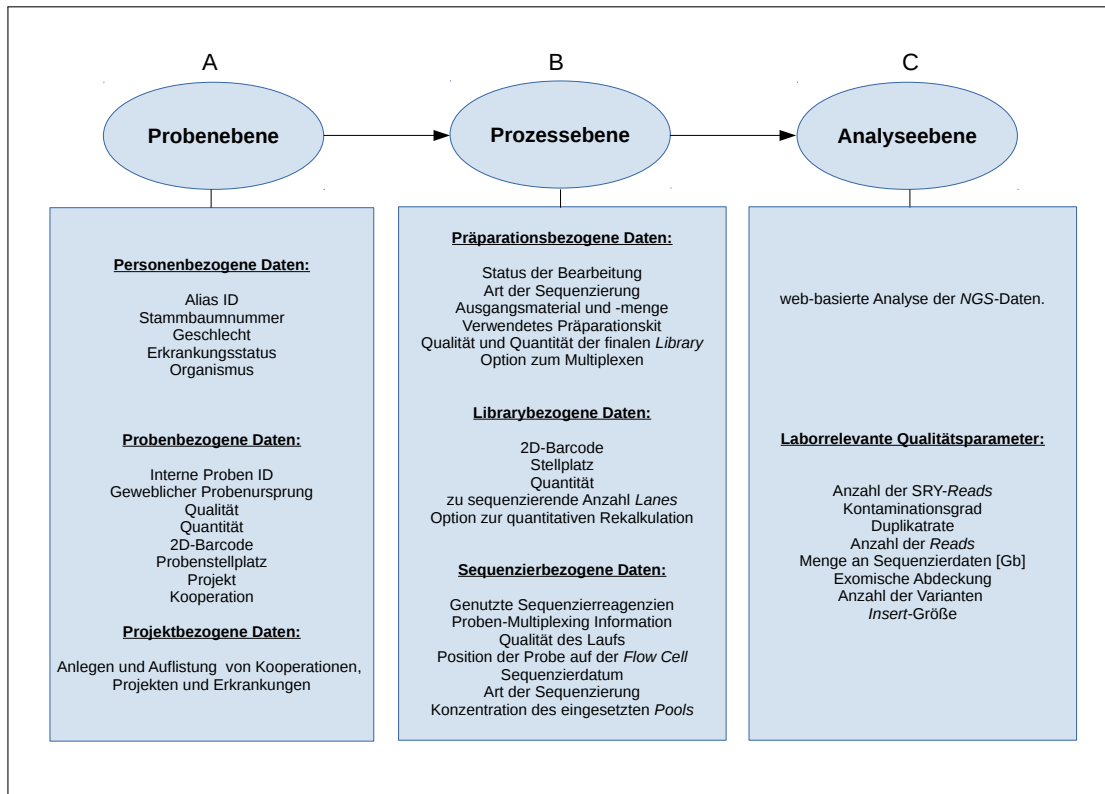


Abbildung 6: Die Architektur des LIMS. (A) Innerhalb der Probenebene werden personen-, proben- und projektspezifische Informationen registriert. (B) Das Prozess-Management findet über die Prozessebene des LIMS statt. (C) Die Analyseebene enthält laborrelevante Qualitätsparameter, dient aber hauptsächlich als Web-Anwendung zur NGS-Datenanalyse durch interne und externe Kooperationspartner.

sierung enthalten. Des Weiteren sollten LIMS-Anwendungen entwickelt werden, welche die Qualitätssicherung unterstützen (*Sample Manifest, Sample Tracking Sheet, 2D-Barcodes, Berechnungen zur Rekalkulation von Sequenzier-Pools* sowie eine Kontaminationsdetektion).

Bei der Verarbeitung patientenbezogener Daten ist die Einhaltung des Datenschutzes von großer Wichtigkeit. Der Zugriff auf die Datenbank wird ausschließlich autorisierten Nutzern über eine Nutzerkennung und ein Passwort gewährleistet. Die Anwendungen innerhalb der Proben- und Prozessebene, ist den Mitgliedern der NGS-Gruppe vorbehalten, während die Analyseebene auch Projektpartnern zur Auswertung ihrer Proben zugänglich gemacht werden kann. Der Zugang erfordert neben einer Nutzerkennung und dem Passwort zusätzlich die Eingabe eines Einmalpasswortes über den Yubikey. Die Zwei-Wege-Authentifizierung wird zur Erhöhung der Sicherheit verwendet.

Search samples	Overview	New sample	Import internal samples	Create libraries	New disease	List disease	New cooperation	List cooperation	New Project	List Project	New invoice	List invoice	Import external samples	Import mtDNA samples	Statistics	Login / Logout
----------------	----------	------------	-------------------------	------------------	-------------	--------------	-----------------	------------------	-------------	--------------	-------------	--------------	-------------------------	----------------------	------------	----------------

Abbildung 7: Die Menüleiste der Probenebene mit ihren einzelnen Funktionen.

Abfragen über die Datenbank zu Patienten, Proben und Ergebnissen erfolgen in Suchmasken. Die Ausgabe der Suchergebnisse erfolgt in tabellarischer Form. Die einzelnen Funktionsmöglichkeiten der Applikation können in der oberen Menüleiste der jeweiligen Ebene ausgewählt werden (Abb.: 7). Insgesamt entstand auf Grund der vielen Anforderungen und der umfassenden Abläufe im Bereich des *NGS* eine komplexe Applikation, deren einzelne Funktionen im Folgenden dargestellt sind.

3.1.2 Die Probenebene

Die Probenebene dient der Probenregistrierung und -verwaltung. Hier finden sich zu jeder eingesandten Probe alle zur Prozessierung notwendigen Informationen. Neben dem Probeneingangsdatum wird die zu untersuchende DNA oder RNA mit einer internen Identifikationsnummer (ID) und einem individuellen Probenstellplatz definiert. Zusätzlich werden zu jeder Einsendung weitere Informationen vom Einsender abgefragt, welche für die finale Datenanalyse essentiell sind: die Proben-ID des Einsenders, die Stammbaummer, das Geschlecht, der Erkrankungsstatus (betroffen oder nicht betroffen), der Organismus, aus welchem Gewebe das Probenmaterial isoliert wurde, welche Erkrankung vorliegt sowie Parameter zur Nukleinsäure Quantität und Qualität (Abb.: 8). Die Patientendaten werden in pseudonymisierter Form angefordert und in der Datenbank hinterlegt. Jede Probe wird einem Projekt und Kooperationspartner zugeordnet um die Datenanalyse zugänglich zu machen und die Kosten abzurechnen.

Über die Suchfunktion können Proben über ihre ID-Nummer aufgerufen werden und Stammdaten korrigiert oder Kommentare hinterlegt werden. Neben der Registrierung und Verwaltung von Proben dient die Probenebene zum Einsehen und Anlegen neuer Stammdaten, wie Erkrankungen, Kooperationspartner und Projekte (Abb.: 9). Im Bereich *Overview* können Projekte mit den zugehörigen Proben und deren Bearbeitungsstatus abgerufen werden (Abb.: 10) um einen schnellen Überblick über erledigte oder noch ausstehende Proben zu erhalten.

Search samples	Overview	New sample	Import internal samples	Create libraries	New disease	List disease	New cooperation	List cooperation	New Project
New Sample									
Id	<input type="text"/>								
Sample id	<input type="text"/>								
Foreign ID	<input type="text"/>								
Pedigree	<input type="text"/>								
Father	<input type="text"/>								
Mother	<input type="text"/>								
Comment	<input type="text"/>								
Sex	<input checked="" type="radio"/> unknown <input type="radio"/> male <input type="radio"/> female								
Affected	<input checked="" type="radio"/> unknown <input type="radio"/> affected <input type="radio"/> unaffected								
Organism	<input type="text"/>								
Tissue	<input type="text"/>								
Project	<input type="text"/>								
Cooperation	<input type="text"/>								
Concentration (ng/ul)	<input type="text"/>								
Volume (ul)	<input type="text"/>								
A260/280	<input type="text"/>								
A260/230	<input type="text"/>								
Barcode	<input type="text"/>								
Plate	<input type="text"/>								
Row	<input type="text"/>								
Column	<input type="text"/>								
Analysis	<input type="radio"/> other <input checked="" type="radio"/> exome								
Sequencing	<input checked="" type="radio"/> To sequence <input type="radio"/> Not to sequence								
BAM file	<input type="text"/>								
Invoice	<input type="text"/>								
Entered from	<input type="text"/>								
Entered	<input type="text" value="2017-02-07"/>								
<input type="button" value="Submit"/> <input type="button" value="Reset"/>									

Abbildung 8: Der Menüpunkt *New Sample*. Über *New Sample* kann eine Probe manuell in die Datenbank eingetragen werden. Eine nachträgliche Änderung an den Stammdaten ist jederzeit möglich.

A New Disease

Id:

Name:

Symbol:

OmimID:

Disease group:

Submit Reset

B New Cooperation

Id:

Name:

Prenome:

Institution:

Email:

Phone:

Submit Reset

C New Project

Project ID:

Project Name:

Description:

Comment:

Contact:

Entered from:

Date (yyyy-mm-dd):

Menuflag: False True

Submit Reset

Abbildung 9: Die Menüpunkte *New Disease*, *New Cooperation* und *New Project*. **A** Unter *New Disease* wird eine neue Erkrankung, mit der zugehörigen Abkürzung und, wenn vorhanden, der zugehörige OMIM-ID (*Online Mendelian Inheritance in Man*) eingetragen. Zusätzlich erfolgt die Einteilung in eine Erkrankungsgruppe wie zum Beispiel neurologische Erkrankungen oder Erkrankungen des Herzens. **B** Über den Menüpunkt *New Cooperation* kann ein Stammblatt eines neuen Kooperationspartners ausgefüllt werden. Dabei werden die Institution und die Kontaktdaten hinterlegt. **C** Über *New Project* werden Projekte in fortlaufender Nummerierung zusammen mit dem Kooperationspartner und Forschungsgebiet angelegt.

n	Id	Name	Cooperation	Organism	Libtype	Libpair	Samples	Libraries	to do	Results
1	S9999	Controls		human	genomic	paired-end	8	9	0	8
2	S9999	Controls		human	RNA	paired-end	4	4	0	4
3	S9999	Controls		human	exomic	paired-end	1	1	0	0

Abbildung 10: Der Menüpunkt *Overview*. Sortiert nach Projekten kann die Anzahl der Proben und *Libraries* eingesehen werden. Nicht bearbeitete Proben werden rot unterlegt.

3.1.3 Die Prozessebene

Die Prozessebene ist eine umfassende Struktur, welche die Verfolgung der Probe, von der Freigabe zur *Library*-Präparation bis hin zur eigentlichen Sequenzierung, ermöglicht (Abb.: 11).



Abbildung 11: Übersicht über die Funktionen der Prozessebene. Hier finden sich alle Anwendungen von der zu präparierenden *Library* bis hin zur Sequenzierung sowie der Reagenzien-Bestandskontrolle.

Zur Prozesskontrolle der Laborabläufe wird die *Library* im Verlauf des Präparations-Prozesses unterschiedlich gekennzeichnet. Beim anlegen einer *Library* in der Datenbank erhält die Probe den Status *to do*. Während der Herstellung der *Library* wird der tag *lib in process* gesetzt. Ist die *Library* fertig gestellt erscheint *Library prepared*, nach dem Pooling *pooled* und nach der Sequenzierung *sequenced*. Proben, die nicht bei uns im Labor bearbeitet wurden und von denen wir Sequenzierdaten erhalten, werden als *external* gekennzeichnet. Es kann sowohl über die Proben-ID direkt, als auch über den Proben-Status eine Suchabfrage erfolgen um den Bearbeitungsstand zu ermitteln.

Entspricht das eingesandte Material den Qualitäts- und Quantitätsvorgaben, wird die Probe freigegeben und eine *Library* angelegt (Abb.: 12). Dabei wird definiert welche Analyse durchgeführt wird (Exom-, Genom-, Transkriptom-Analyse), welches Kit zur Herstellung verwendet wird, sowie die Länge und Art der Sequenzierung.

Im Anschluss an die *Library*-Präparation werden alle wichtigen Qualitäts- und Quantitäts-Parameter sowie der Index der *Library* im LIMS hinterlegt und zeitgleich der Status *Library prepared* vergeben. Zusätzlich wird festgelegt, ob es sich um Proben handelt, die vor der Sequenzierung *gepoolt* werden sollen. Alle zu *poolenden* Proben werden über den Menüpunkt *Pooling* ausgegeben und die nötigen Volumina mit der hinterlegten Gleichung berechnet (Abb.: 13 A-D). Wird ein *Library-Pool* in der Datenbank angelegt, wird der Status *pooled* vergeben und die Proben für die Sequenzierung freigegeben.

Zur Sequenzierung der Proben unterstützt die Applikation mit den Berechnungen zur Verdünnung der einzelnen Pools, sowie zur Belegung der *Flow Cell*. Im Menüpunkt *New Run* wird die finale Belegung der *Flow Cell* definiert.

Search samples
Overview
New sample
Import internal samples
Create library
New disease

Search

DNA ID
 Foreign ID
 Pedigree
 Type material
 Library All Without library
 Disease
 Status
 For pool No Yes All
 Cooperation
 Project
 Order Sample Pedigree Status

Create Libraries for Exome
Create Libraries for Genome
Create Libraries for RNAseq
Create Libraries for ChIPseq

Description
 Comment
 Exome kit
 Type material
 Type library
 Type assay
 Date (yyyy-mm-dd)
 Label library 'preparation in process' No Yes

n	Box	id	DNA	Plate	Row	Column	Foreign id	Pedigree	Sex	Disease	Analysis
1	<input type="checkbox"/>	13465	55966G1				controlDNA	55966G1	female		other
2	<input type="checkbox"/>	13465	55966G1				controlDNA	55966G1	female		other
3	<input type="checkbox"/>	15249	55966G2				controlDNA	55966G2	female		other

Abbildung 12: Der Menüpunkt *Create Library*. Über *Create Library* wird aus der registrierten Probe eine *Library* generiert. An dieser Stelle entscheidet sich, aus welcher Probe eine Präparation der *Library* erfolgt. Mit dem Anlegen der *Libraries* werden weitere Parameter (Kit, Material, Art der *Library*) definiert und die Probe erhält automatisch den Status *to do*.

A

Pooling

Number of samples	<input type="text" value="12"/>
Final volume (ul)	<input type="text" value="50"/>
Final concentration (nMol)	<input type="text" value="10"/>
Disease	<input type="text"/>
Project	<input type="text"/>
For pool	<input type="radio"/> No <input checked="" type="radio"/> Yes <input type="radio"/> All
Cooperation	<input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

B

$$\mu\text{l library} = \frac{V(p) * C(p)}{\# * C(l)}$$

V(p) = Volumen des Pools
C(p) = Konzentration des Pools
 $\#$ = Anzahl de *libraries*
C(l) = Initiale Konzentration der *library*

C

Pooling Sheet

Label library as pooled: No Yes

n	Box	Library	Pool	Barcode	Plate	Row	Column	Index1	Index2	Status	BA size	BA hs size	PicoGreen (nMol)	qPCR (nMol)	ul for 10 nMol (PicoGreen)	ul for 10 nMol (qPCR)
1	<input checked="" type="checkbox"/>	ULM0291_LIB1	ULM0291_LIB1	0215049565	BOX137	B	04	Agilent02		library prepared	341	0	7.0	0.0	23.9	
2	<input checked="" type="checkbox"/>	103867_LIB1	103867_LIB1	0219878973	BOX159	E	02	Agilent78		library prepared	310		6.9	0.0	24.1	
3	<input checked="" type="checkbox"/>	PAD0011_LIB1	PAD0011_LIB1	0183323198	BOX152	A	01	Agilent91		library prepared	293	0	11.4	0.0	14.7	
4	<input type="checkbox"/>	ERL0029G_LIB1	ERL0029G_LIB1	0226410866				D708	D507	library prepared		0	0.0	10.0		16.7
5	<input type="checkbox"/>	ERL0030G_LIB1	ERL0030G_LIB1	0226410905				D709	D508	library prepared		0	0.0	10.5		15.8

D

Pooling sheet

2017-05-10

14268 35783 ULM0291_LIB1 ULM0291_LIB1 0215049565 BOX137 B 04 Agilent02 library prepared 341 0 7.0 0.0 23.9
 16982 39268 103867_LIB1 103867_LIB1 0219878973 BOX159 E 02 Agilent78 library prepared 310 6.9 0.0 24.1
 14794 36825 PAD0011_LIB1 PAD0011_LIB1 0183323198 BOX152 A 01 Agilent91 library prepared 293 0 11.4 0.0 14.7

n	Library	Pool	Barcode	Plate	Row	Column	Index1	Index2	Status	BA size	BA hs size	PicoGreen (nMol)	qPCR (nMol)	ul for 10 nMol (PicoGreen)	ul for 10 nMol (qPCR)
1	ULM0291_LIB1	ULM0291_LIB1	0215049565	BOX137	B	04	Agilent02		library prepared	341	0	7.0	0.0	23.9	
2	103867_LIB1	103867_LIB1	0219878973	BOX159	E	02	Agilent78		library prepared	310		6.9	0.0	24.1	
3	PAD0011_LIB1	PAD0011_LIB1	0183323198	BOX152	A	01	Agilent91		library prepared	293	0	11.4	0.0	14.7	

Volume PicoGreen 62.7 ul
 Volume qPCR 0 ul

Abbildung 13: Übersicht des Menüpunkts *Pooling*. (A) Über den Menüpunkt *Pooling*, werden die zu *poolenden* Proben definiert und die Variablen der Kalkulation festgelegt. (B) Die Berechnung der nötigen Volumina erfolgt anhand der hinterlegten Formel. (C) Zur Auswahl erscheinen alle zu *poolenden* Proben mit den notwendigen Informationen über den Lagerplatz, den Index und die Quantität. (D) Nach der Probenselektion erfolgt die finale Ausgabe der Proben, Probenvolumina und des Gesamtvolumens.

Search	Import Libsheet	Taqman	Import Taqman	Pooling	Make Pool	Sequencing	New Run	Check Pool Barcode	New Index	List Indices	New Stock
New Run											
Run ID	<input type="text"/>										
Flowcell Barcode	<input type="text"/>										
Description	<input type="text"/>										
Index run	<input checked="" type="radio"/> No <input type="radio"/> Yes										
Comment	<input type="text"/>										
Failed comment	<input type="text"/>										
Lane1 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Lane2 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Lane3 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Lane4 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Lane5 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Lane6 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Lane7 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Lane8 Pool barcode	<input type="text"/>	Pool Description	<input type="text"/>	Pool Id	<input type="text"/>						
Molarity (pM)	<input type="text"/>										
Flow-cell (yyyy-mm-dd)	<input type="text" value="2017-02-07"/>										
Run (yyyy-mm-dd)	<input type="text" value="2017-02-07"/>										
Directory	<input type="text"/>										
Menuflag	<input type="radio"/> False <input checked="" type="radio"/> True										
<input type="button" value="Submit"/> <input type="button" value="Reset"/>											

Abbildung 14: Die Funktion *New Run*. Über den Reiter *New Run* wird die *Flow Cell* mit ihrer ID, der Art des Laufs und der Belegung definiert. Zeitgleich wird das Datum der *Flow Cell*-Generierung sowie der Bearbeiter hinterlegt.

Dazu kommt der *Flow Cell*-Barcode, die Art des Laufs und die genutzte Molarität zum *Clustern* der *Flow Cell* (Abb.: 14). Im Anschluss werden zum Lauf die genutzten Reagenzien hinterlegt. Das dient zum einen der Bestandskontrolle als auch der Registrierung der Reagenzien Chargen-Nummern. So kann im Falle einer nachträglichen Fehlersuche die genutzte Chemie genau definiert werden.

Neben der Prozesskontrolle dient die Prozessebene auch zur Kontrolle der Lagerbestände aller nötigen Reagenzien für die *Library*-Präparation und die Sequenzierung. Wird eine *Library* angelegt, wird zu dieser ein Reagenzien-Kit definiert und eine Reaktion als gebraucht gezählt. Das gleiche geschieht beim Anlegen eines neuen HiSeq-Laufs. Die Bestände können zu jeder Zeit abgefragt und neue Bestellungen ausgelöst werden. Über die Menüpunkte *New Order* und *List Order* können Bestellungen angelegt und eingesehen werden.

Autosomal Chromosomes		Autosomal X-chromosomes		Y-chromosome	De novo	Y-chrom	Genes	Discordance	Coverage	Coverage ratio	Pages	Homozygosity	CNVs	RGND	GMN	IFD	IBS	Comments	Conclusions	Quality control	Samples	PFM	PA	Help	Logout						
Search results																															
Show <input type="text" value="10000"/> entries																															
#	ID	Case	Chromosome	Pedigree	Sex	Foreign ID	SRY	Cooperation	Libraries	libType	libpair	type	Contam.	Duplicates	Optical duplicates	Reads	Mapped	Mapped percent	Properly paired	Seq (Gb)	Read length	on ball	Avg cov (estimated)	Uncovered	Cov 1x	Cov 4x	Cov 8x	Cov 16x	SNV	Indel	Phased
1	TUEB020	1 P066A	ARRP_260	female	27284	2	Wissinger	TUEB020_LIB1	exomic	paired-end	SureSelectHS040bV6	0.01085	15.00	7.00	136549391	136363024	99.86	99.09	13.770	100	77.06	145.51	0.30	99.70	99.57	99.41	98.56	2.99	79921	6680	801
2	TUEB020	1 P066A	ARRP_307	male	28061	631	Wissinger	TUEB020_LIB1	exomic	paired-end	SureSelectHS040bV6	0.01214	17.00	7.00	127781232	127597957	99.82	99.14	12.880	100	77.24	135.73	0.10	99.90	99.81	99.67	98.83	3.07	78763	6536	830
3	TUEB023	1 P066A	ARRP_314	male	27929	628	Wissinger	TUEB023_LIB1	exomic	paired-end	SureSelectHS040bV6	0.01181	14.00	7.00	159924191	159625502	99.81	99.93	16.120	100	75.63	166.79	0.09	99.91	99.84	99.73	99.13	2.99	77421	6853	815
4	TUEB024	1 P066A	ARRP_316	female	27939	0	Wissinger	TUEB024_LIB1	exomic	paired-end	SureSelectHS040bV6	0.01107	19.00	6.00	173019877	172742580	99.84	99.23	26.080	150	79.40	255.56	0.27	99.73	99.63	99.57	99.32	2.99	80031	7252	803
5	TUEB025	1 P066A	ARRP_317	female	28004	0	Wissinger	TUEB025_LIB1	exomic	paired-end	SureSelectHS040bV6	0.01443	27.00	9.00	131103563	130838187	99.80	99.19	13.210	100	75.92	136.93	0.27	99.73	99.61	99.50	98.81	2.93	79195	6752	781
6	TUEB028	1 P066A	ARRP_322	male	28220	691	Wissinger	TUEB028_LIB1	exomic	paired-end	SureSelectHS040bV6	0.04009	20.00	9.00	133395446	133177962	99.84	98.91	13.450	100	77.35	142.14	0.09	99.91	99.82	99.67	98.82	2.99	79069	6566	747

Abbildung 15: Qualitätskontrolle innerhalb der Analyseebene. Die kalkulierten Ergebnisse der Sequenzierung werden hier angezeigt und können validiert werden. Parameter die nicht in den Spezifikationen liegen werden rot unterlegt.

3.1.4 Die Qualitätssicherung innerhalb der Analyseebene

Die Analyseebene dient der Darstellung und Auswertung der Sequenzierungsergebnisse. Sie ist der Teil des LIMS, der für externe Kooperationspartner zugänglich gemacht werden kann. Die Legitimation erfolgt über einen Benutzernamen, ein Kennwort sowie ein Einmalpasswort im Sinne einer Zwei-Wege Authentifizierung. Die Anwendung kann nach der Sequenzierung und Datenanalyse genutzt werden.

Innerhalb Analyseebene können laborrelevante Qualitätsparameter abgerufen werden um die erfolgreiche Sequenzierung einer Probe zu beurteilen. Die Funktion wird über den Menüpunkt *Quality Control* aufgerufen. Die Suche nach entsprechenden Proben kann über einen Zeitraum erfolgen, über die Art der *Library* oder gezielt per DNA-Identifikationsnummer. Im Ergebnis erscheint in tabellarischer Form eine Auflistung aller relevanter Daten (Abb.: 15). Wichtige Qualitätsparameter sind die Anzahl der *SRY-Reads*, Kontamination, Menge der Sequenzierdaten, sowie die Anzahl der identifizierten Varianten. Zur besseren Ansicht werden Werte, die außerhalb des Normbereichs liegen rot markiert. Hier kann ein schneller Überblick über die erfolgreiche Sequenzierung einer Probe gewonnen werden. Hat die Sequenzierung einer Probe zu wenig Datenmenge erzeugt, kann die *Library* nachsequenziert werden. Bei Kontaminationen oder Geschlechtsdiskordanzen wird neues Material für eine erneute Sequenzierung angefordert.

3.1.5 Erweiterung und Entwicklung des LIMS durch zusätzliche Anwendungen

Die Entwicklung des LIMS begann mit dem Einzug der *Next-Generation-Sequenzierung* am Institut für Humangenetik und entwickelte sich durch Veränderungen der Anforderungen stetig weiter. Der steigende Durchsatz der im Labor zu prozessierenden Proben zeigte Schwachstellen auf, die mit neu implementierten LIMS-Anwendungen vermieden wurden. Der folgende Abschnitt beschreibt die Einführung des *Sample Manifest*, des *Sample Tracking Sheets* und einer Implementierung von 2D-Barcode Röhrchen. Diese Erweiterungen dienen der Vermeidung von Fehlern, welche durch das manuelle Eintragen von Parametern in die Datenbank, entstehen können. Eine weitere Ergänzung ist die Einführung der LIMS-gestützten Rekalkulation von Sequenzier-*Pools* um eine homogenen Menge an Sequenzierdaten zu erreichen und einer damit einhergehenden Vermeidung von über- oder unterrepräsentierten Proben. Zusätzlich wurde die Kontaminationsbestimmung anhand von *NGS*-Daten implementiert, um die Identität der sequenzierten Proben sicherzustellen.

3.1.5.1 Einführung des *Sample Manifest*

Zu Beginn der Entwicklung des LIMS wurden alle Probeneingänge manuell über die Probenebene eingetragen (Abb.: 8). Bei der Übernahme von Proben- und Patientendaten kam es zu Fehlern in Form von Zahlendrehern oder falsch ausgefüllten Feldern durch Verwechslungen. Aufgrund des steigenden Probenaufkommens, war die manuelle Eintragung zusätzlich sehr zeitaufwendig. Für eine rasche und fehlerfreie Übernahme der Proben in die Datenbank, wurde ein *Sample Manifest* (SAM) entwickelt, welches alle notwendigen Informationen abfragt und es gleichzeitig ermöglicht mehrere Proben in einem Schritt zu importieren ohne manuelle Eingaben tätigen zu müssen. Das *Sample Manifest* wird dem Einsender im Excel-Format zur Verfügung gestellt und im Dateiformat csv in die Datenbank importiert (Abb.: 16). Die im SAM abgefragten Parameter sind zwingend anzugeben. Dies ermöglicht eine standardisierte Probenaufzeichnung und garantiert, dass alle, zur Analyse notwendigen, Daten vorhanden sind. Die Menge der im *Sample Manifest* mitgeteilten Proben schwankt zwischen zwei und 200. Unabhängig von der Anzahl der Proben ist der Import innerhalb weniger Sekunden getätigt.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Sample ID	Foreign ID	Pedigree	Comment	Sex	Affected	Organism	Tissue	Disease	Concentration (ng/ul)	Volume (ul)	A260/280	A260/230	Barcode	Plate	Row	Column	Analysis	Cooperation	Project	
2		Index	female	1	human	peripheral blood		Alzheim syndrome	163.8883276163	61.0246052148	1.93	2.2	241841251	ExLib134	F	3	exome			
3		Index	male	1	human	peripheral blood		Heart failure	219.6445407314	45.5281063062	2.01	2.13	241841251	ExLib134	F	4	exome			
4		Mother	female	0	human	peripheral blood		Metabolic disease	103.3780666386	96.7323178422	1.81	2.15	241841276	ExLib134	F	5	exome			
5		Mother	female	0	human	peripheral blood		Epilepsy	163.4972217207	61.1631188271	1.95	1.99	241841284	ExLib134	F	9	exome			
6		Index	female	1	human	other		Alzheim	170.3175462879	58.7138566633	1.94	2.04	241841285	ExLib134	H	7	exome			
7		Index	female	1	human	other		Parkinson disease	115.0729172935	86.9014207269	1.94	1.96	241841250	ExLib134	H	8	exome			
8		Mother	female	0	human	peripheral blood		Heart failure	234.3883282181	42.8734349655	1.84	2.14	241841258	ExLib134	H	9	exome			
9		Index	male	0	human	peripheral blood		Heart failure	142.6551122345	70.0991352035	1.83	2.14	241841247	ExLib134	H	10	exome			
10		Index	male	1	human	peripheral blood		Mitochondrial disease	199.1935969189	50.2024169184	1.96	2.1	241841262	ExLib134	H	11	exome			
11		Index	female	1	human	peripheral blood		spastic paresis	144.8770730395	69.0240339958	1.83	2.16	241841266	ExLib135	A	11	exome			
12		Index 2	female	1	human	peripheral blood		spastic paresis	155.3128322401	64.3861801744	1.83	2.11	241841283	ExLib135	A	12	exome			
13		Mother	female	0	human	peripheral blood		developmental disorder	166.3557401055	60.1171427711	1.84	2.2	241841260	ExLib135	B	2	exome			
14		Father	male	0	human	peripheral blood		developmental disorder	114.5714229401	97.261798774	1.83	2.12	241840629	ExLib135	B	3	exome			
15		Index	male	1	human	peripheral blood		developmental disorder	60.4300816433	165.480497925	1.69	1.91	241840632	ExLib135	B	4	exome			
16		Index_edig	male	1	human	peripheral blood		Mitochondrial disease	89.2961023851	111.986970684	1.96	1.91	241840649	ExLib135	B	5	exome			
17		mother	female	0	human	peripheral blood		Mitochondrial disease	129.3755931065	77.2649638642	1.84	2.13	241840623	ExLib135	B	6	exome			
18		father	male	0	human	peripheral blood		Mitochondrial disease	102.7762732844	97.298721577	1.82	2.12	241840624	ExLib135	B	7	exome			
19		Index_edig	female	1	human	peripheral blood		Neuromuscular disease	183.0550504901	54.6262395485	1.89	2.06	241840631	ExLib135	B	8	exome			
20		mother	female	0	human	peripheral blood		Neuromuscular disease	223.9674229033	44.6463936484	1.95	2.12	241840633	ExLib135	B	9	exome			
21		father	male	0	human	peripheral blood		Neuromuscular disease	170.036709394	58.8108299416	1.81	2.11	241840626	ExLib135	B	10	exome			

Abbildung 16: Das Sample Manifest. Es enthält alle relevanten Parameter, die zur Bearbeitung und Auswertung der Proben notwendig sind. Die Identifikationsnummern und Stammbaumnummern der Patienten werden in pseudonymisierter Form hinterlegt und sind aus Datenschutzgründen in der Abbildung nicht dargestellt.

3.1.5.2 Einführung des *Sample Tracking Sheets*

Sollte eine Probe im Labor in den Ablauf der *Library*-Präparation gebracht werden, wurden dies zu Beginn über Tabellendokumente oder mündlich kommuniziert. Die Berechnungen zur initialen Verdünnung der DNAs für die Fragmentierung wurden manuell durchgeführt. Im Anschluss an die *Library*-Präparation wurden die Parameter der Qualitätskontrolle, die verwendeten Indizes und die finale *Library*-Quantität per Hand in die Datenbank eingetragen. Ebenso wie bei der Eintragung der Proben- und Patientendaten kam es hier zu Fehleinträgen. Der Probendurchsatz war zu Anfang der *NG*-Sequenzierung gering und die Probenlisten übersichtlich. Die Automation der *Library*-Präparation erlaubt nunmehr die parallele Herstellung von 96 *Libraries*. Daher spielt neben der Einführung von Fehlern durch die manuelle Eingabe, auch die Bearbeitungszeit eine Rolle.

Für die Bearbeitung der zu präparierenden Proben wird aus der Prozessebene des LIMS ein *Sample Tracking Sheet* (STS) generiert (Abb.: 17). Es enthält Informationen über die Proben-Identifikationsnummer, deren Stellplatz und 2D-Barcode (die Einführung von 2D-Barcode Röhren wird im nächsten Abschnitt beschrieben). Zusätzlich werden die Konzentrationen der Probe ausgegeben und die Verdünnung zur Fragmentierung automatisch kalkuliert. Das STS wird sukzessive während der *Library*-Präparation mit den Informationen zum vergebenen Index sowie den Quantitäten vor und nach der Hybridisierung (im Falle vom Exomen) ausgefüllt. Die Quantitäts- und Qualitätsparameter (ng/ μ l, nM, bp), die vom Bioanalyzer oder dem Caliper LabChip GX ausgegeben werden, können einfach über *copy/paste* in das STS übertragen werden. Am Ende der Prozessierung wird das STS mit allen relevanten Daten in das LIMS reimportiert. Ebenso wie beim SAM werden durch die Einführung des STS manuelle Einträge vermieden und die Bearbeitungszeit stark reduziert.

3.1.5.3 Einführung von 2D-Barcode Röhren

Der wohl kritischste Punkt bei der Prozessierung von Proben im Hochdurchsatz ist der Erhalt der Probenintegrität beginnend bei der Lagerung der Proben, über die Prozessierung, bis hin zur fertigen *Library* und dem *Library-Pool*. Eine Verwechslung von Proben kann falsch-positive oder falsch-negative Assoziationen

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Box	Column	Row	DNA Id	DNA Barcode	DNA Barcode control	Library	Pool	Pool Barcode	Number	Index1	Index2	Nano Drop (ng/ul)	Shearing DNA ul (1000 ng)	Shearing TE ul (50 - DNA)
2	ExLib097	2	E	88689G	226411404		88689G_LIB1	88689G_LIB1		1			236.5	4.2283298097	45.77167019
3	ExLib097	3	E	99593G	226410329		99593G_LIB1	99593G_LIB1		2			455.5	2.1953896817	47.80461032

Abbildung 17: Ausschnitt aus dem LIMS-generierten *Sample Tracking Sheet* (STS) für eine Genom *Library*-Präparation. Das STS wird aus den Daten der Probenebene generiert. Es inkludiert den Stellplatz und den 2D-Barcode der Probe und enthält die Kalkulation zur initialen Verdünnung der DNA (Spalte N und O). Im Laufe des Prozesses wird das STS mit verschiedenen Parametern gefüllt. Als erster Schritt wird der Barcode des DNA-Stocks eingescannt und abgeglichen. Es folgen die genutzten Indizes, Qualität und Quantität der *Libraries*, sowie der Barcode des *Library*-Röhrchens.

nach sich ziehen.

Klassischerweise wurden die DNAs oder RNAs von Patienten- oder Forschungsproben in Schraubröhrchen gelagert, die händisch beschriftet wurden. Das gleiche Prinzip wurde für fertige *Libraries* und *Pools* angewendet. Die Kontrolle, welche Probe prozessiert, *gepooled* oder sequenziert werden soll, fand durch einen Abgleich der Beschriftung mit denen in der Anforderung statt. Zusätzlich ist beim Anlegen der *Pools* und der Sequenzier-Läufe, eine Definition der Proben notwendig. Hier mussten bisher die *Library*-IDs manuell eingetragen werden. Diese Prozesse waren anfällig für Fehler.

Im Jahr 2011 führten wir die Proben- und *Library*-Lagerung in 2D-Barcode Röhrchen ein (Thermo Scientific, Matrix 0,5 ml 2D-Barcode Schraubverschlussröhrchen). Die Anwendung des Systems innerhalb der Probenprozessierung ist in Abbildung 18 dargestellt. Jede Patientenprobe, jede *Library* und jeder *Pool* bekommt ein individuelles Röhrchen mit ID-Beschriftung und 2D-Barcode. Der Barcode der Patientenprobe wird im *Sample Manifest* hinterlegt und importiert.

Zu Beginn der *Library*-Präparation findet ein ersten Kontrollschritt statt. Wird die Probe prozessiert, erscheint der Proben-Barcode im *Sample Tracking Sheet* (Abb.: 17 Spalte E). Durch einscannen der Barcodes in die *DNA Barcode control*-Spalte findet ein visueller Abgleich durch den Bearbeiter statt. Gibt es Diskrepanzen in der Barcode-Nomenklatur, wird zusätzlich beim Reimport des STS einer Fehlermeldung im LIMS ausgegeben.

Die fertige *Library* wird ebenfalls in ein 2D-Barcode Röhrchen überführt und der Barcode im STS hinterlegt. Zum Anlegen des *Library-Pools* im LIMS ist keine manuelle Eintragung der *Library*-IDs mehr nötig. Alle in einem *Pool* befindlichen *Libraries* werden einfach über ein Barcode-Lesegerät eingescannt. Der *Library*-

Pool wiederum erhält ebenfalls ein eigenes 2D-Barcode Röhrchen. Beim Anlegen des Sequenzier-Laufs wird dieser in das LIMS eingescannt und automatisch die Verknüpfung zu den einzelnen *Libraries* im Pool hergestellt.

Das *Sample Manifest* und das *Sample Tracking Sheet*, sowie die Einführung des 2D-Barcode Systems dienen der Vermeidung von Fehleinträgen durch manuelle Eintragungen, vereinfachen und beschleunigen den Prozess des Datenimports und standardisieren die Probenprozessierung.

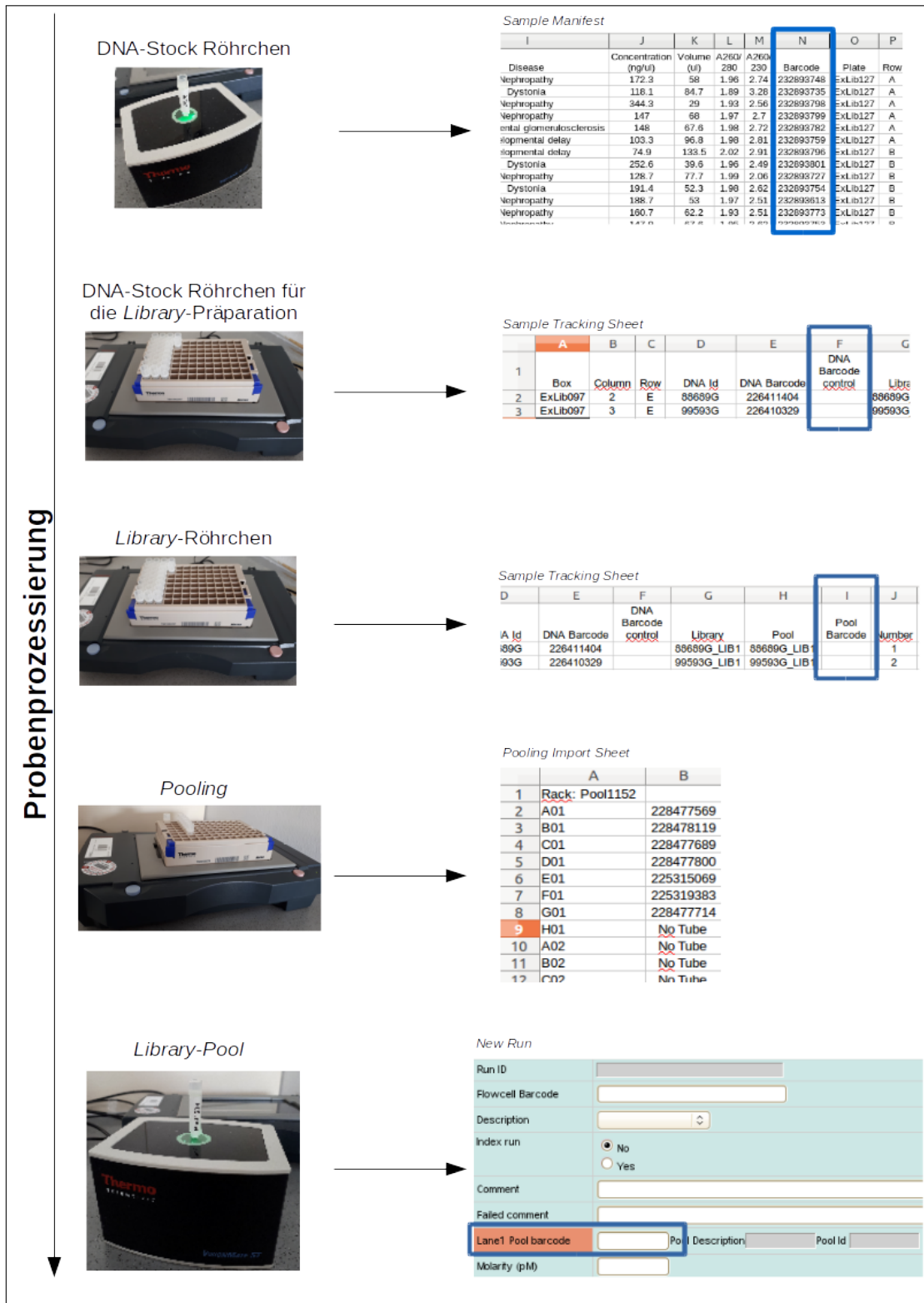


Abbildung 18: Übersicht zur Anwendung von 2D-Barcodes innerhalb der Probenprozessierung.

3.1.5.4 Rekalkulation von genomischen *Library-Pools*

Eine weitere Entwicklung und Verfeinerung des LIMS war die Einführung der Möglichkeit einer molaren Rekalkulation von *Library-Pools*. Der hohe Durchsatz der gegenwärtigen *NG*-Sequenzierplattformen machte es notwendig, das Proben miteinander gemischt und auf eine oder mehrere *Lanes* einer *Flow Cell* verteilt werden (*Pooling*). Zusätzlich ergibt sich dadurch der Effekt, das mögliche *Lane*-abhängige Sequenzierartefakte, wie ungleichmäßiges *Clustern*, besser ausbalanciert werden können.

Das *Pooling* ist ein äquimolares Mischen einzelner *Libraries* (Abb.: 13) zu einem *Library-Pool*. Das Ziel ist es, das jede Probe in gleichem Maße vertreten ist und die gleiche Menge an Sequenzierdaten produziert. Eine homogene Produktion von Daten erhöht die Vergleichbarkeit von Proben und verhindert das Über- oder Untersequenzieren von *Libraries*.

Die Vergangenheit hat gezeigt, das es praktisch nicht möglich ist das *Pooling* äquimolar auszulegen. Das zeigte sich in der ungleichmäßigen Datenmenge, die in der Sequenzierung erzeugt wurde (Abb.: 19A, B). Die Gründe für einen inhomogenen Pool können in der nicht 100 %ig genauen Messbarkeit der Quantität einer *Library*, als auch in der Varianz der Pipettiergenauigkeit des Bearbeiters bei der Herstellung der *Pools* liegen.

Genom-*Libraries* werden mittels qPCR quantifiziert und im Anschluss äquimolar gemischt. Ein Pool aus sechs Genomen wird auf einer HiSeq4000 *Flow Cell* verteilt um eine durchschnittliche Abdeckung von 40-fach pro Genom zu erhalten. Die Ergebnisse der Sequenzierung eines Genom-Pools ohne Rebalancierung sind in Abbildung 19A dargestellt. Die *Libraries* im *Pool* hatten eine durchschnittliche Abdeckung von 47-fach und schwankten zwischen 38-fach bis 61-fach. In Anlehnung an eine Poster-Präsentation des *Broad Institute* auf der AGBT Konferenz (*Advances in Genome Biology and Technology*) im Jahr 2016 [110], führten wir die Rebalancierung von Genom-*Pools* ein. Dazu wird vom *Pool* eine *Lane* einer *Flow Cell* sequenziert und die Datenmenge der einzelnen *Libraries* ermittelt. Abbildung 19B zeigt die durchschnittliche Abdeckung der *Libraries* eines unbalancierten 12-Plex, die zwischen 1,7-fach bis 4-fach schwankte. Für unterrepräsentierte Proben wurden die verhältnismäßigen Volumina der jeweiligen *Library* dem bestehenden Pool hinzugefügt (Abb.: 19C), und die restlichen

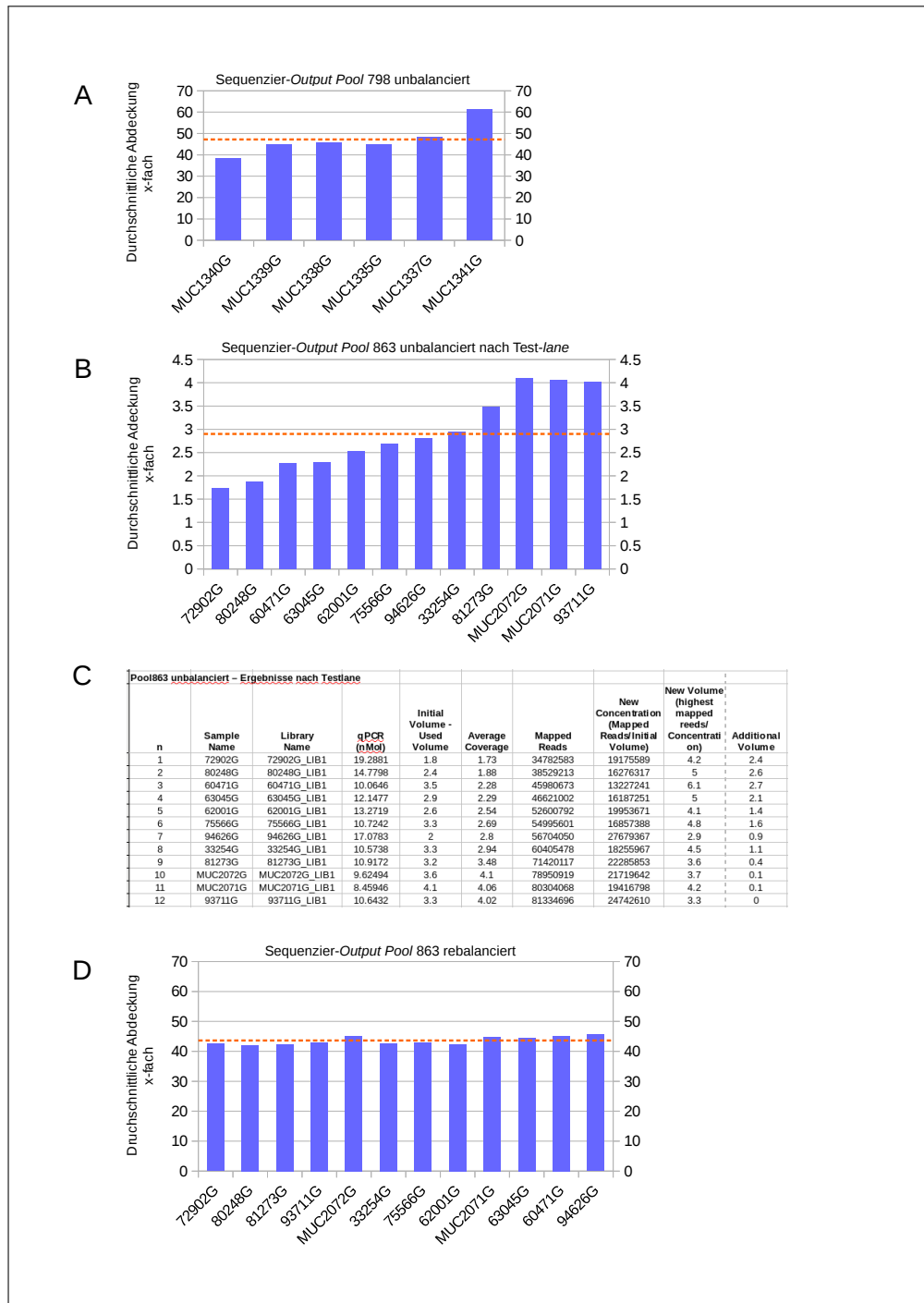


Abbildung 19: Rekalkulation von genomischen *Library-Pools*. Abbildung **A** zeigt die durchschnittliche Abdeckung einzelner *Libraries* eines unbalancierten Genom-Pools. Die Abdeckung schwankt um einen Mittelwert (gestrichelte Linie) von 47-fach von 33,4-fach bis 61,3-fach. Die Abbildung **B** zeigt die Resultate der Sequenzierung einer Test-Lane eines 12-Plex. Noch unbalanciert schwankt auch hier die durchschnittliche Abdeckung pro Genom von 1,3-fach bis 4-fach bei einem Mittelwert von 2,9-fach. **C** Die Volumina der unterrepräsentierten Genom-*Libraries* zur Balancierung des Pools werden durch das LIMS berechnet und ausgegeben. Die Ergebnisse zeigt die Abbildung **D** mit einer homogenen Abdeckung von 43,6-fach im Mittel.

sieben *Lanes* sequenziert. Das Sequenzier-Ergebnis nach der Rekalkulation des Pools ist in Abbildung 19D dargestellt. Es zeigte sich eine homogene Abdeckung der einzelnen Proben zwischen 42,5-fach und 45,6-fach.

Die Rekalkulation von genomischen *Library-Pools* ermöglicht das Multiplexen von mehreren Proben und resultiert in einer homogenen Menge an Sequenzierdaten. Einer Über- oder Unterrepräsentation von einzelnen *Libraries* wird vorgebeugt. Die Methode kann nicht nur auf genomische *Library-Pools* angewendet werden, sondern ist auf alle Arten von *Libraries*, wie Exome und Transkriptome erweiterbar. Ein Nachteil der Anwendung ist die verlängerte Bearbeitungszeit von durchschnittlich 4-5 Tagen durch die zweistufige Sequenzierung.

3.1.5.5 Einführung der Kontaminationsbestimmung bei humane Proben

Ein wichtiger Qualitätsparameter bei der Sequenzierung von exomischen und genomischen *Libraries* ist die Identität der Patientenprobe. DNAs werden im 96er-Ansatz bearbeitet und es sind viele Pipettierschritte notwendig von der DNA bis hin zur finalen *Library*. Kontaminationen führen zu einer reduzierten Genotyp-Qualität und zu falsch positiven Assoziationen [58]. Zur Bestimmung des Grades der Kontamination einer humanen Proben wurde die Methodik nach Jun *et al.* [58] eingeführt. Zusammenfassend werden dabei die Allel-Frequenzen ($MAF > 5\%$) einer Kontrollpopulation (hier die der HapMap-CEU-Population) mit dem Genotyp der sequenzierten Probe verglichen. Basierend auf der Wahrscheinlichkeit des Auftretens von SNPs kann über ein statistisches Verfahren ein Kontaminationsgrad berechnet werden.

Die Kontaminationswerte können innerhalb der Analyseebene abgerufen werden. Proben mit einem Kontaminationsgrad größer als 5 % werden von weiteren Analysen exkludiert, neue DNA oder Blut angefordert und die *Library*-Präparation und Sequenzierung wiederholt. Diese Art Anwendung der Kontaminationsbestimmung ist die Naheliegendste und eine der Wichtigsten, zur Aufrechterhaltung der Datenqualität und Vergleichbarkeit.

Es zeigte sich in der Vergangenheit zusätzlich, dass Vergleiche der Kontaminationen über einen Zeitverlauf auf Probleme bei der Prozessierung der Proben hinweisen können. So wurde zu Beginn des Jahres 2016 eine Erhöhung der Kon-

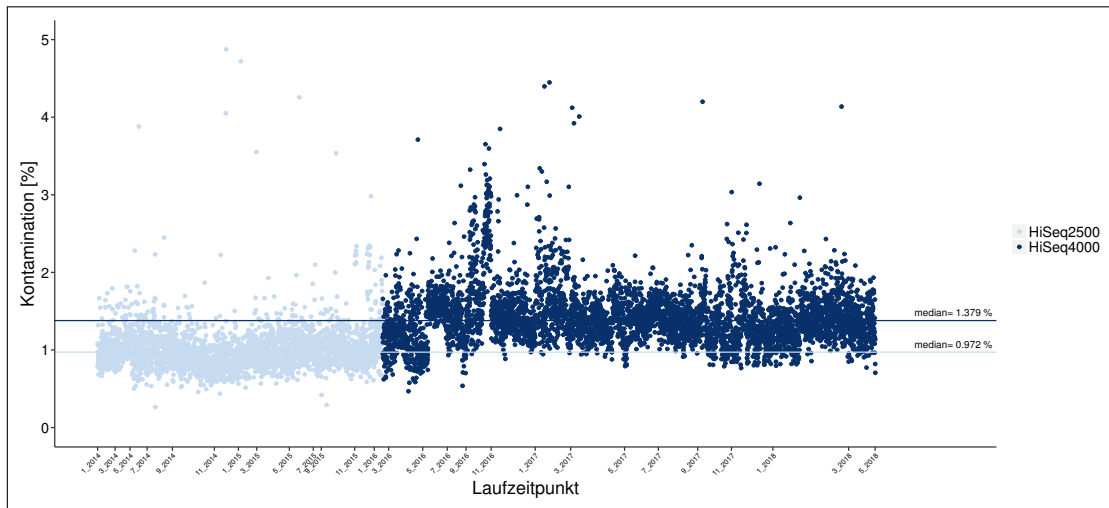


Abbildung 20: Zeitlicher Verlauf der prozentualen Kontamination von Exom-*Libraries*. Die Grafik zeigt die Kontaminationen der einzelnen *Libraries* (als Punkt dargestellt), nach der Sequenzierung auf dem HiSeq2500 (hellblau) und dem HiSeq4000 (dunkelblau). Die waagerechten Striche kennzeichnen den Median der prozentualen Kontamination bei der Sequenzierung auf den beiden unterschiedlichen Plattformen.

tamination von Exom-*Libraries* festgestellt. Eine Evaluation der Labor- und Prozessabläufe der *Library*-Präparation zeigte keine eingeführten Veränderungen. Es konnte keine Kontamination von humaner DNA in den genutzten Lösungen wie Wasser oder Elutionspuffer nachgewiesen werden (*KAPA hgDNA Quantification and QC Kit*). Eine zusätzliche Reinigung aller im Labor verwendeten Oberflächen, Geräte und Pipetten mit einer Dekontaminationslösung (DNA-ExitusPlus IF) brachte ebenfalls keine Reduktion der Kontamination.

Vergleichende Analysen über Kontaminationsplots zeigten, dass mit der Einführung der HiSeq4000-Sequenzierplattform und der damit einhergehenden *Patterned-Flow Cell*-Technologie, die Kontamination der Exom-Proben von rund 1 % auf rund 1.4 % (Median) (Abb.: 20 und 21) stieg.

Was zu einer Erhöhung der Kontaminationsrate mit der Nutzung der HiSeq4000 führte, war zu diesem Zeitpunkt gänzlich unklar. Kurze Zeit später veröffentlichte eine Forschergruppe aus Stanford eine Publikation [19], die den Effekt des *Adapter Hoppings* oder auch *Index Switchings* beschreibt, welcher zu dem registrierten Anstieg des Kontaminationsgrades führt (Abb.: 22).

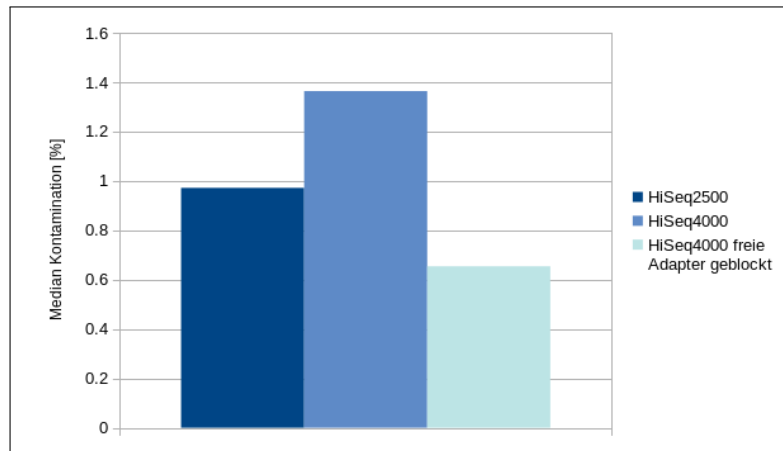


Abbildung 21: Kontamination exomischer *Libraries* auf dem HiSeq2500 im Vergleich zum HiSeq4000. Vergleichend ist der Median der Kontamination in % von Exom-*Libraries* auf dem HiSeq2500 und dem HiSeq4000 dargestellt. Zusätzlich dazu findet sich der Median der Kontamination von Exom-*Libraries* auf dem HiSeq4000, die zuvor mit einem *Blocking*-Reagenz behandelt wurden.

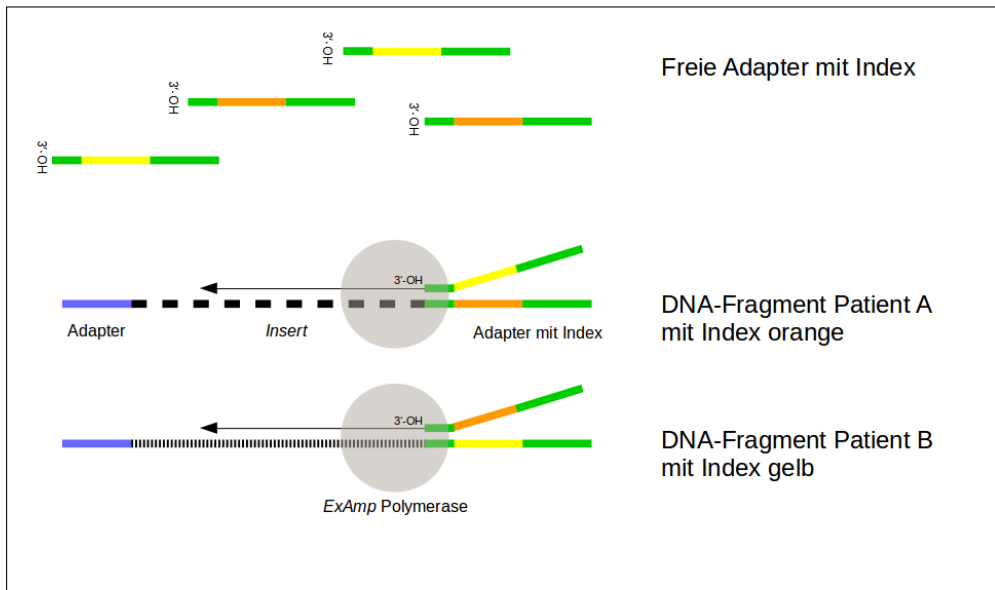


Abbildung 22: Schematische Darstellung des *Adapter Hoppings*. Bei der Herstellung von *Libraries* verbleiben freie Adapter in der Lösung. Werden *Libraries* zu einem *Pool* gemischt, befindet sich auch eine Mischung freier Adapter in der Lösung. Diese werden durch die *ExAmp*-Polymerase (grauer Kreis) als Primer für die Synthese des Zweitstrangs genutzt, wobei es zum Austausch der Indizes kommt. In dem gezeigten Beispiel sind die *Library*-Fragmente von Patient A mit einem orangenen Index und Patient B mit einem gelben Index dargestellt. Werden diese beiden Proben gemischt, befinden sich ebenfalls von beiden *Library*-Präparationen nicht aufgebrauchte Indizes in der Lösung. So kann der Adapter von Patient B als Primer-Oligo an das Fragment von Patient A gebunden und elongiert werden. Das Insert von Patient A bekommt nun den Index von Patient B und es entsteht eine Kontamination.

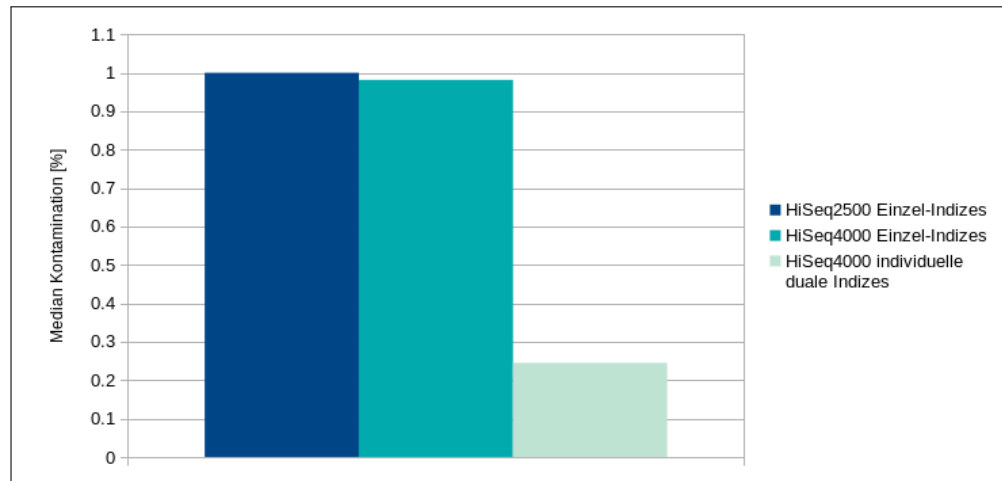


Abbildung 23: Kontamination genomischer *Libraries*. Vergleichend sind die prozentualen Kontaminationswerte (Median) von genomischen *Libraries* nach der Sequenzierung auf dem HiSeq2500 und dem HiSeq4000 dargestellt. Zusätzlich findet sich der Kontaminationsgrad bei *Libraries*, die über zwei individuelle Indizes markiert wurden.

Ein Anstieg der Kontaminationen war bei genomischen *Libraries* nicht sichtbar (Abb.: 23). Das kann daran liegen, dass zum Einen der Durchsatz genomischer *Libraries* um ein vielfaches geringer ist, als der von exomischen *Libraries* und somit der Effekt nicht sichtbar wurde. Zum Anderen kann es auch an der Anzahl der *Libraries* in einem *Pool* liegen, die bei Genom-Sequenzierungen meist im einstelligen Bereich liegt im Vergleich zu Exomen, bei denen 40 Proben *gepoolt* werden.

Für die Genom-Sequenzierung war es möglich individuelle duale Indizes zu verwenden [19, 111]. Dazu wird sowohl am 5'-Ende als auch am 3'-Ende einer *Library* ein individueller Index eingeführt. Diese beiden Indizes kommen im gesamten *Library-Pool* nicht noch einmal vor. Da die, durch *Adapter Hopping* entstandenen, Index-Neukombinationen nicht zu einem Patienten definiert wurden, werden die betroffenen Fragmente bioinformatisch herausgefiltert. Dies führte zu einer Reduktion der Kontaminationsrate bei Genomen auf 0.24 % (Median) (Abb.: 23).

Für das derzeitige Standard-Protokoll zur Herstellung von exomischen *Libraries* sind keine individuellen dualen Indizes verfügbar. Jedoch können durch die Behandlung des *Library-Pools* mit dem *Illumina Free Adapter Blocking Reagent* vor der eigentlichen Sequenzierung, die freien Adapter blockiert werden (Abb.: 24). Die Anwendung des Reagenz bei Exom-*Libraries* führte zu einer Reduktion

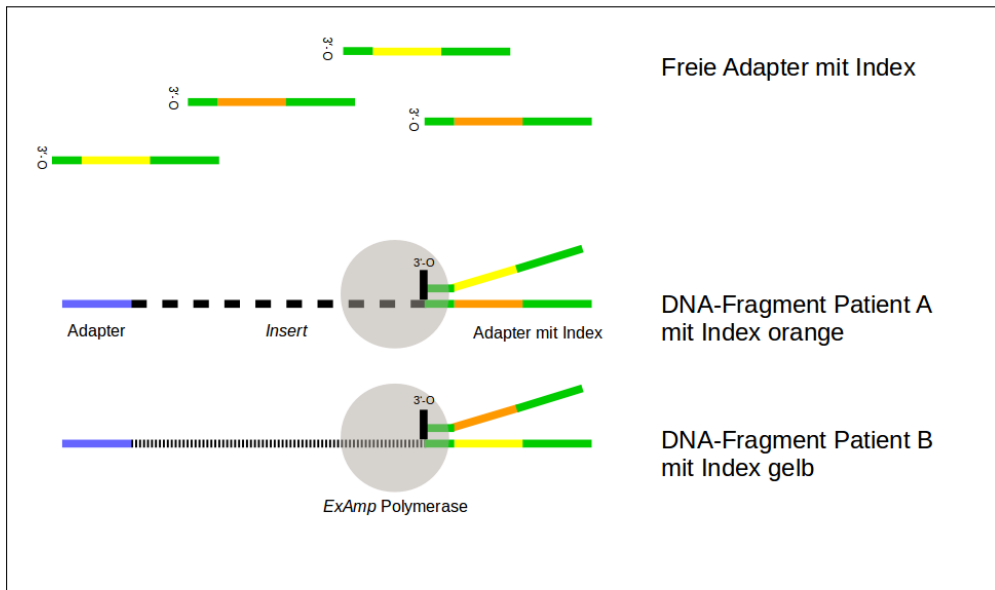


Abbildung 24: Schematische Darstellung des Adapter-*Blocking*. Durch die Behandlung mit einem *Blocking*-Reagenz werden die 3'OH-Enden der freien Adapter in ein 3'O-Enden umgewandelt. Dadurch stehen sie der Polymerase nicht mehr zur Elongation zur Verfügung.

der Kontamination von 1,4 % auf 0,7 % (Median) (Abb.: 21).

Sowohl die Verwendung von individuellen dualen Indizes bei Genomen, als auch das Blockieren freier Adapter führte zu einer deutlichen Reduktion der Kontaminationsrate sogar unter die Werte, die bei der Sequenzierung mit dem HiSeq2500 erreicht wurden. Dies kann darauf hindeuten, dass es auch bei der *Non-Patterned-Flow Cell*-Technologie zum Phänomen des *Adapter Hoppings* kommt [19, 111].

3.2 Etablierung der automatisierten Exom-*Library*-Präparation

Der folgende Abschnitt beschreibt die Einführung und Weiterentwicklung der Exom-Sequenzierung am Institut für Humangenetik. Der Prozess gliedert sich in drei Schritte: (a) die Einführung der manuellen Exom-*Library*-Präparation im niedrigen Durchsatz, (b) die Entwicklung einer Semi-Automation zur parallelen Herstellung von 48 Exom-*Libraries* und (c) der Erweiterung der Semi-Automation zur weitgehenden Automation und der damit einhergehende Hochdurchsatz der *Library*-Herstellung (Abb.: 45). Neben der Entwicklung einer Automation beschreibt der Abschnitt die Optimierung der Laborabläufe sowie der Qualitätssicherung.

3.2.1 Einführung der manuellen Exom-*Library*-Präparation

Im Jahr 2009 wurden die ersten Exom-*Libraries* hergestellt. Die manuelle Herstellung erfolgte nach dem Protokoll *SureSelect Human All Exon Kit Illumina Paired-End Sequencing Library Prep Version 1.0.1, October 2009*. Die händische Präparation war ein zeitaufwendiger und anfälliger Prozess. Von der DNA bis hin zur fertigen *Library* waren etliche Pipettierschritte notwendig, die gewissenhaft und konzentriert ausgeführt werden mussten. Die einzelnen Schritte der manuellen *Library*-Herstellung sind in Abbildung 25A dargestellt. Einer der zeit- und arbeitsintensivsten Schritte ist die säulchenbasierten Aufreinigung über das *QIAquick PCR Purification Kit* bzw. *Qiagen MinElute Kit* (insgesamt 5 mal im Prozess). Diese finden nach jedem enzymatischen Schritt statt und sind notwendig um die DNA-Lösung von Enzymen und Puffern zu reinigen und so für den nächsten Schritt vorzubereiten. Limitierend ist das Arbeiten mit der Einkanalpipette und die Zentrifugationsschritte während des Prozesses. Gleichzeitig führen die säulchenbasierten Aufreinigungsschritte zu Verlusten von Material [112]. Durch die Prozessierung mehrerer Proben in einzelnen Reaktionsgefäßen kann es zu Vertauschungen von Probematerial kommen. Über mehrere Tage konnten bei hohem Personalaufwand bis zu 12 *Libraries* hergestellt werden.

Bis zur Einführung der Semi-Automation Ende 2010 wurden 121 Exom-*Libraries* hergestellt und sequenziert. Das Ziel, zur Steigerung des Durchsatzes, war

die Etablierung einer automatisierten *Library*-Präparation.

3.2.2 Einführung der semi-automatisierten Exom-*Library*-Präparation

Mit der Einführung von Exom-Anreicherungstechniken im Jahr 2009 stieg die Anfrage nach *NG*-Sequenzierungen stark an. Im Vergleich zur Genom-Sequenzierung, ist die Exom-Sequenzierung ein kostengünstiger und erfolgversprechender Ansatz zur Identifikation pathogener Varianten in kodierenden Bereichen des Genoms [23, 113, 114, 115, 78]. Um der steigenden Nachfrage nachkommen zu können wurde eine Semi-Automation der *Library*-Präparation entwickelt.

Die Grundlage für eine Semi-Automation bot das optimierte manuelle *Library*-Präparations-Protokoll *SureSelect Target Enrichment System for Illumina Paired-End Sequencing Library, Version 2.0.1, May 2010* (Abb.: 25 B). Darin wurden die säulchenbasierten Aufreinigungsschritte, so wie sie zur Einführung der Exom-*Library*-Präparation genutzt wurden, durch eine Aufreinigung mit AMPure XP *Beads* ersetzt. Zusammen mit der Anschaffung der *Bravo Automated Liquid Handling Platform* (Agilent) sollte ein Automationsprotokoll, welches bis dahin nicht verfügbar war, entwickelt werden. Die Entwicklungsarbeit übernahm ich gemeinsam mit Dr. Ingo Poleschak (Agilent Technologies, Inc.), der die gewünschten Programmierungen durchführte.

Im Ergebnis wurden die AMPure XP *Bead*-Aufreinigung, der Schritt der Exom-Anreicherung sowie die Elution nach der Anreicherung automatisiert. Die Automation sollte so gestaltet sein, das nach dem Bestücken des Bravo mit den benötigten Verbrauchsmaterialien, die jeweiligen Protokolle ohne Zutun des Bearbeiters durchlaufen konnten. Das wurde durch ein Stapeln der Reagenzienplatten erreicht. Der Durchsatz wurde um ein 4-faches im Vergleich zur manuellen *Library*-Präparation gesteigert. Es konnten 48 Proben parallel bearbeitet werden, bei geringerer Bearbeitungszeit durch den Anwender. Die Qualität der erhaltenen Sequenzier-Daten verbesserte sich im Hinblick auf die durchschnittliche Abdeckung des Exoms (Abb.: 26) sowie der *On-Bait*-Rate (Abb.: 27), während sich die Duplikatrate der einzelnen *Libraries* verdoppelte (Abb.: 28). In den folgenden Abschnitten werden die entwickelten Automationsprotokolle beschrieben.

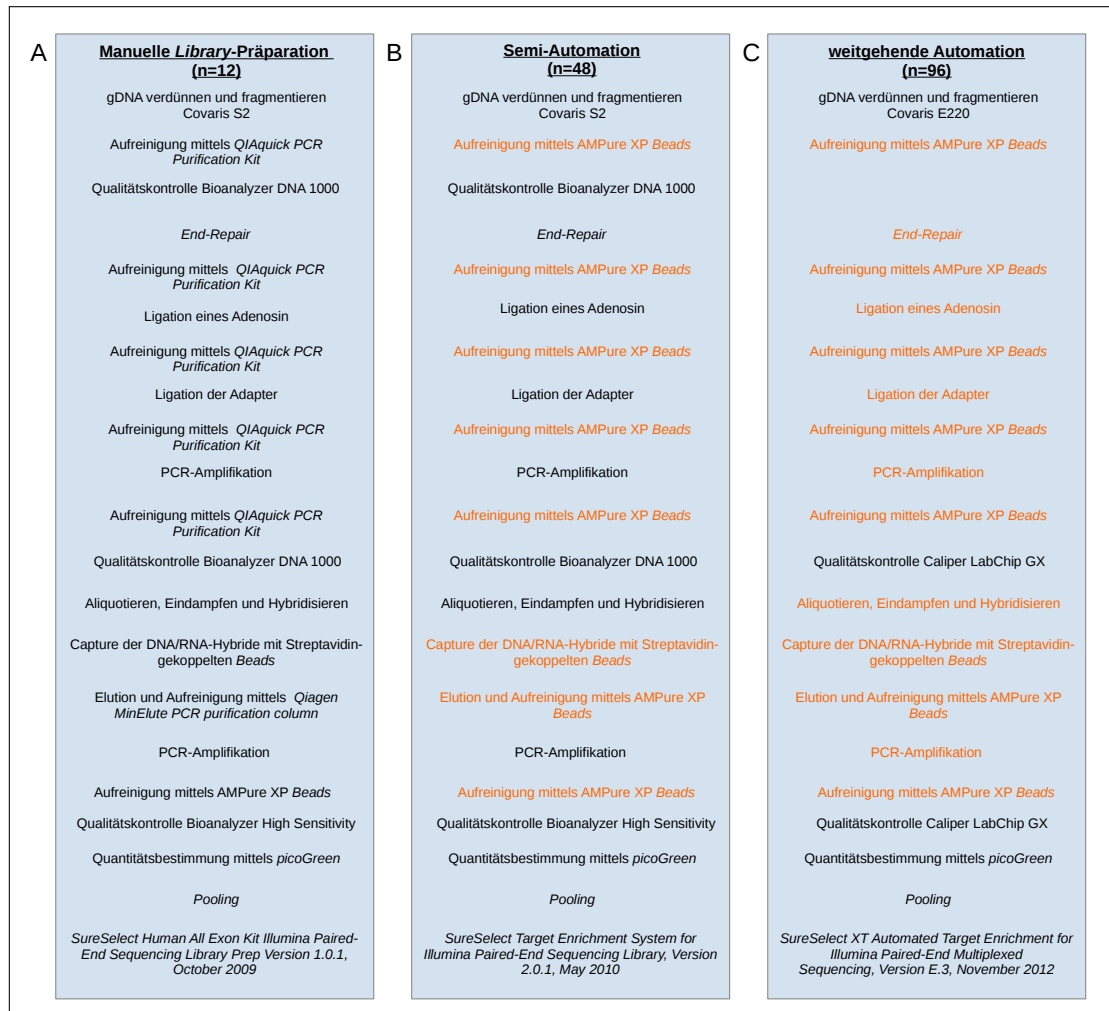


Abbildung 25: Übersicht der Arbeitsschritte einer *Library*-Präparation im manuellen, semi-automatisierten und weitgehend automatisierten Modus. In schwarz sind alle nicht-automatisierten Schritte dargestellt. In orange finden sich alle auf der *Bravo Liquid Handling Platform* automatisierten Schritte. Innerhalb der Semi-Automation werden alle AMPure XP *Bead* Aufreinigungsschritte, sowie die Exom-Anreicherung und Elution der *Libraries* automatisiert durchgeführt. Innerhalb der weitgehenden Automation finden nur noch die Fragmentierung sowie die QC-Schritte auf anderen Plattformen statt.

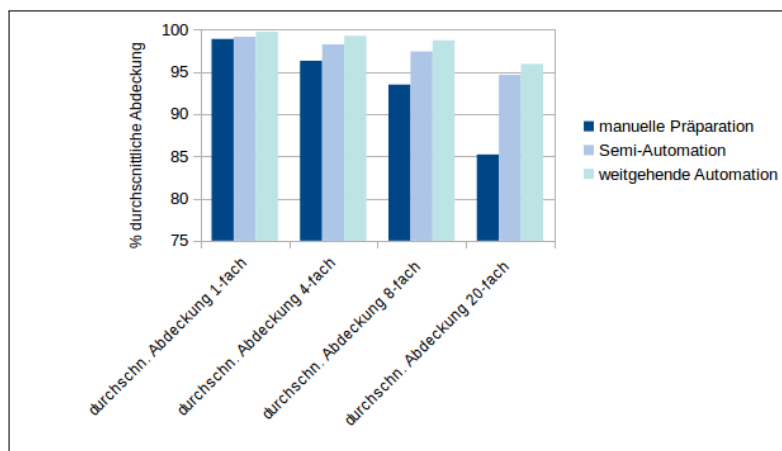


Abbildung 26: Vergleichende graphische Darstellung der durchschnittlichen Abdeckung eines Exons. Mit dem manuellen Protokoll konnte eine 20-fache Abdeckung für 85 % der Basen erzeugt werden. Die Semi-Automation steigerte diesen Wert auf 94 % und die weitgehende Automation erreicht einen Wert von 96 %.

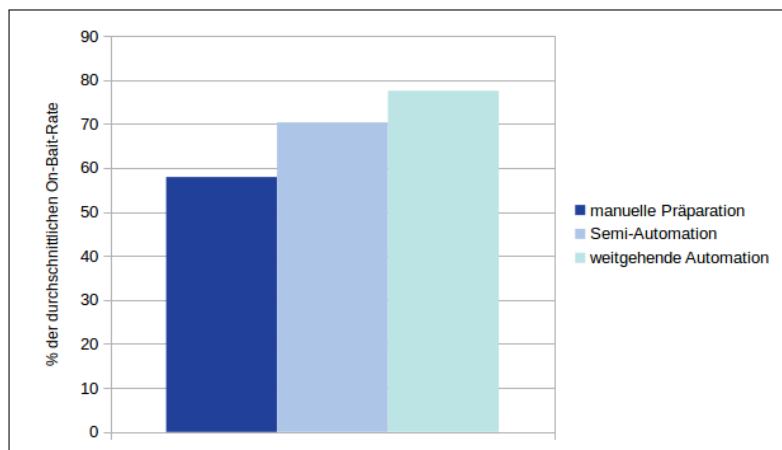


Abbildung 27: Übersicht zur *On-Bait-Rate*. Vergleichend sind die prozentualen *On-Bait-Raten* in Abhängigkeit von der Präparationsmethode der *Libraries* dargestellt.

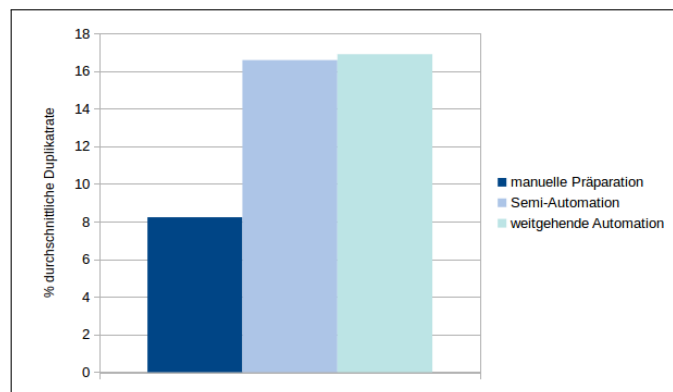


Abbildung 28: Die Duplikatrate im Überblick. Vergleichend sind die durchschnittlichen Duplikatraten der exomischen *Libraries* dargestellt. Während die manuelle *Library*-Präparation nur wenige Duplikate erzeugt, verdoppelt sich die Anzahl auf 16 % bei einer automatisierten Herstellung.

3.2.2.1 Automatisierte AMPure XP *Bead*-basierte Aufreinigung

Mit der Einführung der AMPure XP *Beads* (Beckman Coulter, Inc.) zur Aufreinigung von enzymatischen Reaktionen entfallen alle notwendigen Zentrifugationsschritte [112, 116]. Es kann mit Mehrkanalpipetten gearbeitet werden und der Verlust von DNA, durch den Wegfall der Filtration über Kieselgelmembranen (*silica gel*), wird gemindert. Bisher wurden AMPure XP *Beads* lediglich zur Reinigung des PCR-Ansatzes innerhalb der *Library*-Präparation eingesetzt. Sie arbeiten nach dem *Solid Phase Reversible Immobilization Prinzip* (SPRI) [117]. Es sind magnetische Partikel, die auf ihrer Oberfläche freie Carboxygruppen (-COOH) tragen. Sie erlauben eine reversible Bindung von DNA in Anwesenheit von Polyethylenglycol (PEG) und Natriumchlorid (NaCl), die bei Zugabe von Wasser oder Elutionspuffer (Niedrigsalzpuffer) gelöst werden kann [117]. Die Anwendung von AMPure XP *Beads* ermöglichte alle DNA-Aufreinigungsschritte innerhalb der exomischen *Library*-Herstellung zu automatisieren. Die Aufreinigung besteht aus nur wenigen Schritten, die einfach zu automatisieren sind (Abb.: 29, 1-6).

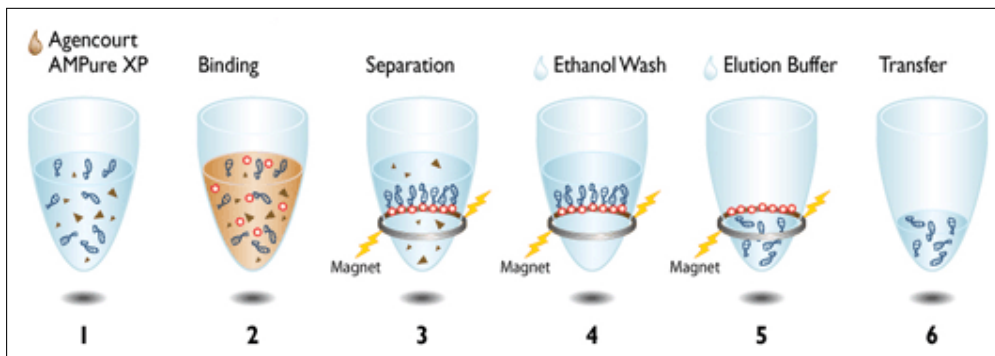


Abbildung 29: Die Schritte der AMPure XP *Bead* Aufreinigung [118]. (1) Zugabe eines definierten AMPure XP *Bead* Volumens zur DNA-Lösung. (2) Mischen der Lösung. Die DNA-Fragmente binden an die *Beads*. (3) Separation der *Bead*/DNA-Hybride über einen Magneten. Der Überstand enthält nicht gebundene DNA-Fragmente und Reste der vorangegangenen enzymatischen Reaktion. (4) 2-maliges Waschen des Pellets mit 80 % Ethanol. Die Platte verbleibt dabei auf dem Magnetständer. (5) Elution der DNA-Fragmente durch Zugabe von Elutionspuffer nachdem der Ethanol abgetrocknet ist. Die Platte wird hierzu vom Magnetständer genommen. (6) Die Platte wird auf den Magnetständer gestellt. Nach Bildung des *Bead*-Pellets wird der Überstand, in dem nun die DNA-Fragmente gelöst sind, abgenommen und in eine frische Platte pipettiert.

Für eine automatisierte AMPure XP *Bead*-Aufreinigung, standen uns sieben der neun Plattenpositionen auf dem Bravo zur Verfügung (Abb.: 30). Die Positio-

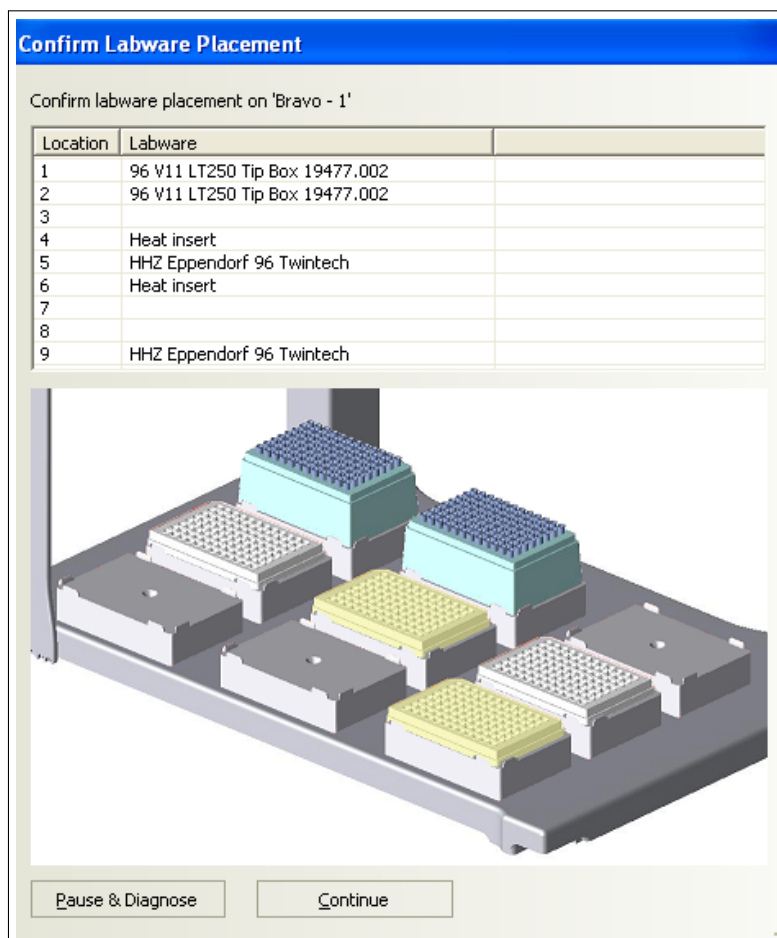


Abbildung 30: Plattenanordnung zur AMPure XP *Bead*-Aufreinigung. Die Deckpositionen werden von links oben nach rechts unten von eins bis neun durchnummeriert.

nen vier und sechs können gekühlt oder beheizt werden und sind dauerhaft mit Adaptern für *Deep Well*-Platten belegt. Position sieben wird dauerhaft durch die Magnetplatte belegt.

Um alle notwendigen Reagenzienplatten zu positionieren wurden sie auf Position neun der Reihenfolge ihrer Nutzung nach gestapelt (Abb.: 31). Position eins wird durch eine leere Spitzenbox belegt, Position 2 durch eine volle Spitzenbox. Auf Position fünf stehen die vorpipettierten AMPure XP *Beads*. Die benötigten Platten werden zur Nutzung jeweils von Position neun auf Position acht gestellt. Der Plattenstapel wird im Anschluss der Nutzung auf Position drei umgestapelt. Entgegen dem manuellen Protokoll wird innerhalb der Automation statt in 1 ml Tubes mit PCR-Platten gearbeitet, die maximal 200 μ l fassen können. Die Volumina der AMPure XP *Beads* und des Ethanol für den Waschschrift wurden



Abbildung 31: Aufbau der Bravo-Oberfläche zur AMPure XP *Bead*-Aufreinigung. Eine Abfrage und Kontrolle der Positionen erfolgt vor dem Start des Protokolls durch die Software.

aufgrund dessen in Anlehnung an Fisher *et al.* nach unten skaliert [112].

Die Aufreinigungsschritte über AMPure XP *Beads* finden nach jedem enzymatischen Schritt innerhalb der Exom-*Library*-Präparation statt. Dabei sind abhängig vom Volumen der DNA-Lösung sowie dem Elutionsvolumen unterschiedliche Parameter zu setzen. Um das Automationsprotokoll für die Aufreinigung flexibel zu gestalten, erfolgt zu Beginn eine Abfrage über die anzuwendenden Parameter (Abb.: 32). Das Protokoll hat eine Dauer von 50 Minuten für bis zu 48 Proben. Im Anschluss an die Reinigung erfolgt der nächste enzymatische Schritt. Hierzu bedarf es lediglich der Herstellung des enzymatischen Mastermixes und dessen Zugabe per Mehrkanalpipette zu den Proben. Nach vorgegebener Inkubation erfolgt der nächste Reinigungsschritt auf der Automationsplattform.

Please enter parameters for the purification process

Variable Name	Initial Value
dna_to_bead_ratio	1.8
dna_vol	50
elute_vol	40
samples	16

OK

Abbildung 32: Abfrage der Parameter zur automatisierten AMPure XP *Bead* Aufreinigung. Die DNA- und Elutions-Volumina, das Verhältnis von DNA und AMPure XP *Beads*, sowie die Probenanzahl werden an dieser Stelle definiert.

3.2.2.1.1 Prozessschritte der AMPure XP *Bead*-Aufreinigung auf dem Bravo

Im Folgenden ist der Ablauf der einzelnen Schritte, die auf dem Bravo während der Aufreinigung ausgeführt werden, dargestellt. Diese gliedern sich in: (1) Bindung der DNA-Fragmente, (2) Entfernen des Überstandes und waschen des *Bead*-Pellets sowie (3) der Elution.

1. Bindung der DNA-Fragmente an die AMPure XP *Beads*
 - (a) Umstellen der DNA-Platte von Position 9 auf Position 8.
 - (b) Aufnahme frischer Spitzen von Position 2.
 - (c) Aufnahme des Volumens aus der DNA-Platte von Position 8 und Abgabe des Volumens in die AMPure XP *Bead* Platte auf Position 5.
 - (d) Durchmischen der DNA-*Bead* Suspension mit 15 Mixzyklen auf Position 5.
 - (e) Abgabe der gebrauchten Spitzen auf Position 1.
 - (f) 5 minütige Inkubation des DNA-*Bead* Gemisches auf Position 5 zur Bindung der DNA-Fragmente an die *Beads*.
 - (g) Umstellen der DNA-*Bead*-Platte von Position 5 auf Position 7.
 - (h) 5 minütige Inkubation auf Position 7 zur Bildung des *Bead*-Pellets auf dem Magnetständer.
2. Entfernen des Überstandes und waschen des DNA-*Bead*-Pellets mit Ethanol
 - (a) Aufnahme frischer Spitzen von Position 2.
 - (b) Aufnahme des Überstandes aus der DNA-*Bead*-Platte von Position 7 und Abgabe des Überstandes in die DNA-Platte auf Position 8.
 - (c) Abgabe der gebrauchten Spitzen auf Position 1.
 - (d) Umstellen der DNA-Platte von Position 8 auf Position 3.
 - (e) Ethanol Platte 1 von Position 9 auf Position 8 stellen.
 - (f) Aufnahme frischer Spitzen von Position 2.

- (g) Aufnahme von 100 μ l 70 % Ethanol und Abgabe des Volumens in die DNA-*Bead*-Platte auf Position 7.
- (h) 1 Minute Inkubation.
- (i) Aufnahme von 105 μ l 70 % Ethanol aus der DNA-*Bead*-Platte auf Position 7 und Abgabe des Volumens in die Ethanol Platte 1 auf Position 8.
- (j) Abgabe der gebrauchten Spitzen auf Position 1.
- (k) Ethanol Platte 1 von Position 8 auf Position 3 stellen.
- (l) Ethanol Platte 2 von Position 9 auf Position 8 stellen.
- (m) Aufnahme frischer Spitzen von Position 2.
- (n) Aufnahme von 100 μ l 70 % Ethanol und Abgabe des Volumens in die DNA-*Bead*-Platte auf Position 7.
- (o) 1 Minute Inkubation.
- (p) Aufnahme von 105 μ l 70 % Ethanol aus der DNA-*Bead*-Platte auf Position 7 und Abgabe des Volumens in die Ethanol-Platte 2 auf Position 8.
- (q) Abgabe der gebrauchten Spitzen auf Position 1.
- (r) Ethanol Platte 2 von Position 8 auf Position 3 stellen.

3. Elution der DNA-Fragmente

- (a) Umstellen der DNA-*Bead*-Platte auf Position 5.
- (b) Vier minütige Inkubation zum Abtrocknen des Ethanols.
- (c) Elutionspuffer-Platte von Position 9 auf Position 8 stellen.
- (d) Aufnahme frischer Spitzen von Position 2.
- (e) Aufnahme des Elutionsvolumens aus der Elutionspuffer-Platte auf Position 8 und Abgabe des Elutionsvolumens in die Platte auf Position 5.
- (f) Durchmischen der Suspension mit 30 Mix-Zyklen.
- (g) Abgabe der gebrauchten Spitzen auf Position 1.
- (h) Zwei minütige Inkubation auf Position 5.

- (i) Umstellen der DNA-*Bead*-Platte von Position 5 auf Position 7.
- (j) Drei minütige Inkubation auf Position 7 zur Bildung des Pellets.
- (k) Aufnahme frischer Spitzen von Position 2.
- (l) Aufnahme des Elutionsvolumens aus der DNA-*Bead*-Platte auf Position 7 und Abgabe des Elutionsvolumens in die Elutionsplatte auf Position 9.
- (m) Abgabe der gebrauchten Spitzen auf Position 1.
- (n) Umstellen der Elutionspuffer-Platte von Position 8 auf Position 3.
- (o) Umstellen der leeren DNA-*Bead*-Platte von Position 7 auf Position 5.
- (p) Die eluierte DNA ist auf Position 9 bereit für den nächsten Schritt.

3.2.2.1.2 Vergleich der manuellen und automatisierten AMPure XP *Bead*-Aufreinigung

Die Ausführung der automatisierten AMPure XP *Bead*-Aufreinigung kann direkt gegen die Ergebnisse der manuellen Aufreinigung verglichen werden. Dazu wurden von vier DNAs 6 µg Material in 260 µl Wasser verdünnt, der Ansatz geteilt und über den Covaris E220 nach Standard-Protokoll fragmentiert. Die eine Hälfte der Proben wurde automatisiert über AMPure XP *Beads* aufgereinigt, während die zweite Hälfte nach dem manuellen Protokoll bearbeitet wurde. Im Anschluss an die Aufreinigung wurden die Proben mittels Bioanalyzer (DNA 1000 Kit) quantifiziert. Alle automatisiert aufgereinigten Proben zeigen dabei eine geringere Ausbeute (5-29 %) nach der Reinigung im Vergleich zur gepaarten Probe, die der manuellen Bearbeitung unterzogen wurde (Abb.: 33). Der Verlust von Material bei der automatisierten Aufreinigung führt zu einer Verminderung der Komplexität der *Library*, was sich in einem Anstieg der Duplikatrate zeigt (Abb.: 18). Aufgrund der angestrebten Sequenziertiefe von 8 Gb (120-fache Abdeckung des Exoms) wird dieser Anstieg ausgeglichen und eine ausreichende 20-fache Abdeckung von durchschnittlich 95 % erreicht (Abb.: 26).

Ursächlich für den Materialverlust während der automatisierten AMPure XP *Bead*-Aufreinigung ist die nicht vorhandene Anpassungsfähigkeit der Automationsplattform. Eine händische Bearbeitung hat im Vergleich dazu den Vorteil

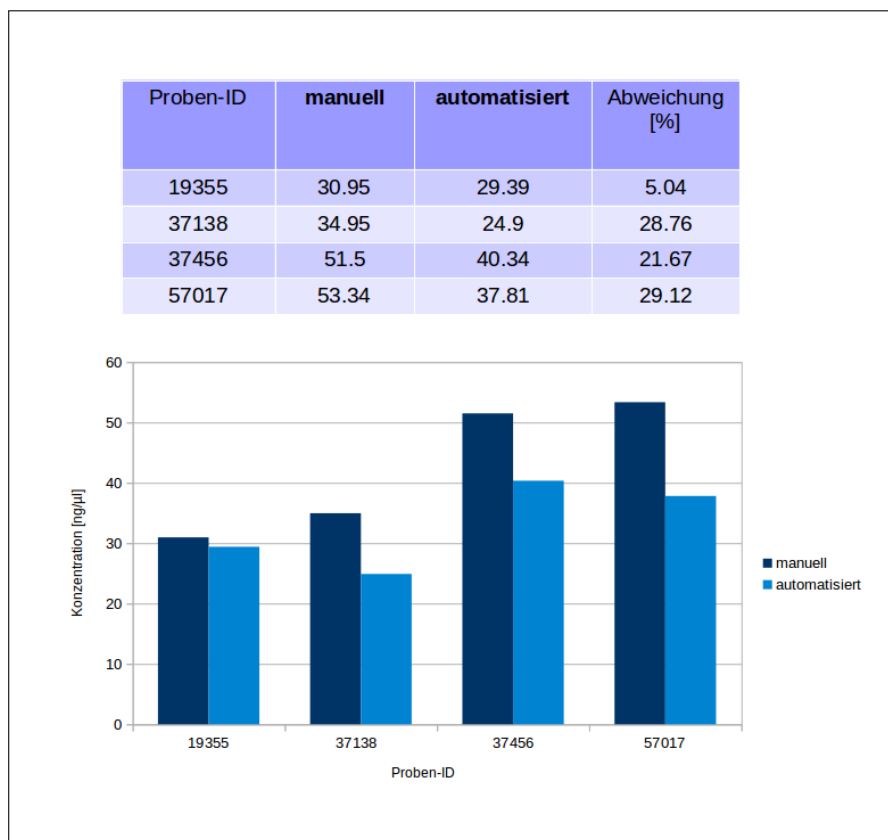


Abbildung 33: Vergleichende Darstellung der finalen Konzentrationen nach automatisierter versus manueller AMPure XP *Bead*-Aufreinigung.

flexibel auf den Prozess eingehen zu können. So kann zum Beispiel die Trocknungszeit der *Beads* durch visuelle Inspektion genau angepasst werden. Diese ist abhängig von der eingesetzten Menge an AMPure XP *Beads* und der Raumtemperatur. Werden die *Beads* Übertrocknet, ist keine effiziente Elution mehr möglich. Die Automation hingegen kann nur anhand der Menge der eingesetzten *Beads* programmiert werden und kann nicht flexibel auf die Raumtemperatur eingehen. Dadurch kommt es zu starren Trocknungszeiten, die abhängig von den äußeren Faktoren suboptimal sein können.

3.2.2.2 Automation der Exom-Anreicherung und der Elution

Die Anreicherung exomischer Bereiche einer DNA-*Library* erfolgt in Lösung über eine komplementäre Bindung an RNA-*Baits* [28]. Diese sind biotinyliert und werden im Anschluss an die Hybridisierungsreaktion an magnetische Streptavidin-*Beads* gebunden, während der Überstand weggewaschen wird. Die Hybride werden eluiert und mittels PCR amplifiziert (Abb.: 4). Innerhalb des Prozesses wird vier mal mit zwei verschiedenen Lösungen gewaschen. Der händische Durchsatz liegt bei 12 Proben pro vier Stunden. Da in einzelnen Reaktionsgefäßen gearbeitet wird, ist die Gefahr von Pipettierfehlern dieser Stelle besonders groß. Die Automation ermöglicht die Anreicherung von 48 Proben in vier Stunden. Die Automation inkludiert das einmalige Waschen mit Waschpuffer 1 und das dreimalige Waschen mit Waschpuffer 2. Die Volumina der Waschpuffer wurden nach Fisher *et al.* [112] angepasst, um in *Deep Well*-Platten arbeiten zu können. Die Anordnung der Platten auf der Automationsoberfläche wird bebildert und schriftlich vor dem Start des Protokolls abgefragt und kontrolliert (Abb.: 34). Die Elution erfordert ein erneutes Bestücken des Bravo mit Reagenzien und Platten.

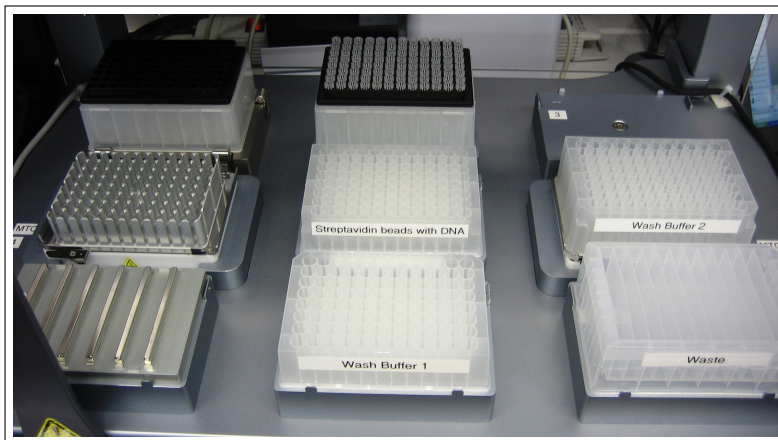


Abbildung 34: Anordnung der Platten vor dem Start des Anreicherungs-Protokolls. Die Positionen 4 und 6 sind auf 65 °C vorgeheizt.

3.2.2.2.1 Automatisierte Prozessschritte der Exom-Anreicherung und Elution auf dem Bravo

Im Folgenden ist der Ablauf der einzelnen Schritte, die auf dem Bravo während der Exom-Anreicherung und Elution ausgeführt werden, dargestellt. Der Prozess gliedert sich in drei Schritte: (1) Bindung der DNA/RNA-Hybride an Streptavidin-*Beads*. Über die biotynilierten RNA-*Baits* werden die DNA/RNA-Hybride an Streptavidin-*Beads* gebunden. DNA-Fragmente, die nicht an eine komplementäre RNA gebunden sind, bleiben in Lösung. Der Prozess findet unter ständiger Bewegung der Flüssigkeit innerhalb von 30 Minuten statt. Die Bewegung der Suspension wird durch auf- und abpipettieren gewährleistet. Im Anschluss an die Bindung wird nach der Inkubation auf der Magnetplatte der Überstand mit den ungebundenen DNA-Fragmenten abgenommen und verworfen. (2) Waschen mit Waschpuffer 1 zum Entfernen von Pufferresten aus Schritt 1. Das *Bead*-Pellet wird in Waschpuffer 1 resuspendiert und 15 Minuten inkubiert. Im Anschluss wird der Waschpuffer 1 nach erneuter Bildung des *Bead*-Pellets abgenommen und verworfen. (3) Waschen mit Waschpuffer 2. Das Waschen mit Waschpuffer 2 dient zur Entfernung unspezifisch gebundener DNA-Fragmente. Dazu wird der Puffer auf 65 °C erhitzt und zum *Bead*-Pellet gegeben. Nach 10 minütiger Inkubation der *Bead*-Platte bei 65 °C wird nach der Bildung des *Bead*-Pellets der Waschpuffer 2 wieder abgenommen und verworfen. Der Waschschrift mit Waschpuffer 2 wird insgesamt drei mal wiederholt. Final erfolgt die Elution der angereicherten DNA mit Elutions- und Neutralisationspuffer.

1. Schritt: Bindung der DNA/RNA-Hybride an Streptavidin-*Beads*
 - (a) Aufnahme frischer Spitzen von Position 2.
 - (b) Mixen der Streptavidin-*Bead*-Suspension in der *Capture*-Platte mit 180 µl 200 mal.
 - (c) Abgabe der gebrauchten Spitzen auf Position 1.
 - (d) Umstellen der *Capture*-Platte von Position 5 auf Position 7.
 - (e) 5 minütige Inkubation auf Position 7 zur Bildung des *Bead*-Pellets auf dem Magnetständer.
 - (f) Aufnahme frischer Spitzen von Position 2.

- (g) Öffnen eines *Loops* zur Abnahme des Überstandes. Der *Loop* wird zwei mal durchgeführt.
 - i. Aufnahme von 115 μl Überstand aus der *Capture*-Platte und Abgabe des Volumens in die Abfall-Platte auf Position 9.
 - (h) Schließen des *Loops*.
 - (i) Abgabe der gebrauchten Spitzen auf Position 1.
2. Schritt : Waschen mit Waschpuffer 1 zum Entfernen von Pufferresten aus Schritt 1.
- (a) Umstellen der *Capture*-Platte von Position 7 auf Position 5.
 - (b) Aufnahme frischer Spitzen von Position 2.
 - (c) Öffnen eines *Loops*. Der *Loop* wird drei mal durchgeführt.
 - i. Aufnahme von 166.6 μl Waschpuffer 1 auf Position 8 und Abgabe des Volumens in die *Capture*-Platte von Position 5.
 - ii. 4 minütige Inkubation auf Position 5.
 - iii. Mixen der Suspension in der *Capture*-Platte mit 155 μl 10 mal.
 - (d) Schließen des *Loops*.
 - (e) Mixen der Suspension in der *Capture*-Platte mit 190 μl 90 mal.
 - (f) Abgabe der gebrauchten Spitzen auf Position 1.
 - (g) Umstellen der *Capture*-Platte von Position 5 auf Position 7.
 - (h) 5 minütige Inkubation auf Position 7.
 - (i) Aufnahme frischer Spitzen von Position 2.
 - (j) Öffnen eines *Loops* zur Abnahme des Überstands. Der *Loop* wird drei mal durchgeführt.
 - i. Aufnahme von 170 μl Überstand aus der *Capture*-Platte auf Position 7 und Abgabe des Volumens in die Abfall-Platte von Position 9.
 - (k) Schließen des *Loops*.
 - (l) Abgabe der gebrauchten Spitzen auf Position 1.

-
3. Schritt: Waschen mit Waschpuffer 2 zum Entfernen unspezifischer Hybridisierungsprodukte bei 65 °C.
- (a) Umstellen der *Capture*-Platte von Position 7 auf Position 5.
 - (b) Öffnen eines *Loops*. Der *Loop* wird drei mal durchgeführt.
 - i. Aufnahme frischer Spitzen von Position 2.
 - ii. Öffnen eines zweiten *Loops*. Der *Loop* wird drei mal durchgeführt.
 - A. Aufnahme von 166 µl Waschpuffer 2 auf Position 6 und Abgabe des Volumens in *Capture*-Platte auf Position 5.
 - iii. Schließen des *Loops*.
 - iv. Mischen der Suspension in der *Capture*-Platte mit 190 µl 15 mal.
 - v. Abgabe der gebrauchten Spitzen auf Position 2.
 - vi. Umstellen der Platte von Position 5 auf die vorgeheizte Position 4.
 - vii. Wiederaufnahme der Spitzen von Position 2.
 - viii. Durchmischen der Suspension in der *Capture*-Platte mit 190 µl 15 mal.
 - ix. 10 minütige Inkubation der *Capture* Platte auf Position 4.
 - x. Durchmischen der Suspension in der *Capture*-Platte mit 190 µl 15 mal.
 - xi. Abgabe der gebrauchten Spitzen auf Position 2.
 - xii. Umstellen der Platte von Position 4 auf die vorgeheizte Position 7.
 - xiii. 5 minütige Inkubation auf Position 7 zur Bildung des *Bead*-Pellets auf dem Magnetständer.
 - xiv. Wiederaufnahme der Spitzen von Position 2.
 - xv. Öffnen eines *Loops* zur Abnahme des Überstandes.
 - A. Aufnahme von 170 µl Überstand aus der *Capture*-Platte auf Position 7 und Abgabe des Volumens in die Abfall-Platte auf Position 9.
 - xvi. Schließen des *Loops*.
 - xvii. Abgabe der gebrauchten Spitzen auf Position 1.

- xviii. Umstellen der *Capture*-Platte von Position 7 auf Position 5.
 - (c) Schließen des *Loops*.
 - (d) Umstellen der *Capture*-Platte von Position 7 auf Position 3.
4. Schritt: Elution der DNA.
- (a) Aufnahme frischer Spitzen von Position 2.
 - (b) Aufnahme von 50 μ l Elutionspuffer von Position 8 und Abgabe des Volumens in die *Capture*-Platte auf Position 3.
 - (c) Durchmischen der Suspension in der *Capture*-Platte mit 30 μ l 15 mal.
 - (d) 10 minütige Inkubation der *Capture*-Platte auf Position 3.
 - (e) Durchmischen der Suspension in der *Capture*-Platte mit 30 μ l 15 mal.
 - (f) Abgabe der gebrauchten Spitzen auf Position 1.
 - (g) Umstellen der *Capture*-Platte von Position 3 auf Position 7.
 - (h) 5 minütige Inkubation auf Position 7 zur Bildung des *Bead*-Pellets auf dem Magnetständer.
 - (i) Aufnahme frischer Spitzen von Position 2.
 - (j) Aufnahme von 50 μ l Elutionsvolumen von Position 7 und Abgabe des Volumens in die Elutionsplatte auf Position 9.
 - (k) Durchmischen des Elutionsvolumens mit dem Neutralisationspuffer in der Elutionsplatte auf Position 9 mit 80 μ l 15 mal.
 - (l) Abgabe der gebrauchten Spitzen auf Position 1.

Die DNA-Fragmente werden im Anschluss an ihre Anreicherung amplifiziert. Die PCR-Reaktion wird noch einmal automatisiert über AMPure XP *Beads* aufgereinigt. Die nun fertige *Exom-Library* wird final quantifiziert, ihre Qualität beurteilt und kann dann sequenziert werden.

3.2.2.3 Anwendung der automatisierten AMPure XP *Bead*-Aufreinigung zur Herstellung von genomischen *Libraries*

Zur Präparation von Genom-*Libraries* sind die gleichen Schritte notwendig, wie zur Präparation einer Exom-*Library* bis zum Zeitpunkt vor der Hybridisierung. Die genomische DNA wird fragmentiert, ihre Enden repariert, ein Adenin am 3'-Ende gebunden und die plattformspezifischen Adapter ligiert (vgl. Abb.: 1, A-D). Nach der Fragmentierung und zwischen den enzymatischen Schritten erfolgt, wie bei der Exom-*Library*-Präparation, die Aufreinigung über AMPure XP *Beads*. Die etablierten Automationprotokolle wurden für die Genom-*Library*-Präparation übernommen. Die Anwendung erlaubt auch hier die Aufskalierung auf 48 Proben pro Durchgang und verhindert durch das Arbeiten in PCR-Platten ein mögliches Verwecheln von Proben. Die Volumina der AMPure XP *Beads* und der Elution innerhalb der automatisierten Aufreinigung wurden dem *TruSeq DNA Sample Preparation Guide* entnommen.

3.2.2.4 Kontrollpunkte während der Automation

Die Semi-Automation ermöglicht eine Aufskalierung des manuellen Durchsatzes von 12 auf 48 Proben. Es muss gewährleistet sein, dass die Protokolle ohne Fehler durchlaufen um *Libraries* mit hoher Sequenzierqualität herzustellen [112]. Die Automation selbst stoppt nur dann, wenn es Probleme mit der Hardware gibt. Steht zum Beispiel keine Platte auf der entsprechenden Position erkennt die Automation das Fehlen und gibt eine Fehlermeldung aus. Sie erkennt hingegen nicht, ob die definierte Platte an der richtigen Position steht oder ob Flüssigkeiten wie vorgeschrieben aufgesaugt werden. Hier ist eine manuelle Kontrolle durch den Anwender notwendig. Dazu wurden Kontrollpunkte eingeführt: (a) vor jedem Automationsstart wird über die Software noch einmal die Platten-Konfiguration visualisiert und abgefragt (Abb.: 31). (b) Die eingegebenen Volumen-Parameter innerhalb einer *Bead*-basierten Aufreinigung müssen am Ende noch einmal bestätigt werden (Abb.: 32). (c) Während des Automationsprotokolls kann sporadisch die Gleichmäßigkeit der pipettierten Volumina visuell kontrolliert werden. Mit diesen einfachen Maßnahmen werden Ausfälle in der *Library*-Präparation vermindert und möglichen Defekten an der Automation selbst vorgebeugt.

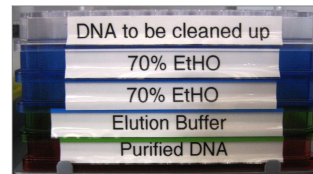
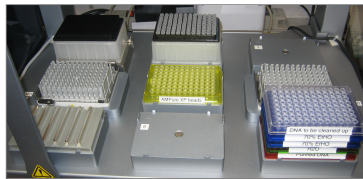
3.2.2.5 Automationsprotokolle und *Lab Tracking*-Formulare

Zur Steuerung der Prozesse und der Software innerhalb der Semi-Automation wurden Arbeitsanleitungen erstellt. Diese fassten die manuellen und automatisierten Arbeitsschritte einer *Library*-Präparationen zusammen (Abb.: 35 zeigt einen Ausschnitt und Anhang 5.1 sowie 5.2 die vollständigen Protokolle).

Die Anleitungen für eine semi-automatisierte Exom- und Genom-*Library*-Präparation, entstanden in Anlehnung an das manuelle SureSelect Protokoll *SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library*. Zur besseren Übersichtlichkeit und Qualitätskontrolle entwarf ich entsprechende *Lab Tracking*-Formulare (Abb.: 36 zeigt einen Ausschnitt sowie Anhang 5.3 und 5.4 die kompletten Formulare). Darin wird der Bearbeiter und der Bearbeitungszeitraum hinterlegt. Aufgelistet nach den Arbeitsschritten werden die einzelnen Handgriffe benannt und abgefragt. Das Formular erlaubt eine letzte Kontrolle aller Arbeitsschritte, bevor das Automationsprotokoll gestartet wird und verhindert Fehler im Ablauf. Gleichzeitig dient es der Dokumentation von abweichenden Arbeitsschritten oder Auffälligkeiten und kann bei zukünftigen Fehleranalysen aus dem Archiv herangezogen werden.

Step 3. Purify the sheared sample using Agencourt AMPure XP beads 3.0x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **150 µl of homogenous Agencourt AMPure XP** beads in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **55µl of nuclease-free water** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110µl of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing sheared DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "1. Post Shearing SPRI bead Purification protocol stacked Row.pro" protocol and follow the instructions.

12. Enter parameters for the purification step as follows:

- | | |
|----------------------|-----|
| a. DNA to bead ratio | 3.0 |
| b. DNA_vol | 52 |
| c. elute_vol | 50 |
| d. samples | 48 |



13. Once the protocol has finished assess quality with the Agilent 2100 Bioanalyzer.

SAFE STOPPING POINT (-20°C)

Abbildung 35: Ausschnitt aus dem Automationsprotokoll zur Exom-*Library*-Präparation. Die vollständigen Protokolle für die Genom- und Exom-*Library*-Herstellung finden sich im Anhang.

Exome Lab Tracking Form

DNA batch description: _____ Date: _____

PLEASE MAKE A NOTE OF THE USED THERMOCYCLER AT EACH INCUBATION STEP

1 Measurement of DNA stocks

Date: _____
 Operator: _____
 Quality and Quantity control: Start: _____ Stop: _____

NanoDrop cleaned with sodium hypochlorite
 measurement of 1µl DNA stock per sample
 verification of DNA quality and quantity
 (high molecular DNA with a minimum of 60ng/µl with a A260/280 ratio between 1.8 and 2.0 is required)

2 Fragment DNA

Date: _____
 Operator: _____
 Fragment DNA: Start: _____ Stop: _____
 covaris instrument at 7°C and degased at least for 30min
 dilution of 3µg DNA with low TE to a final volume of 50µl
 sheared with Broad settings (Broad SureSelect)
 sheared DNA transferred into an Eppendorf twin.tec plate

3 Perform automated AMPure XP Purification

Date: _____
 Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 150µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 55µl water per sample on position 9 on top of the empty plate
 two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing sheared DNA on top of the stack on position 9
 Bravo protocol 1.SPRI bead purification protocol stacked started

4 Quality Check via Bioanalyzer (DNA 1000 Kit)

Date: _____
 Operator: _____
 Bioanalyzer Start: _____ Stop: _____
 DNA 1000 Kit 30min at RT
 1µl sample per well
 verification of quality and quantity of sheared DNA
 peak from the beginning of the curve to 500bp
 average peak height 135bp
 average concentration 45ng/µl (20ng/µl minimum)
 result tables exported and sample tracking sheet for BA post shearing completed

Abbildung 36: Ausschnitt des *Lab Tracking*-Formulars für die *Exom-Library*-Präparation. Das vollständige Formular, sowie das Formular für die *Genom-Library*-Herstellung finden sich im Anhang.

3.2.3 Einführung einer weitgehenden Automation zur Herstellung von exomischen *Libraries*

Mit der Semi-Automation war es möglich 48 Proben parallel zu bearbeiten. Eine Steigerung des Durchsatzes war auf Grund der begrenzten Plattenpositionen auf dem Bravo nicht möglich. Die Erhöhung des Probendurchsatzes hätte ein deutlich häufigeres Eingreifen des Bearbeiters erfordert und den Prozess ineffektiv gemacht. So wäre es nötig gewesen, zu verschiedenen Zeitpunkten Spitzenboxen zu tauschen und Mikrotiterplatten rein oder raus zustellen. Bis zur Erweiterung des *Bravo* um den Agilent *BenchCel 4R Microplate Handler* und den Agilent *Labware MiniHub* im Jahr 2013 wurde mit den beschriebenen Exom-Automationsprotokollen gearbeitet. Die Erweiterung erlaubt es nunmehr 96 Proben parallel zu Prozessieren. Die Protokolle zu weitgehenden Automation werden von Agilent zur Verfügung gestellt und bedürfen keiner weiteren Modifikation (*SureSelect XT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing, Version E.3, November 2012*). Bis auf die initiale Verdünnung der DNA, der Qualitätskontrolle auf dem *Caliper LabChip GX* vor der Hybridisierung und am Ende der *Library*-Präparation, finden alle Schritte weitgehend automatisiert auf der *Bravo Liquid Handling Station NGS Option B* statt (Abb.: 25 C). Mit der Einführung einer weitgehenden Automation konnte die durchschnittliche 20-fache Abdeckung einer Base auf 96 % und die *On-Bait*-Rate auf 78 % erhöht werden, während die Duplikatrate annähernd gleich bleibt (Abb.: 26, 27 und 28).

3.2.4 Qualitätskontrolle und Prozessoptimierung der *Library*-Präparation

Neben der Entwicklung einer automatisierten *Library*-Präparation hatte die vorliegende Arbeit auch zum Ziel, die laborrelevanten Prozesse zu optimieren und zu standardisieren.

Der folgende Abschnitt beschreibt die Anforderungen an die DNA-Qualitätskontrolle als wichtigen Parameter zur Aufrechterhaltung der Reproduzierbarkeit sowie dem Erhalt und der Vergleichbarkeit von qualitativ hochwertigen Sequenzierdaten. Es sollte abgeleitet werden, welche Anforderungen an eingesendetes DNA-Material gestellt werden müssen, um eine erfolgreiche *Library*-Herstellung zu gewährleisten.

Im Weiteren werden Optimierungen innerhalb des Laborablaufs beschrieben, um die Arbeitszeiten weiter zu verkürzen. Die Prozessoptimierung beschreibt zudem die Entwicklung von Methodiken zum Ausschluss von Probenverwechslungen und Kontaminationen innerhalb der *Library*-Präparation.

3.2.4.1 DNA-Qualitätskontrolle

Das Gelingen der zeit- und kostenintensiven *Library*-Präparation sowie die erfolgreiche Sequenzierung einer DNA-Probe sind abhängig von der genauen Quantifizierung, sowie der Qualität des Ausgangsmaterials. Eine hohe Qualität ist zum Erreichen der angestrebten Sequenziertiefe und Sequenziergenauigkeit, sowie zum Erhalt der Reproduzierbarkeit und Vergleichbarkeit notwendig. Die qualitätsbeeinflussenden Faktoren sind die Integrität, die Reinheit sowie die Konzentration der Nukleinsäure. Die Bestimmung der einzelnen Parameter einer DNA zu Beginn einer jeder *Library*-Präparation ist ein kritischer Schritt, der standardisierte Abläufe und Methoden erfordert (Abb.: 37).

3.2.4.1.1 Bestimmung der DNA-Konzentration

Eine genaue Bestimmung der Konzentration von doppelsträngiger genomischer DNA vor der *Library*-Präparation ist essentiell um eine *Library* generieren zu können, eine homogene Sequenziertiefe zu erreichen und eine Verzerrung der Sequenzierdaten zu vermeiden. Die DNA-Konzentrationsbestimmung wurde in den ersten Monaten mit dem UV-Vis Spektralphotometer NanoDrop (Thermo Scien-

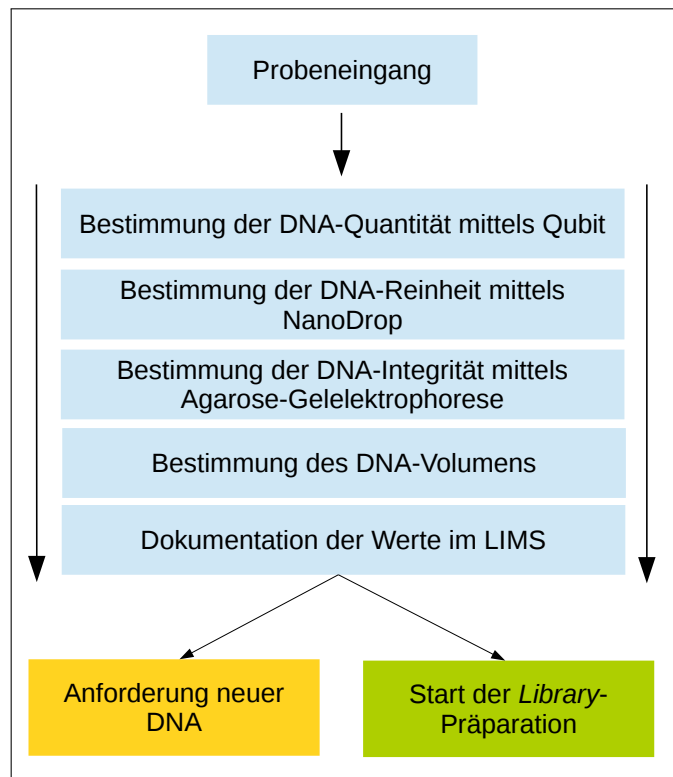


Abbildung 37: Arbeitsablauf zur Bestimmung der DNA-Qualifikation für die *NG*-Sequenzierung. Die DNA wird quantifiziert, auf ihre Reinheit und Integrität getestet und das Volumen des vorhandenen Materials bestimmt. Im Anschluss des Qualifizierungsablaufs wird entschieden ob neues Material angefordert werden muss oder die *Library*-Präparation starten kann.

tific) durchgeführt [119]. Dabei wird die Extinktion der gereinigten DNA bei 260 nm gemessen. Die Anwendung des NanoDrop zur Konzentrationsbestimmung ist schnell und erfordert keine Vorbehandlung der Probe. Nachteilig hingegen wirkt sich die Ungenauigkeit der Messung bei unreinen DNA-Extraktionen aus. Vor allem eine Überquantifizierung [120, 121, 122] der DNA führte zum Ausfall der Probe während der *Library*-Präparation oder zu einer wenig komplexen *Library* mit entsprechend heterogener Abdeckung und hoher Duplikatrate. Dabei lag der Prozentsatz der zu wiederholenden Proben bei 10 %.

Eine Überquantifizierung entsteht bei Verunreinigungen der DNA-Probe mit Proteinen, RNA und einzelsträngiger DNA, die Licht ebenfalls bei 260 nm absorbieren. Für eine genauere Bestimmung der Quantität von doppelsträngiger, genomischer DNA wurde eine fluoreszenzbasierte Messmethodik eingeführt. Die DNA wird mit einem Fluoreszenzfarbstoff gemischt und im Vergleich zu zwei

Tabelle 4: Vergleich der gemessenen Quantitäten zwischen UV-Vis- und fluoreszenzbasierten Messmethodiken an repräsentativen Beispielen.

Sample ID	NanoDrop [ng/ μ l]	Qubit [ng/ μ l]	Abweichung in %
PARIS0022G	44	15,9	63,8
PARIS125G	244	47	80,7
WIEN0971	442	254	42,5
TUEB014	417	338	18,9

Standards im Fluoreszenzphotometer Qubit (Life Technologies) gemessen. Das genutzte Fluorochrom PicoGreen (2-(n-bis-(3-dimethylaminopropyl)-amino)-4-(2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-methylidene)-1-phenyl-quinolinium) interkaliert dabei über elektrostatische Interaktionen mit dem negativ geladenen Phosphatrückgrat an doppelsträngige DNA Moleküle [123]. PicoGreen bindet nur marginal an RNA und einzelsträngige DNA-Moleküle, während Proteine gar nicht mit dem Fluorochrom interagieren [124]. Vergleichende Messungen zwischen der klassischen UV-Vis Spektralphotometrie und der fluoreszenzbasierten Technik zeigen eine abweichende Konzentration um bis zu 80 % (Tab.: 4). Nach Umstellung der Messmethodik zur Quantitätsbestimmung konnte die Ausfallrate durch zu wenig Einsatzmenge auf 0 bis 1 % gemindert werden.

3.2.4.1.2 Bestimmung der Reinheit von DNA-Proben

Zusätzlich zur Quantität, kann die Reinheit der Probe ausschlaggebend sein. DNA-Extraktionen, die mit Proteinen, Salzen oder Lösungsmittel verunreinigt sind, können die enzymatischen Reaktionen innerhalb des *Library*-Präparationsprozesses negativ beeinflussen und zum Ausfall einer Probe oder zur Komplexitätsminderung führen. So wird in den kommerziellen Protokollen zur Herstellung von exomischen *Libraries* empfohlen, die Reinheit der Probe vor dem Beginn einer *Library*-Präparation zu bestimmen.

Die Reinheit der Probe lässt sich über die Absorptionswerte bei 260 nm, 280 nm und 230 nm mit dem NanoDrop beurteilen. Das Verhältnis der Absorption bei 260 und 280 nm sollte zwischen 1,8 und 2,0 liegen. Werte unterhalb von 1,8 weisen typischerweise auf Protein-Kontaminationen hin, welche die weiteren Präparationsschritte negativ beeinflussen können. Verunreinigungen mit Lösungsmitteln, wie Phenol oder Trizol, führen zu einem verminderten A₂₆₀/A₂₃₀-Verhältnis, welches bei neutralen DNA-Extraktionen zwischen 2,0 und 2,2 liegen sollte. Ver-

Tabelle 5: Übersicht zur Sequenzier-Statistik von Exomen mit unterschiedlichen Reinheiten. Im Vergleich dazu sind die Qualitätsparameter von Exomen aus DNA, die innerhalb der Spezifikationen liegen, aufgezeigt.

	A260/280 > 3	A260/280 <1	A260/230 > 3	A260/230 < 1	innerhalb der Spec.
Anzahl der Proben	160	2682	430	1238	332
<i>Mapped Reads</i> [%]	99,5	99,2	99,7	99,5	99,7
Daten [Gb]	11,5	11,3	14	14,1	14,2
Ø Abdeckung	137	142	147	153	148
20-fache Abdeckung [%]	97,5	95,6	98,7	98,3	98,7
nicht-abgedeckt [%]	0,19	0,31	0,16	0,24	0,17

ringerte Werte zeigen zudem Kontaminationen mit organischen Substanzen wie Kohlenhydraten oder Salzen an [125].

Eine vergleichende Auswertung der Qualitätsparameter von Exomen aus DNA mit unterschiedlichen Reinheiten ist in Tabelle 5 dargestellt. Proben mit Phenol- oder Proteinkontaminationen zeigen einen verringerten A260/280 Wert. Dieser sollte zwischen 1,8 und 2,0 liegen. Liegt der Wert unter eins kommt es zu einer geringfügig verringerten 20-fachen Abdeckung auf durchschnittlich 95,6 % der Basen. Zusätzlich steigt die Rate der unabgedeckten Regionen leicht an (Ø 0,3 %) im Vergleich zu Exomen, die aus reinen DNA-Extraktionen hergestellt wurden. Steigt der A260/280 Wert über 3 und befindet sich somit ebenfalls außerhalb der vorgegebenen Spezifikationen, ist kein Effekt sichtbar. Wurden für eine *Library*-Präparation Proben verwendet, die einen verringerten A260/230 Wert aufweisen (< 1) steigt ebenfalls die Rate der unabgedeckten Regionen leicht an (Ø 0,23 %). Des Weiteren sind die Proben im Vergleich zu reiner DNA unauffällig. Eine Steigerung der A260/280 und A260/230 Verhältnisse über einen Wert von 3 hat keine Auswirkungen auf die zu erwartenden Qualitätsparameter einer Exom-Sequenzierung. Auch eine Selektive Betrachtung von DNA-Extraktionen mit extremen Werten von über 15 oder unter -15 für die beiden jeweiligen Absorptionsverhältnisse zeigt keinen Effekt auf die erhaltene Qualität der Sequenzierdaten.

Zusammenfassend kann gesagt werden, dass zu niedrige A260/280 und A260/230 Verhältnisse, welche Kontaminationen mit Proteinen oder organischen Substanzen anzeigen, nur sehr leichte Auswirkungen auf die Datenqualität haben. Diese Effekte befinden sich innerhalb des Toleranzbereiches der Qualitätsparameter ei-

ner Exom-Sequenzierung und sind daher vernachlässigbar.

3.2.4.1.3 Integritätskontrolle von DNA-Proben

Die Integrität der DNA ist ausschlaggebend für die Komplexität der finalen *Libraries*. Eine degradierte DNA kann zu einem kompletten Ausfall einer *Library*-Präparation führen, oder aber, durch den Verlust kleiner Moleküle während der einzelnen Präparationsschritte, ihren Informationsgehalt (Komplexität) verlieren. Zur Beurteilung der Integrität von eingesandten DNA-Proben wurde die Agarose-Gelelektrophorese genutzt. Der Vorteil dieser Methode ist eine schnelle und kostengünstige Beurteilung der DNA. Dazu werden 50 bis 100 ng DNA auf ein 0,7 %iges Agarosegel aufgetragen. Bei 90 V beträgt die Laufzeit 20 min. Im Anschluss wird die an den Fluoreszenzfarbstoff Ethidiumbromid gekoppelte DNA mittels UV-Licht detektiert und der Integritätsgrad beurteilt (Abb.: 38 F). Die subjektive Einschätzung gliedert sich in intakte DNA, leicht degradierte DNA und degradierte DNA. Nachteilig wirkt sich bei dieser Methode die Arbeit mit dem kanzerogenen Stoff Ethidiumbromid aus, sowie die subjektive Einschätzung des Degradationsgrades. Eine Alternative zur Agarose-Gelelektrophorese bietet die Messung mit dem *Caliper LabChip GX* (Perkin Elmer, Abb.: 38 A-E). Der Vorteil dabei ist die zeitgleiche Bestimmung der DNA-Qualität und -Quantität der Probe. Von der DNA wird eine Verdünnung von 60 ng pro 30 μ l (2 ng/ μ l) hergestellt. Der Grad der Degradation des Probenmaterials wird über den sogenannten *Genomic DNA Quality Score* (GQS) über die zugehörige Software ausgegeben. Intakte DNA hat einen GQS um 5, während partiell degradierte DNA einen erniedrigten GQS Wert um 2,5 zeigt. Für degradierte DNA wird ein GQS von 0 ausgegeben. Diese Methode bietet eine skalierbare und standardisierbare Messung der DNA Qualität. Nachteilig wirkt sich der Preis pro Probe und eine notwendige vorab-Bestimmung der Quantität der Probe aus. Die Quantifizierung der Probe vor der Messung ist notwendig um die Verdünnung der Probe in den empfohlenen Rahmen von 2 ng/ μ l zu bringen.

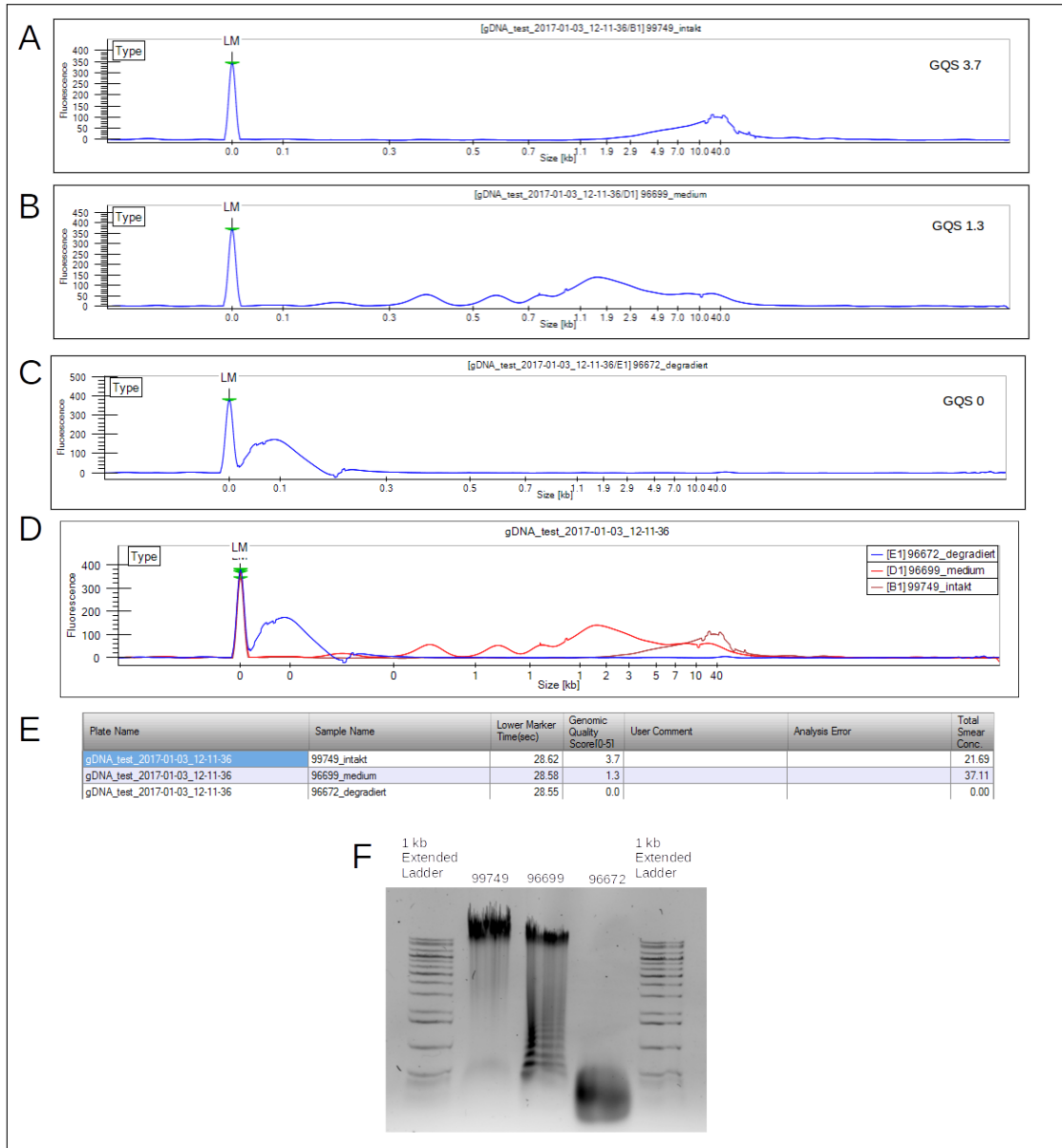


Abbildung 38: Ergebnis der *Caliper LabChip GX* Messung von genomischer DNA in unterschiedlichen Qualitäten im Vergleich zur Agarosegel-Elektrophorese. (A) Intakte genomische DNA. (B) Leicht degradierte DNA. (C) Degradierte DNA. (D) Überlagerung der drei Elektropherogramme. (E) Ergebnistabelle mit der Konzentration und den genomischen Qualitätswert (GQS) der DNAs. (F) Die gleichen Proben aufgetragen auf ein 0,7 % Agarosegel. Als Standard diente der 1 kb *Extended Ladder* von Invitrogen.

DNA Extraktionen, die keine Anzeichen von Degradation aufweisen, können ohne weitere Vorkehrungen in die *Library*-Präparationsschritte gehen. Bei Proben mit Anzeichen von leichter Degradation ist es möglich die Einsatzmenge des Probenmaterials zu erhöhen um die Komplexität der Probe zu erhalten beziehungsweise auf ein Normalniveau zu bringen. Degradierete Proben gelten als Sonderfall und werden nicht in den standardisierten *Library*-Präparationsablauf integriert.

Der höhere Preis, die aufwendige Vorbereitung und die damit verbundene Zeit, um eine Probe mittels *Caliper LabChip GX* zu messen, machen die Anwendung unattraktiv für den Hochdurchsatz. Die Kontrolle der Integrität über ein Agarosegel, die Bestimmung der Reinheit mittels NanoDrop, sowie die Messung der Quantität mittels Qubit reichen aus, um eine Aussage über die Qualifizierung einer DNA für die *Library*-Präparation zu treffen. Die Werte zur Konzentration, Integrität und Reinheit werden als Qualitätsmerkmale im LIMS zur jeweiligen Probe protokolliert und können bei auffälligen Sequenzier-Qualitäten zur Fehlersuche herangezogen werden.

3.2.4.2 Ergebnis der Exom-Sequenzierung von eingesandter DNA versus intern-präparierter DNA

Für eine Studie bei Patienten mit amyotropher Lateralsklerose wurden uns 66 DNAs zur Exom-Sequenzierung eingesandt. Die DNAs wurden im Labor des Kooperationspartners manuell und ohne Anwendung eines kommerziellen Kits durch Aussalzung und Ethanol-Präzipitation isoliert und in Wasser eluiert [126]. Eine Quantifizierung erfolgte mittels NanoDrop ebenfalls durch den Einsender.

Die DNAs wurden wie in Abbildung 37 gezeigt, fluoreszenzbasiert quantifiziert und die Integrität über eine Agarose-Gel geprüft. Die Quantitäten wichen dabei zum Teil erheblich von den Werten der NanoDrop Quantifizierung ab (siehe auch Abschnitt 3.2.4.1.1 Bestimmung der DNA-Konzentration). Die Integrität der Proben war hoch und keine Degradation sichtbar. Wir begannen die *Library*-Präparation mit dem Standard-Protokoll (*SureSelect XT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing, Version E.3, November 2012*) und dem *SureSelect Human All Exon v5 kit*. Der experimentelle Ablauf und die Ergebnisse der Studie sind in Abbildung 39 illustriert.

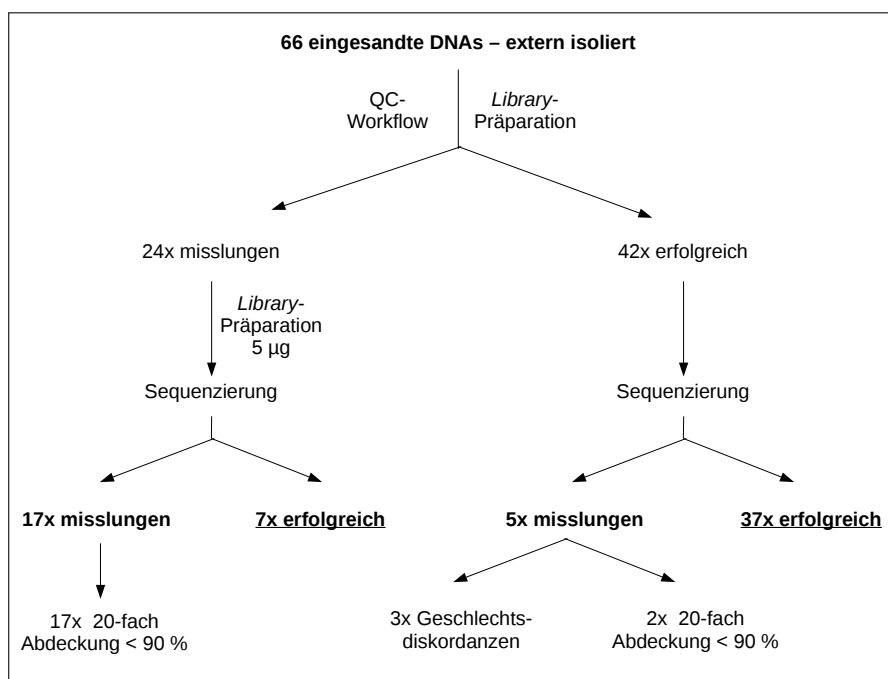


Abbildung 39: Die Prozessschritte und Resultate der *Library*-Präparation und Sequenzierung von eingesandten DNAs.

Für 24 der 66 DNAs konnte keine ausreichende Menge *DNA-Library* (750 ng) für die Hybridisierung erzeugt werden. Die verbliebenen 42 DNAs wurden in eine exomische *Library* konvertiert. Die 24 ausgefallenen DNAs wurden wiederholt in die *Library*-Präparation genommen. In Abweichung zum Standard-Protokoll wurden anstelle von 3000 ng Startmaterial 5000 ng eingesetzt. Damit gelang auch die Herstellung der restlichen *Libraries*. Mittels eines 100 bp *paired-end* Laufs auf dem HiSeq2500 wurden die Proben sequenziert. Bei der Überprüfung der Qualität der *Libraries* innerhalb der Analyseebene des LIMS fiel auf, dass ein Drittel der Proben die geforderte 20-fache Abdeckung des Exoms von mindestens 90 % der Basen trotz ausreichender Anzahl an *Reads* nicht erreicht. Die prozentuale Anzahl der Basen schwankte dabei zwischen 62 und 87 %. Damit einhergehend wurden zwischen 0,9 und 4,7 % der exomischen Sequenz nicht abgedeckt. Dieser Wert schwankt bei qualitativ hochwertigen Sequenzen zwischen 0 bis 0,2 %.

Von vier der 22 misslungenen Exom-Sequenzierungen stand uns EDTA-Blut zur Verfügung. Um einen möglichen Zusammenhang zwischen dem Ausgangsmaterial und den Sequenzier-Ergebnissen zu finden isolierten wir die DNA aus den Blutröhrchen über eine *Magnetic Bead*-basierte Automationsplattform, dem

Chemagic 360 der Firma Perkin Elmer. Die Eignung der DNA wurde im Anschluss über die beschriebene Qualitätsprüfung ermittelt, eine *Library* nach dem Standard-Protokoll mit 3000 ng Startmaterial hergestellt und in einem 100 bp *paired-end* Lauf auf dem HiSeq2500 sequenziert. Die erhaltenen Daten entsprachen einem qualitativ hochwertigem Exom mit vergleichbaren Qualitätswerten zu einem Durchschnitts-Exom.

Der beschriebene Effekt der misslungenen *Library*-Präparation und/oder Sequenzierung konnte im Verlauf der Jahre auch bei anderen Projekten, wenn auch in kleinerem Maßstab, beobachtet werden. Warum manche DNA-Extraktionen zu schlechten Qualitätsparametern nach der Exom-Sequenzierung führen, obwohl sie den qualitativen und quantitativen Vorgaben entsprachen, kann am Ende nicht zufriedenstellend geklärt werden. Es könnten sich weiterhin Inhibitoren wie Antikoagulanzen, divalente Kationen (Kalzium) oder Enzyme in der DNA-Lösung befinden [127], die sich möglicherweise über eine Absorptionsmessung zur Bestimmung der Reinheit nicht detektieren lassen, aber Effekte auf die *Library*-Präparation haben.

Im Zuge eines anderen Projektes wurde DNA zweier Patienten für die *NG*-Sequenzierung zur Verfügung gestellt. Der Einsender konnte keine Informationen über die DNA-Isolationsmethode geben. Die Qualitäts- und Quantitätskontrolle der Proben wurde wie beschrieben durchgeführt und die DNAs qualifizierten sich für eine *Library*-Präparation. Die Analyse der Daten im Anschluss an die Sequenzierung zeigte eine geringe 20-fache Abdeckung von 73 bzw. 79 %. Ein bis 1,6 % der Basen wurden nicht abgedeckt und der *CNV Noise* war sehr hoch. Dem Einsender bestand keine Möglichkeit EDTA-Blut zu senden. Daher wurden die DNA über das *Genomic DNA Clean & Concentrator* Kit (Zymo Research) aufgereinigt um mögliche Inhibitoren zu entfernen. Von der aufgereinigten DNA wurde im Anschluss noch einmal eine exomische *Library* präpariert und sequenziert. Die Ergebnisse unterschieden sich nur marginal von denen aus der ersten Präparationsrunde. Eine Aufreinigung der DNAs hat zu keiner Verbesserung der finalen Daten geführt (Tab.: 6). Daher kann auch in diesem Fall die Einsendung von EDTA-Blut empfohlen werden. Ob die Aufreinigung der DNA nur in diesem Fall nicht gelungen ist, oder generell davon abgesehen werden sollte, kann an dieser Stelle nicht beantwortet werden. Dazu müssen in Zukunft mehr solcher auffälliger Proben aufgereinigt und ein zweites Mal analysiert werden.

Tabelle 6: Übersicht zur Sequenzier-Statistik von unaufgereinigter DNA versus aufgereinigter DNA. Die DNA von exomischen *Libraries* mit schlechten Qualitätsparametern nach der Sequenzierung (11592 und 11598) wurde über das *Genomic DNA Clean & Concentrator* Kit (Zymo Research) aufgereinigt und noch einmal sequenziert (11592W und 11598W). In der ersten Spalte befinden sich die durchschnittlichen Werte von 100 qualitativ hochwertigen Exomen im Vergleich.

Parameter	Ø Exom	11592	11592W	11598	11598W
<i>Mapped Reads</i> [%]	99,7	99,8	99,8	99,8	99,8
Daten [Gb]	14,2	10,07	11,17	9,6	9,78
<i>On-Bait</i> [%]	77,8	69,99	71,87	71,93	71,47
Ø Abdeckung	161	95	108	93	94
20-fache Abdeckung [%]	98,7	79,25	81,77	73,17	73,09
<i>CNV Noise</i>	1.65	9,47	9,81	10,34	10,41

3.2.4.2.1 Ableitung der Anforderungen an eingesandtes Material für eine Exom-Sequenzierung

Aus der Sammlung und dem Vergleich von Parametern zur Qualität, Reinheit und Quantität einer DNA und den zugehörigen Qualitätsparametern der Sequenzierung, können Anforderungen an das eingesandte Material für eine Exom-Sequenzierung abgeleitet werden: (a) wenn die Möglichkeit besteht, sollte vom Einsender EDTA-Blut anstelle von DNA zur Verfügung gestellt werden. Die Isolation von DNA aus peripheralem Blut mit dem *Chemagic 360* kann am Institut für Humangenetik durchgeführt werden. Erfahrungsgemäß lassen sich aus diesen DNA-Isolationen erfolgreich exomische und genomische *Libraries* mit hoher Qualität herstellen. Kann ein Einsender kein Blut, sondern DNA zur Verfügung stellen, muss gewährleistet sein, dass diese eine hohe Integrität besitzt und fluoreszenzbasiert quantifiziert wurde. Kann von einer eingesendeten DNA keine *Library* mit den entsprechenden Qualitätsparametern hergestellt werden, obwohl die Qualitätskontrolle unauffällig war, wird empfohlen EDTA-Blut einzusenden. Eine Aufreinigung von DNA über ein kommerzielles Kit zur Verbesserung der Qualität kann nicht generell empfohlen werden. Die Anwendung muss an weiteren Proben getestet werden, um ihre Auswirkungen genauer bestimmen zu können.

3.2.4.3 Optimierung des Laborablaufs zur Steigerung des Durchsatzes

Das Ziel im Bereich der Hochdurchsatzsequenzierung ist es, die Laborabläufe so aufzustellen, dass ein möglichst hoher Probendurchsatz gewährleistet ist. Neben der Einführung der automatisierten *Library*-Präparation konnten weitere Engpässe im Labor identifiziert und optimiert werden (Abb.: 7).

Zur initialen Fragmentierung nutzen wir den Covaris S2. Dabei wurden die Probengefäße nacheinander in das Gerät gestellt, fragmentiert, entnommen und eine neue Probe platziert. Der Covaris S2 wurde mit der Einführung des Hochdurchsatzes durch den Covaris E220 ersetzt, der mit 96 Proben beladen werden kann, die nacheinander im Gerät fragmentiert werden. Von der Bearbeitung einer Probe pro 6 Minuten mit hohem Personalaufwand konnte der Durchsatz somit auf 96 Proben mit minimaler manueller Arbeit gesteigert werden. Gleichzeitig war es möglich die Proben über Nacht zu fragmentieren und einen Arbeitstag einzusparen.

In Vorbereitung zur Hybridisierung sieht es das *Library*-Präparationsprotokoll vor, 750 ng einer jeweiligen *Library* auf 3,4 µl einzudampfen. Dazu nutzen wir zu Beginn die Savant SpeedVac Zentrifuge (Thermo Fisher), mit der 24 Proben in Röhrchen in 3 Stunden verdampft werden konnten. Mit der Anschaffung eines Plattenrotors konnten bis zu 96 Proben in 3 Stunden eingedampft werden. Der Ersatz der Savant SpeedVac durch den Eppendorf Concentrator Plus erlaubt es nunmehr 96 Proben in 35 Minuten einzudampfen. Die Vorbereitungszeit verkürzte sich damit so stark, dass es nun möglich ist am Morgen eines Arbeitstages die Proben für die Hybridisierungsreaktion einzudampfen und im Anschluss die Reaktion zu starten. Insgesamt ergab dies eine Reduktion von einem Arbeitstag.

Zur Bestimmung der Qualität und Quantität (QC) der Probe innerhalb und am Ende der *Library*-Präparation nutzen wir den Bioanalyser 2100 (Agilent). Für die Bestimmung der Konzentration und Größe einer *Library* von zwölf Proben wurde eine Stunde benötigt. Für einen Hochdurchsatz von 96 Proben würde dies einen ganzen Arbeitstag in Anspruch nehmen. Zur Steigerung des Durchsatzes tauschten wir den Bioanalyser gegen den *Caliper LabChip GX* (Perkin Elmer). 96 Proben können in 2 Stunden analysiert und ausgewertet werden. Dabei ist der personelle Aufwand gering. Zusätzlich wurden die Qualitätskontrollschritte

Tabelle 7: Optimierung der Laborausstattung zur Erhöhung des Durchsatzes und zur Verminderung der Anwenderzeit.

Prozess	Gerät	Nachteil	Verbesserung	Vorteil
<i>Library</i> -Präparation	Bravo	48 Proben	Bravo NGS Option B	96 Proben
Fragmentierung	Covaris S2	Einzeltubes	Covaris E220	Parallelisierung
Eindampfen	Savant Speed-Vac	24 Einzeltubes	Eppendorf Concentrator	96-well Platte
Qualitätskontrolle	Bioanalyzer	12 Einzelproben	Caliper Lab Chip GX	96-well Platte

auf dem *Caliper LabChip GX* von drei auf zwei reduziert. Neben den zwingend notwendigen Qualitätskontrollen vor der Hybridisierung und am Ende der Präparation, konnte die Analyse nach der initialen Fragmentierung entfallen. Diese diente lediglich der Sicherstellung einer gelungenen Fragmentierung. Erfahrungsgemäß ist die Ausfallrate, nachdem die DNA vor Beginn der Präparation als qualifiziert eingestuft wurde, an dieser Stelle sehr gering und vernachlässigbar, so dass auf diesen Qualitätskontrollschritt verzichtet werden kann.

Insgesamt ergibt sich durch die Optimierung eine Einsparung von mehreren Arbeitstagen. 96 Exome können in 7 Werktagen produziert werden. 96 Genome benötigen eine Arbeitszeit von 5 Werktagen, von Beginn der *Library*-Präparation bis hin zum äquimolaren Mischen der Proben. Der limitierende Faktor ist hierbei die notwendige qPCR zur Bestimmung der amplifizierbaren DNA-Moleküle innerhalb der PCR-freien *Library*-Präparation. Sie erfordert eine konzentrierte und genaue Arbeitsweise des Bearbeiters. Dadurch verringert sich der Durchsatz auf maximal 24 *Libraries* pro qPCR-Ansatz und zwei qPCRs an einem Tag. Das Pipettieren der qPCR soll in Zukunft automatisiert über die *Bravo Liquid Handling Platform* stattfinden um reproduzierbarer und in höherem Durchsatz, bei weniger Personalaufwand arbeiten zu können.

3.2.4.4 Entwicklungen von Methodiken zum Ausschluss von Probenverwechslungen und Kontaminationen

Der Ablauf der *Next-Generation-Sequenzierung*, vom Probeneingang bis hin zur Sequenzierung erfolgt in vielen manuellen Schritten. Das birgt die Gefahr der Probenverwechslung und Kontamination. Die Sicherstellung der Probenidentität ist von herausragender Bedeutung. Eine Vertauschung oder Verunreinigung von Proben ist ein schwerwiegendes Problem und führt zu falsch-positiven Assoziationen. Eine Kontamination von Proben führt zudem zur Verringerung der Genotypqualität und ist vor allem bei Sequenzierprojekten mit geringer Basenabdeckung problematisch [58].

Eine Vertauschung oder Kontamination von Proben kann vor oder während der *Library*-Präparation stattfinden. Neben den klassischen Empfehlungen zur Laborarbeit, wie dem Tragen von Handschuhen und der Dekontamination der Arbeitsplätze- und Geräte, wurden weitere Maßnahmen unternommen um die Probenintegrität auch im Hochdurchsatz zu erhalten (Tab.: 8).

Tabelle 8: Übersicht zum Qualitätsmanagement.

Prozessschritt	Maßnahme
prä- <i>Library</i> -Präparation	<ul style="list-style-type: none"> - Überführung der DNA/RNA in 2D-Barcode Röhrrchen - Lagerung an einem definierten Stellplatz in einer 96er-Platte - Dokumentation im LIMS - Amelogenin-PCR zur Geschlechtsbestimmung - Verwendung von <i>Sample Tracking Sheets</i> - Barcode Kontrolle vor der <i>Library</i>-Präparation
<i>Library</i> -Präparation	<ul style="list-style-type: none"> - Probenprozessierung in 96er Ansätzen - Format und Plattenposition wird bei jedem Plattentransfer beibehalten - notwendige manuelle Pipettierschritte über 8- oder 12-Kanal Pipetten - fertige <i>Libraries</i> werden in 2D-Barcode Röhrrchen gelagert - Dokumentation der definierten Plattenposition des <i>Library</i>-Röhrrchens im LIMS - Index-Kontrolle für den <i>Pool</i> durch das LIMS - Anlegen der <i>Pools</i> durch Einscannen der 2D-Barcode Röhrrchen - fertige <i>Pools</i> werden in 2D-Barcode Röhrrchen gelagert - Dokumentation der definierten Plattenposition des <i>Pool</i>-Röhrrchens im LIMS - Festlegung der <i>Flow Cell</i>-Beladung durch einscannen der 2D-Barcode Röhrrchen im LIMS
post- <i>Library</i> -Präparation	<ul style="list-style-type: none"> - Überführung der DNA/RNA in 2D-Barcode Röhrrchen - Bestimmung der Anzahl der <i>SRY-Reads</i> - Kontaminationsbestimmung nach Jun <i>et al.</i> [58]

Ein wichtiger Schritt war die Einführung von von 2D-Barcode Röhrrchen wie unter Abschnitt 3.1.5.3 beschrieben um eine Probenverwechslung zu minimieren. Manuelle Einträge werden durch das einfache Einscannen der 2D-Barcode Röhrrchen vermieden und dadurch Eingabefehler minimiert. Daneben wurden weite-

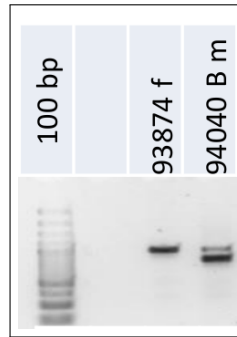


Abbildung 40: Ergebnisse einer Amelogenin-PCR. Das homologe Amelogenin-Gen liegt sowohl auf dem X- (*AMELX*) als auch auf dem Y-Chromosom (*AMELY*). Die beiden homologen Bereiche unterscheiden sich in der Länge des Intron 3 [128]. Eine PCR mit einem Primerpaar in Intron 3 ist ausreichend zur Geschlechtsbestimmung. Für weibliche DNA (93874f) entsteht ein Produkt mit 977 bp Größe, während bei männlicher DNA (94040Bm) zwei Produkte bei 977 bp und 790 bp amplifiziert werden.

re Methodiken zur Vermeidung und Detektion von Probenkontaminationen und Probenverwechslungen eingeführt, die im Folgenden dargestellt sind.

Eine Möglichkeit zur Bestimmung von Probenverwechslungen vor der *Library*-Präparation bietet die Geschlechtsbestimmung per Amelogenin-PCR (Abb.: 40) [128]. Die Ergebnisse werden mit den Angaben in den medizinischen Unterlagen verglichen. Es kann allerdings nur ein Vertauschen zwischen beiden Geschlechtern detektiert werden. Der Austausch zweier Proben innerhalb einer Geschlechtergruppe bleibt dabei unbemerkt. Gleichzeitig dient die Methodik nur dem Ausschluss von Verwechslungen vor der *Library*-Präparation, aber nicht innerhalb des Prozesses. Dennoch bietet die Anwendung den Vorteil einer schnellen und kostengünstigen Geschlechtsbestimmung und einer ersten Qualitätskontrolle. Die Bestimmung einer Proben-Kontamination ist mit dieser Untersuchung nicht möglich.

Während der *Library*-Präparation kann es zu Vertauschungen von Probengefäßen kommen. Durch das Arbeiten in 96-well PCR-Platten und dem Einsatz von 8- oder 12-Kanal Pipetten, wird ein Vertauschen minimiert oder kann nachträglich systematisch zugeordnet werden. Die Proben bleiben von Beginn der Präparation bis zum Ende an der gleichen Position einer Mikrotiter-Platte. Die Position wird im *Sample Tracking Sheet* hinterlegt.

Eine Kontamination von Proben kann im gesamten Ablauf der *Library*-Herstellung stattfinden. Es ist daher zwingend notwendig geeignete Vorsichtsmaßnahmen zu treffen. Dazu gehören die Desinfektion des Arbeitsplatzes und der

Arbeitsmittel mit DNA und RNA dekontaminierenden Mitteln, sowie das Tragen von Handschuhen. Eine sorgfältige Arbeitsweise verhindert ein Überschwappen von Proben innerhalb der Mikrotiterplatte.

Eine Bestimmung von Probenverwechslungen, vor allem aber Kontaminationen, kann anhand der Sequenzdaten im Anschluss an die Sequenzierung erfolgen. Dabei können drei mögliche Strategien angewendet werden: der Abgleich von Sequenzierdaten mit Array-basierten Genotypdaten, über eine Analyse der Sequenzierdaten oder der Array-basierten Genotypdaten allein [58]. In den meisten Fällen stehen zu den einzelnen Proben keine Genotypdaten zur Verfügung. Nach einer statistischen Methode, entwickelt von Jun *et al.* [58], werden die Exom- und Genom-Daten, wie in Abschnitt 3.1.5.5 beschrieben, auf Kontaminationen überprüft und der Grad einer Kontamination in Prozent ausgegeben. Zusätzlich dazu erfolgt die Bestimmung der Anzahl von *SRY-Reads*. Eine zu geringe Anzahl an *SRY-Reads* bei Männern sowie auftretende *SRY-Reads* bei Frauen deuten auf eine Kontamination hin. Proben mit einer Verunreinigung von mehr als 5 % werden von den weiteren Analysen exkludiert. Erfahrungsgemäß ist eine hohe Kontamination schon vor Beginn der *Library*-Präparation vorhanden. Eine wiederholte Untersuchung des Patientenmaterials kann nur mit einer neuen Einsendung von Blut oder neu isolierter DNA erfolgen.

3.2.5 Austestung einer Exom-*Library*-Präparation mit weniger Einsatzmenge und größerer *Insert*-Größe

3.2.5.1 Hintergrund

Das im Institut für Humangenetik angewendete Protokoll zur Anreicherung exomischer Sequenzen, ist das *SureSelect XT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing, Version E.3, November 2012*. Das Protokoll arbeitet mit einer Einsatzmenge von 3000 ng und einer durchschnittlichen *Insert*-Größe von 200 Basenpaaren. Als *Insert*-Größe wird die Größe der zu untersuchenden genomischen DNA innerhalb der 3'- und 5'-Adaptersequenz bezeichnet (Abb.: 41).

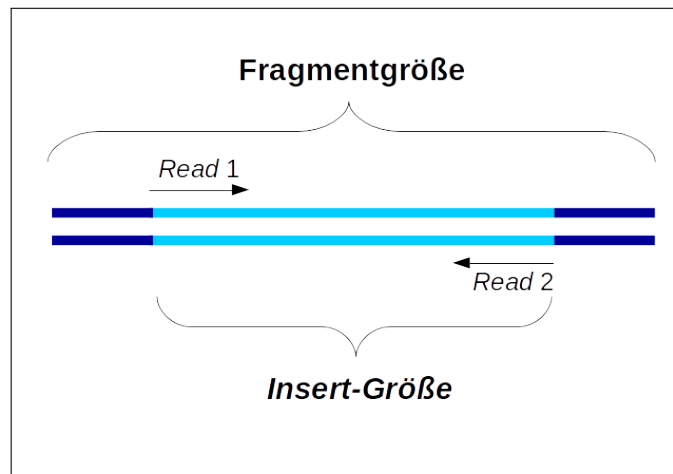


Abbildung 41: Schematische Darstellung eines *Library*-Moleküls. Das *Library*-Fragment besteht aus der genomischen DNA (hellblau), die an plattformspezifische Adapter (dunkelblau) ligiert ist. Beide Elemente summieren sich zur Fragmentgröße, während das Stück genomische DNA als *Insert* bezeichnet wird. Die Sequenzierung beginnt mit dem ersten 5'-Nukleotid des *Inserts*.

Exom-*Libraries* mit einer durchschnittlichen *Insert*-Größe von 200 Basenpaaren zeigen bei einem 100 bp *paired-end* Lauf bis zu 40 % überlappende *Reads* (Abb.: 42B und 43A). Bei *Insert*-Größen kleiner als 100 Basen endet der *Read* im Adapter (Abb.: 42A). Überlappende *Reads* führen zu keinem neuen Datengewinn. Daher sollte als Teil dieser Arbeit zum Einen die *Insert*-Größe der exomischen *Libraries* vergrößert werden, um die Anzahl der überlappenden *Reads* zu reduzieren. Zum Anderen sollte die Einsatzmenge von 3000 ng auf 500 ng reduziert werden um auch Proben prozessieren zu können, von denen keine 3000 ng zur Verfügung stehen. Begrenzte Mengen an DNA stammen zum Beispiel von Tumorproben

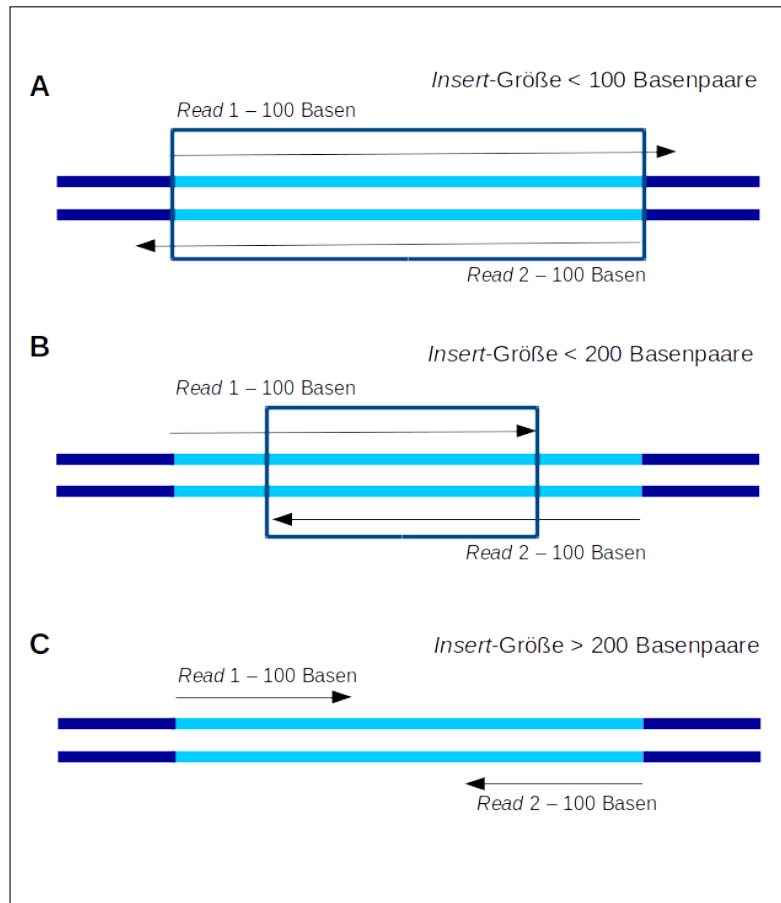


Abbildung 42: Schematische Darstellung des Effekts verschiedener *Insert*-Größen auf die Sequenzierung. **A** Bei *Insert*-Größen (türkis) kleiner als 100 Basen würden die *Reads* 1 und 2 (schwarze Pfeile) in der Adaptersequenz (dunkelblau) enden. **B** Bei *Insert*-Größen kleiner als 200 Basen kommt es zu einer Überlappung der *Reads*. **C** Bei *Insert*-Größen von mehr als 200 Basen kommt es zu einer Annäherung der *Reads* ohne Überlappung.

oder Trockenblutkarten wie sie beim Neugeborenen-Screening verwendet werden.

3.2.5.2 Methodik und Ergebnisse

Für Herstellung von exomischen *Libraries* aus 500 ng Startmaterial mit einer Insert-Größe von 350 Basenpaaren wurden drei kommerziell verfügbare Protokolle miteinander kombiniert. Zum einen das Standard-Protokoll von Agilent (*SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library*) und zum anderen die Illumina Protokolle *TruSeq DNA PCR Free* und *TruSeq Nano DNA*. Das *SureSelect XT* Protokoll arbeitet mit 3000 ng Startmaterial und 200 Basenpaaren Insert-Größe und das *TruSeq Nano* Protokoll mit 100 ng Startmaterial und einer Insert-Größe von 350 Basenpaaren. Um ausreichend Material

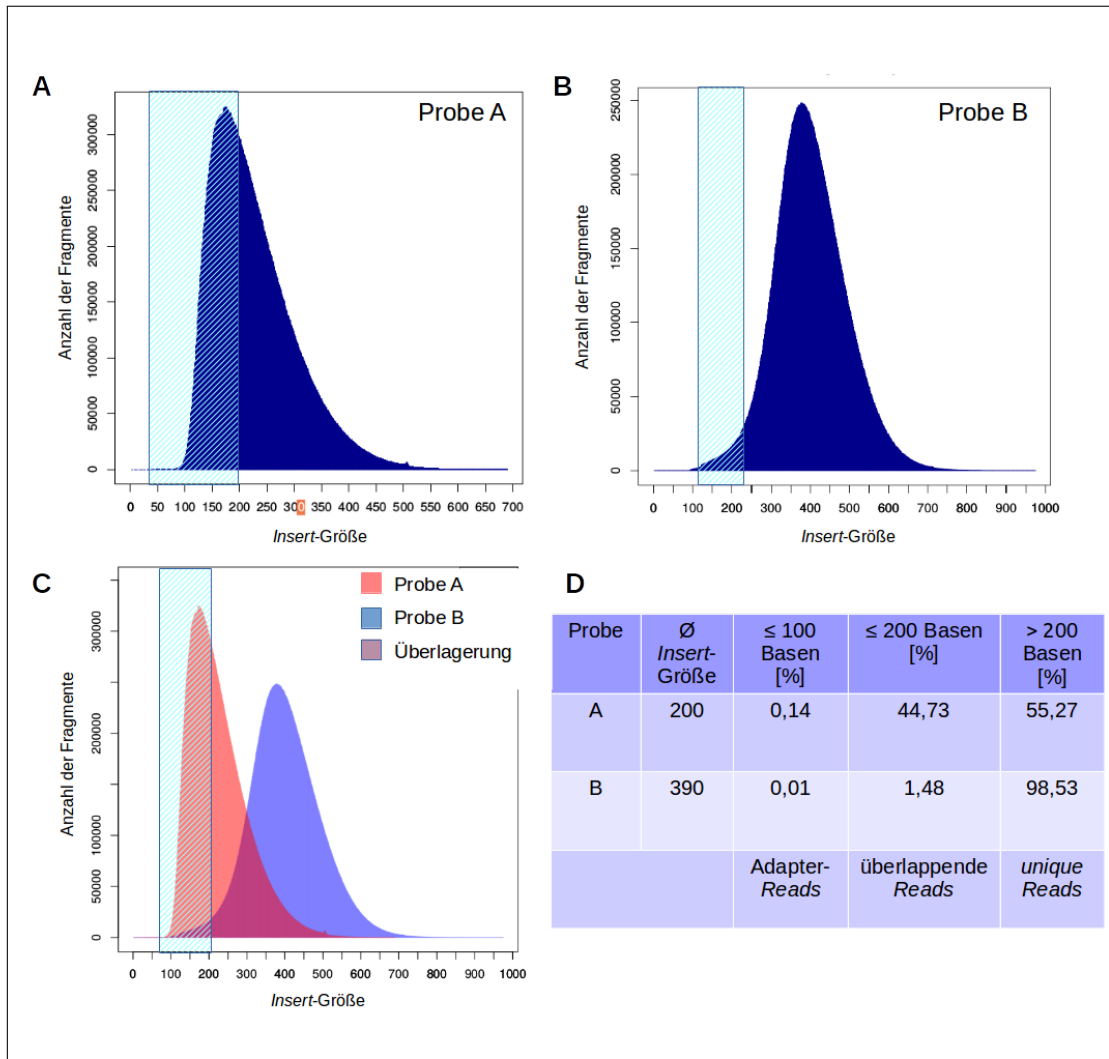


Abbildung 43: Größenverteilung einer exomischen *Library*. **A** Die *Insert*-Größen einer Standard-*Library* (Probe A) verteilen sich auf 90 bis 570 Basenpaare. 40 % der Fragmente sind kleiner/gleich 200 Basen und resultieren in überlappenden *Reads* (hellblau schraffierter Bereich). **B** Wird die *Insert*-Größe von 200 Basenpaaren auf 390 Basenpaare erhöht (Probe B), verringert sich die Zahl der überlappenden *Reads* (hellblau schraffierter Bereich) auf unter 1,5 %. **C** Überlagerung der *Insert*-Größen-Verteilung bei der Anwendung des Standard-Protokolls (rosa) und des adaptierten Protokolls (blau). **D** Vergleichende Aufstellung der prozentualen Anteile verschiedener Fragmentlängen für Probe A und Probe B.

für die Hybridisierungsreaktion zu bekommen (1500 ng), wurden 500 ng in die Test-Präparation eingesetzt. Aus dem *TruSeq DNA PCR Free* Protokoll wurde der Größen-Selektionsschritt übernommen. In der Tabelle 9 ist die Kombination der drei Protokolle für die jeweiligen Prozessschritte aufgegliedert.

Tabelle 9: Kombinatorische Anwendung dreier kommerzieller *Library*-Präparationskits zur Herstellung von exomischen-*libraries* mit weniger Einsatzmenge und größer *Insert*-Größe. In der ersten Spalte sind die einzelnen Prozessschritte aufgegliedert. Das X kennzeichnet, aus welchem Protokoll die Parameter für die *Library*-Präparation entnommen wurden.

Prozessschritt	<i>SureSelect XT</i>	<i>TruSeq Nano DNA</i>	<i>TruSeq DNA PCR Free</i>
Fragmentierung		X	
AMPure XP <i>Bead</i> -Aufreinigung			X
<i>End-Repair</i>			X
Größenselektion via AMPure XP <i>Beads</i>			X
Addition einer A-Base			X
AMPure XP <i>Bead</i> -Aufreinigung			X
Adapter-Ligation	X		
AMPure XP <i>Bead</i> -Aufreinigung		X	
Amplifikation	X mit 10 Zyklen		
Hybridisierung bis zur finalen <i>Library</i>	X mit 1500 ng		

Für den Versuch wurden acht verschiedene DNAs verwendet. Die finalen *Libraries* wurden in einem 100 bp *paired-end* Lauf auf dem HiSeq2500 sequenziert. In Abbildung 43B ist exemplarisch die Verteilung der *Insert*-Größen dargestellt. Die durchschnittliche Größe der acht Proben belief sich auf 390 Basenpaare und war damit größer als auf die gezielten 350 Basenpaare. Lediglich 1,5 % der *Inserts* war kleiner als 200 Basenpaare. Mehr als 98 % der *Reads* waren nicht-überlappend.

3.2.5.3 Effekt der gesteigerten *Insert*-Größe und DNA-Einsatzmenge auf die Qualitätsparameter der Sequenzierung

Ein Vergleich der Qualitätsparameter zwischen Standard-Exomen und den 350 Basenpaar-Exomen zeigt eine Reduktion der durchschnittlichen exomischen Abdeckung unter die geforderten 100-fach und eine Reduktion der *On-Bait*-Rate um mehr als 20 % (Abb.: 44). 95 % der Basen sind mehr als 20-fach abgedeckt. Bedingt durch die größeren *Inserts* liegen mehr *Reads* außerhalb der definierten exomischen Bereiche, was zu einer Verminderung der exomischen Abdeckung führt.

Durch die Verringerung der DNA-Einsatzmenge war eine Erhöhung der Duplikatrate im Vergleich zum Standard-Exom zu erwarten. Tatsächlich unterschied sich die durchschnittliche Duplikatrate bei den 500 ng-Exomen nur um 2,5 %, was einen akzeptablen Wert darstellt (Abb.: 44D).

Das Test-Protokoll bildet einen guten Ausgangspunkt für die weitere Optimierung der *Insert*-Größe bei gleichbleibender Einsatzmenge von 500 ng DNA. Es ist anzunehmen das eine moderate Erhöhung der *Insert*-Größe von 200 Basenpaaren auf 250 oder 300 Basenpaare, die durchschnittliche exomische Abdeckung auf mehr als 100-fach erhöhen, sowie die *On-Bait*-Rate verbessern würde. Dabei ist ein Kompromiss zwischen der Anzahl überlappender *Reads* und dem Erreichen ausreichender Qualitäten zu finden. Mit dem hier beschriebenen Protokoll wurden zwar die überlappenden *Reads* minimiert, aber nicht die gewünschte exomische Abdeckung erreicht. Die Stellschrauben zur *Insert*-Größe für folgende Experimente sollte die initiale Fragmentierung und die Größenselektion über AMPure XP *Beads* sein, um Größen zwischen 250 und 300 Basenpaaren zu testen.

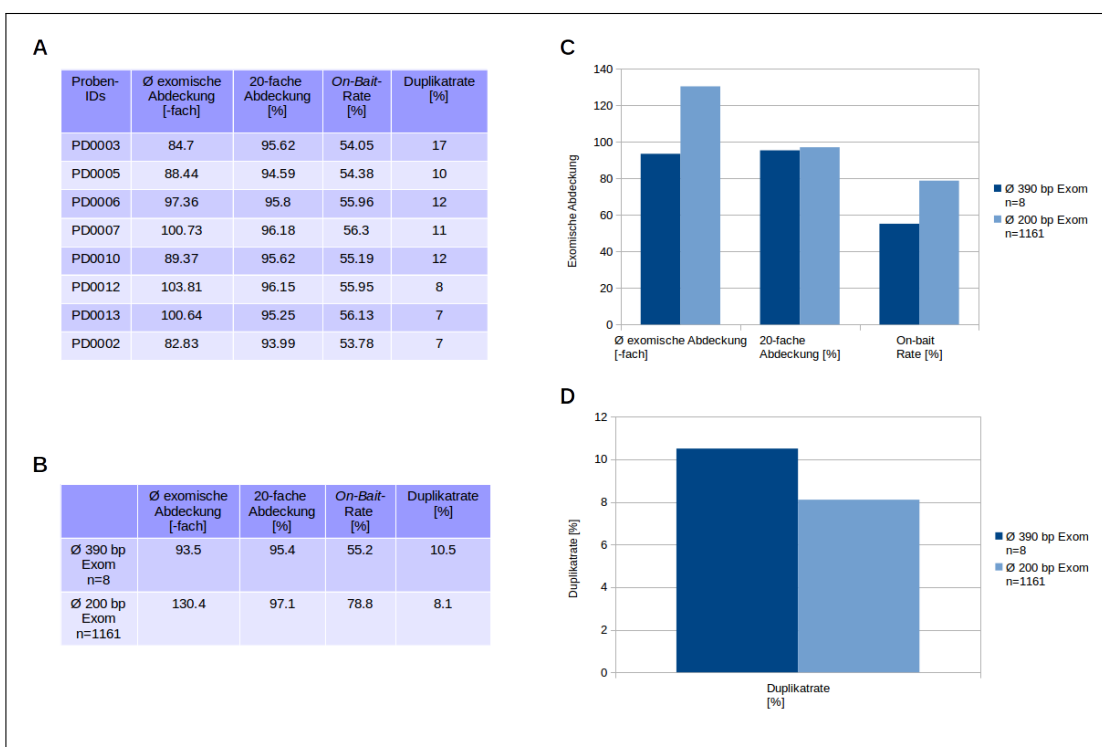


Abbildung 44: Resultate der Exom-Sequenzierung von *Libraries* mit 500 ng DNA-Einsatzmenge und höherer *Insert*-Größe. Für den Versuch wurden acht verschiedene genomische DNAs genutzt. **A** Die Qualitätsparameter der Exom-Sequenzierung von acht Test-Proben. **B** Die Durchschnittswerte der 350 Basenpaar-Exome im Vergleich zu 1000 Standard-Exomen mit einer 200 bp *Insert*-Größe. **C** und **D** Die Sequenzierung zeigt eine reduzierte exomische Abdeckung und eine Verminderung der *On-Bait*-Rate. Dagegen ist die durchschnittliche 20-fache Abdeckung gering vermindert und ebenso wie die leicht erhöhte Dupliktrate akzeptabel.

3.2.6 Evolution der NGS-Plattform am Institut für Humangenetik

Mit meiner Einstellung am Institut für Humangenetik im Jahr 2009, war ich maßgeblich an der Etablierung und Entwicklung der NGS-Plattform beteiligt. Mit der Anschaffung zweier Illumina *Genome Analyzer II* und der Genom-Sequenzierung von *Caenorhabditis elegans* und Hausrindern begann 2008 die Anwendung der NG-Sequenzierung. Ein Jahr später führte ich im Rahmen des NGFN-Ringversuchs (Nationales Genomforschungsnetz) die erste Exom-*Library*-Präparation durch. Mit dem Beginn der Exom-Sequenzierung Ende 2009 war ein geringer Durchsatz der *Library*-Präparation und Sequenzierung ausreichend. Die stetig fallenden Kosten einer Sequenzierung und die Erfolgsaussichten einer NGS-Studie machen die Exom- und Genom-Sequenzierung zu gefragten Anwendungen. Eine Aufskalierung des Durchsatzes der *Library*-Präparation und Sequenzierung war unumgänglich. Die Entwicklung der NGS-Plattform hat sich der Nachfrage der vergangenen Jahre durch die stetige Weiterentwicklung und Optimierung angepasst (Abb.: 45).

Neben den beschriebenen Entwicklungen der Exom- und Genom-*Library*-Präparation führte ich 2014 eine Automation der mRNA-*Library*-Präparation ein, die es erlaubt 96 RNA-Proben parallel zu prozessieren. Derzeit stehen die drei Hauptanwendungen der NG-Sequenzierung im Hochdurchsatz zur Verfügung. Insgesamt wurden seit 2008 18.083 Exom-, Genom- und Transkriptom- *Libraries* präpariert mit einem Durchsatz von circa 5000 *Libraries* in diesem Jahr (2017) (Abb.: 46 und 47).

Neben den Entwicklungen im Laborbereich wurden auch die Sequenziermaschinen aktuell gehalten. Seit dem Jahr 2016 sequenzieren wir auf zwei Illumina *HiSeq4000* Maschinen, mit einem Durchsatz von bis zu 1,5 Tb pro Maschine pro Lauf. In diesem Jahr generierten wir 105,495 Tb Daten und insgesamt 313,568 Tb seit Beginn der NG-Sequenzierung (Abb.: 48).

Die Anwendung erstreckt sich dabei über verschiedene Spezies wie Tiere (Maus, Huhn, Rind, Pferd, Zebrafisch und Andere), Pflanzen (Mais, Ambrosia, Arabidopsis) und Bakterien (*Helicobacter*, *Chlamydia*, *Sulfolobus*). Im Vordergrund steht die Sequenzierung humaner DNA und RNA von Patienten mit seltenen und häufigen Erkrankungen. Den Hauptteil bilden Entwicklungsstörungen, neurologische Erkrankungen, Mitochondriopathien und Herzerkrankungen (Abb.: 49).

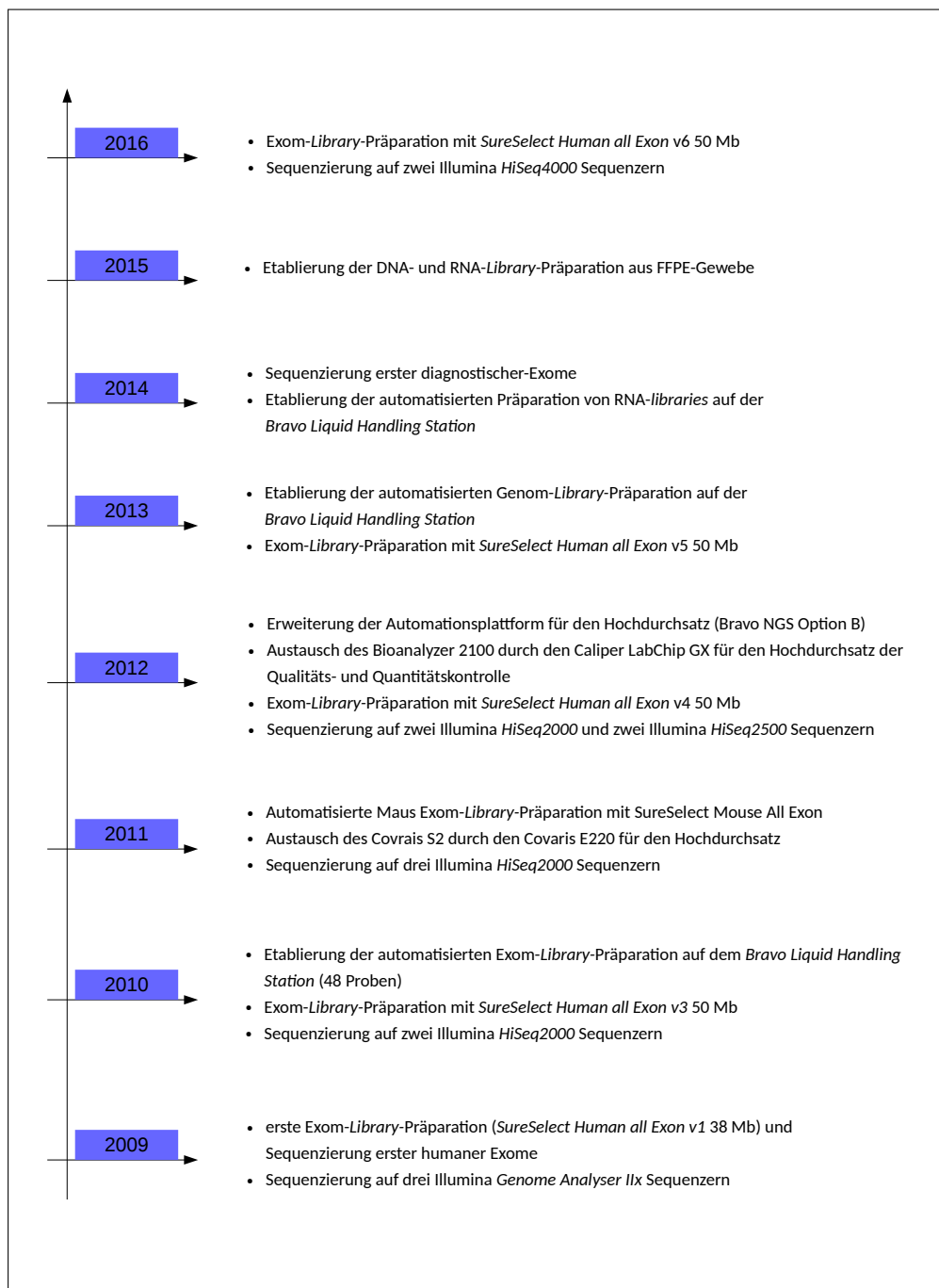


Abbildung 45: Zeitachse der Evolution der *NGS*-Plattform am Institut für Humangenetik.

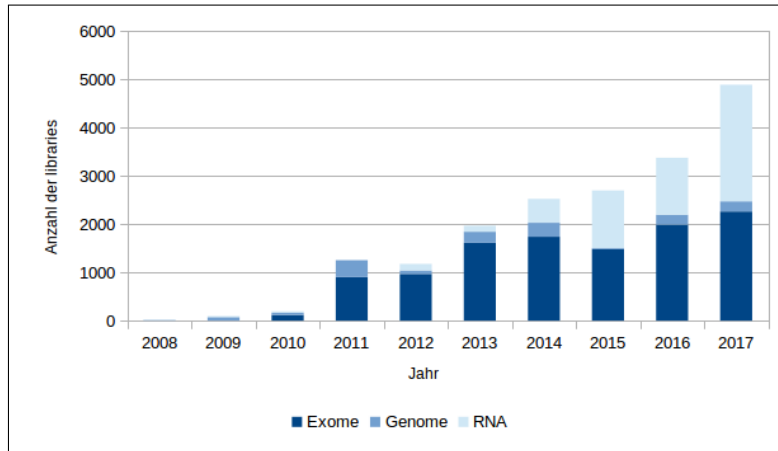


Abbildung 46: Übersicht zum Durchsatz der jährlichen *Library*-Präparation.

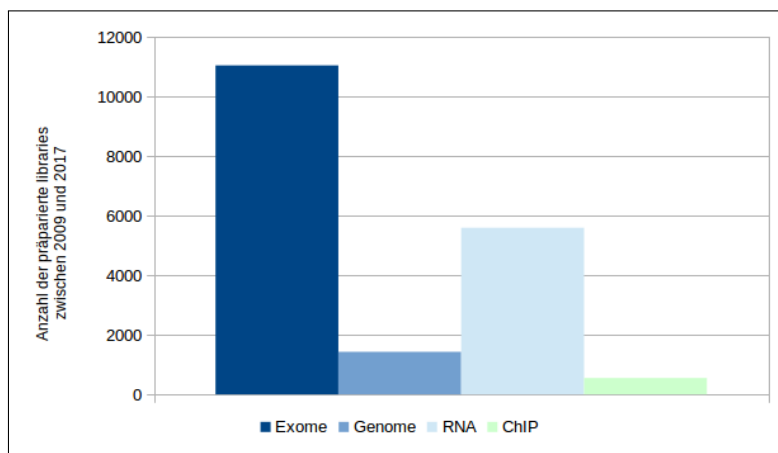


Abbildung 47: Anzahl der präparierten *Libraries* von 2009 bis November 2017.

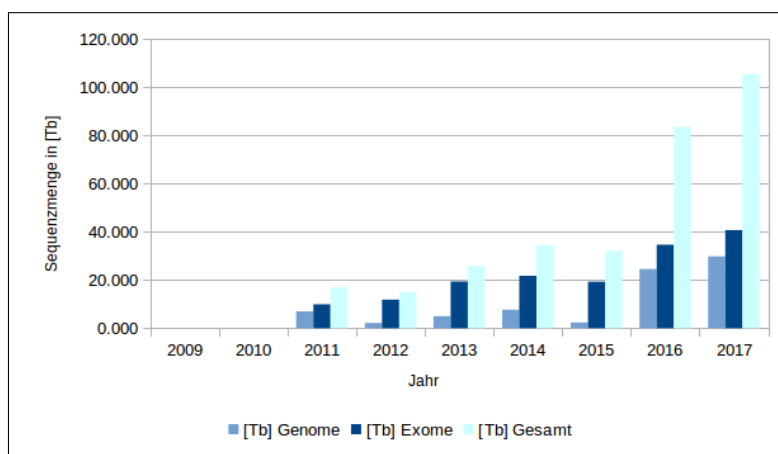


Abbildung 48: Generierte Datenmenge pro Jahr von 2009 bis November 2017.

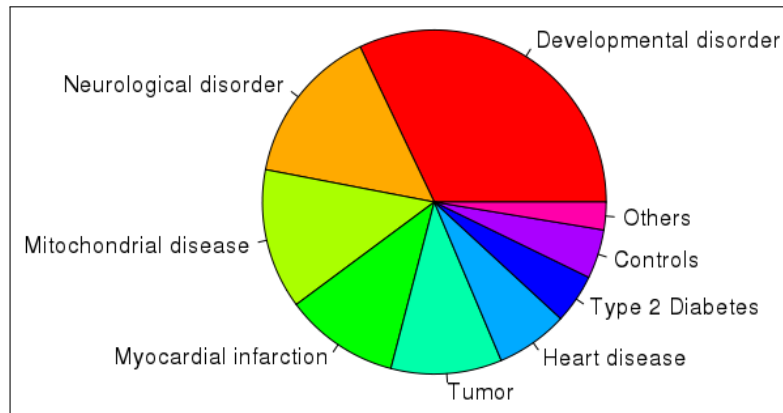


Abbildung 49: Übersicht zur Verteilung der untersuchten Erkrankungen innerhalb der NGS-Plattform am Institut für Humangenetik.

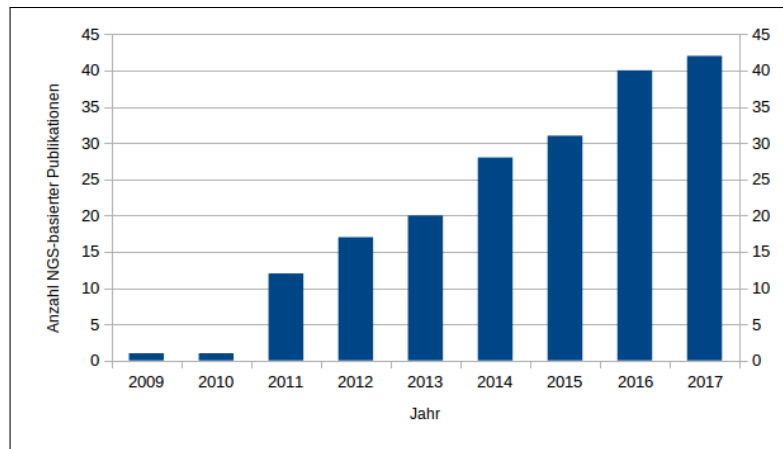


Abbildung 50: Anzahl NGS-basierter Publikationen aus dem Institut für Humangenetik pro Jahr. Bis November 2017 wurden mehr als 190 NGS-basierte Studien publiziert.

Die Entwicklung des LIMS sowie die Einführung der automatisierten *Library*-Präparation zusammen mit den Prozessoptimierungen ermöglichte eine erfolgreiche Anwendbarkeit der Exom-, Genom- und Transkriptom-Sequenzierung am Institut für Humangenetik. Dies zeigt sich an dem deutlichen Anstieg NGS-basierter Publikationen innerhalb der letzten Jahre (Abb.: 50).

3.3 Anwendung der *Next-Generation-Sequenzierung*

Der folgende Abschnitt beschreibt exemplarisch drei Anwendung der automatisierten *Library*-Präparation und des NGS-LIMS:

Im ersten Abschnitt wird beschrieben, wie die etablierte Automation genutzt wurde, um exomische *Libraries* aus FFPE-DNA herzustellen. Es wird erläutert welche Schwierigkeiten bei der Verwendung von FFPE-DNA innerhalb der *Library*-Präparation auftraten, sowie deren Lösung. Final wird gezeigt, dass qualitativ hochwertige und mit Normalgewebe vergleichbare Exom-Daten aus FFPE-DNA generiert werden konnten.

Der zweite Abschnitt beschreibt eine publizierte Studie zur Sequenzierung früher Exome aus dem Jahr 2011, die 2010 über die Semi-Automationsplattform hergestellt wurden. Es wird exemplarisch dargestellt, wie die Exom-Sequenzierung zur Erkennung von ursächlichen Varianten bei monogenen Erkrankungen genutzt werden kann. In der Studie konnte die ursächliche genetische Variante bei drei Familien mit Morbus Parkinson identifiziert werden.

Eine zweite publizierte Studie aus dem Jahr 2014 demonstriert die Limitationen der Exom-Sequenzierung. Im dritten Teil wird gezeigt, dass die Exom-Sequenzierung bei Patienten mit Burn-McKeown-Syndrom allein nicht ausreichte um die *compound*-heterozygote Veränderung zu identifizieren, die zum Phänotyp bei den Betroffenen führte. Mit einer zusätzlichen Sequenzierung der Genome der Patienten konnte die genetische Ursache der Erkrankung aufgeklärt werden. Die Studie repräsentiert eine Exom- und Genomuntersuchung, die mit Hilfe einer ausgereiften weitgehenden Automation im Jahr 2012 hergestellt wurden.

3.3.1 Anwendung der etablierten Automation zur Herstellung von Exom-*Libraries* aus FFPE-DNA

In dieser Studie sollte untersucht werden, ob DNA aus FFPE-Gewebe erfolgreich für eine Exom-*Library*-Präparation genutzt werden kann. Durch den Fixierungs- und Isolationsprozess ist FFPE-DNA anfällig für Degradation. Es sollte geklärt werden, wie sich das Material während der *Library*-Präparation verhält, vor allem im Hinblick auf die zu erwartende geringe Quantität und Qualität von FFPE-DNA. Konnte eine *Library* erfolgreich hergestellt werden, sollte untersucht werden wie sich diese, im Vergleich zu nicht-FFPE-Exomen, bezüglich der Datenqualität in der Sequenzierung verhält. Es sollten Qualitäts- und Quantitätsstandards der FFPE-DNAs für zukünftige Projekte definiert werden.

3.3.1.1 Hintergrund

Formalin-fixierte Paraffin-eingebettete (FFPE) Gewebe dienen zur Untersuchung von pathogenen Mutationen in Tumoren [129]. Die Identifikation der ursächlichen Mutation in Tumoren führt zum Verständnis der molekularen Grundlagen der Erkrankung, kann aber auch die Therapiemöglichkeiten beeinflussen. So wird bei Patienten mit metastasierenden kolorektalen Karzinomen der Mutationssatus des *KRAS*-Gens bestimmt. Aktivierende Mutationen in diesem Gen führen zu einer Resistenz gegen die Behandlung mit den EGFR-Antikörpern Cetuximab und Panitumumab [130].

Eine Fixierung von Tumormaterial mittels Formalin (4 %) gilt als Standardprozedur in der Pathologie und dient zur Beibehaltung der morphologischen Gewebe- und Zellstruktur [129]. Gleichzeitig erlaubt die Fixierung und Einbettung eine Lagerung bei Raumtemperatur über eine lange Zeit [129]. Durch den Fixierungs- und Einbettungsprozess kommt es zum Qualitätsverlust der DNA. Es treten Quervernetzungen und extensive Fragmentierungen der DNA auf. Abasische Stellen führen zur Denaturierung und Dopplestrangbrüchen der DNA-Helix. Desaminierungen von Cytosin zu Uracil erzeugen Mutationsartefakte (C>T) [131]. Die Veränderungen der DNA durch den Fixierungsprozess erschweren sowohl die Herstellung einer qualitativ hochwertigen exomischen *Library*, als auch die anschließende Datenauswertung. So wurde in vergangenen Studien gezeigt, dass die Anzahl von Varianten in FFPE-DNA höher ist als in gepaarten gefrorenen Geweben, was zu einer Erhöhung der falsch-positiv Rate führt [129].

3.3.1.2 Methodik und Ergebnisse

Die DNA wurde aus FFPE-Blöcken von Kopf-Hals-Karzinom Patientenbiopsien gewonnen. Die DNA-Extraktion fand mit dem *AllPrep DNA/RNA FFPE Kit* der Firma Qiagen durch den Kooperationspartner statt und wurde uns zur Verfügung gestellt.

3.3.1.2.1 Einfluss der DNA-Qualität auf die *Library*-Präparation

In einem ersten Versuch sollte untersucht werden, welche Qualität die FFPE-DNA aufweisen muss, um mit dem Standard-Protokoll (*SureSelect XT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing, Version E.3, November 2012*) eine *Library* zu generieren.

Die DNA wurde hinsichtlich ihrer Qualität auf dem Bioanalyzer mit dem *DNA 12000 Kit* überprüft. Die Quantität der DNA wurde mit dem *Qubit dsDNA Broad-Range Kit* ermittelt, während die Reinheit der Probe mit dem NanoDrop gemessen wurde (Tab.: 10 und Abb.: 51).

Die FFPE-DNA wurde anhand ihrer Fragmentlänge in die drei Qualitäten gut, mäßig und schlecht eingeteilt. Hochmolekulare genomische DNA hat eine Größe von mehr als 40 kbp. Eine qualitativ gute FFPE-DNA hat Fragmente größer als 10 kbp, eine mäßige Qualität liegt zwischen 500 bp und 1 kbp und eine schlechte Qualität zeichnet sich durch Fragmentlängen kleiner als 500 bp aus (Abb.: 51).

Tabelle 10: Übersicht der Parameter von drei ausgewählten FFPE-DNAs für eine exomische *Library*-Präparation.

Probenname	Integrität	Konzentration	A260/280 Verhältnis
Probe A	gut	169 ng/µl	1,9
Probe B	mäßig	72 ng/µl	1,85
Probe C	schlecht	132 ng/µl	1,9

Von Probe A, B und C wurde eine Exom-*Library* mit dem Standard-Protokoll hergestellt. Da häufig auch die Menge an FFPE-DNA eher gering ist, wurde anstatt mit 3000 ng Startmaterial mit 1500 ng gearbeitet. Als Positivkontrolle wurde eine genomische DNA mit 1500 ng und 3000 ng mitgeführt.

Ein erster Qualitätsschritt erfolgte nach der Fragmentierung der Proben auf dem Bioanalyzer. Die Fragmentierung mit dem Covaris E220 zielt auf eine Grö-

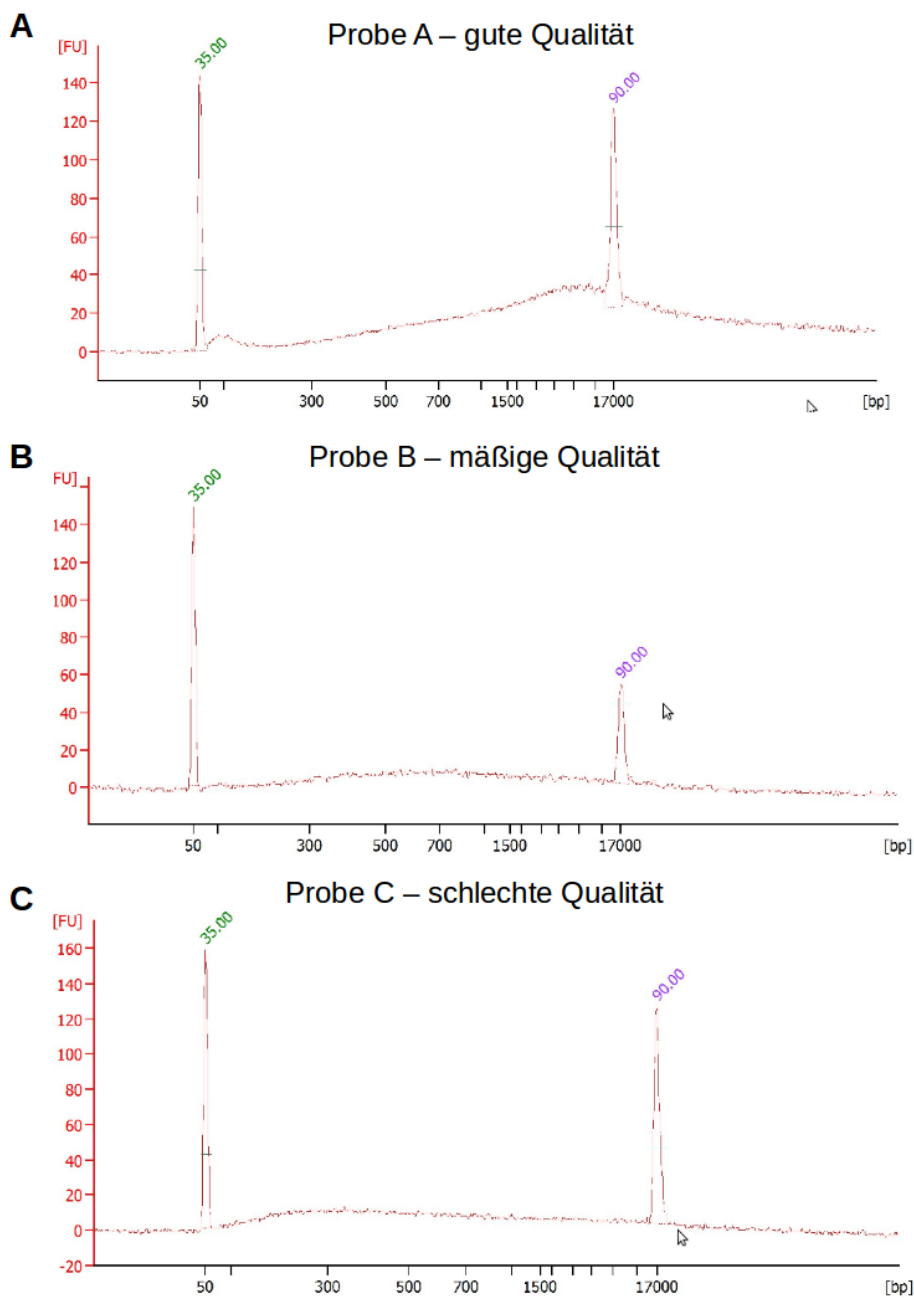


Abbildung 51: Elektropherogramme der unterschiedlichen FFPE-DNA-Qualitäten. Abbildung A zeigt das Elektropherogramm einer qualitativ guten FFPE-DNA mit mehr als 10 kbp Fragmenten. In Abbildung B ist eine qualitativ mäßige FFPE-DNA dargestellt mit einer durchschnittlichen Fragmentgröße um die 700 bp. Abbildung C zeigt eine qualitativ schlechte FFPE-DNA mit Fragmentgrößen kleiner als 500 bp.

ßenverteilung von 150 bis 500 bp mit einem Mittel von 250 bp ab. Die FFPE-DNAs zeigen eine normale Größenverteilung, weisen aber im Vergleich zur genomischen DNA eine um ein sechstel verringerte Quantität auf (Abb.: 52).

Der zweite Qualitätsschritt erfolgte vor der Hybridisierung der *Libraries* ebenfalls über den Bioanalyzer. Zu diesem Zeitpunkt ist die fragmentierte DNA an plattformspezifische Adapter ligiert und wurde wie im Protokoll beschrieben amplifiziert. Im Vergleich zur hochmolekularen DNA zeigen alle drei FFPE-Qualitäten eine verminderte *Library*-Quantität.

Für die folgende Exom-Anreicherung der *Libraries* werden 750 ng benötigt. Lediglich die FFPE-DNA mit guter Qualität erreicht diese Menge (870 ng). Die Mengen der qualitativ mäßigen und schlechten FFPE-DNAs sind nicht ausreichend (Abb.: 53).

Als Fazit wird abgeleitet, dass die Qualität von FFPE-DNAs vor einer *Library*-Präparation über den Bioanalyzer validiert und die Quantität mittels Qubit ermittelt werden muss. Die qualitative Einteilung erfolgt anhand der definierten Basenpaarverteilung. Von qualitativ guter FFPE-DNA werden 1500 ng in die *Library*-Präparation eingesetzt, während von qualitativ schlechter oder mäßiger FFPE-DNA die doppelte Menge genutzt werden muss.

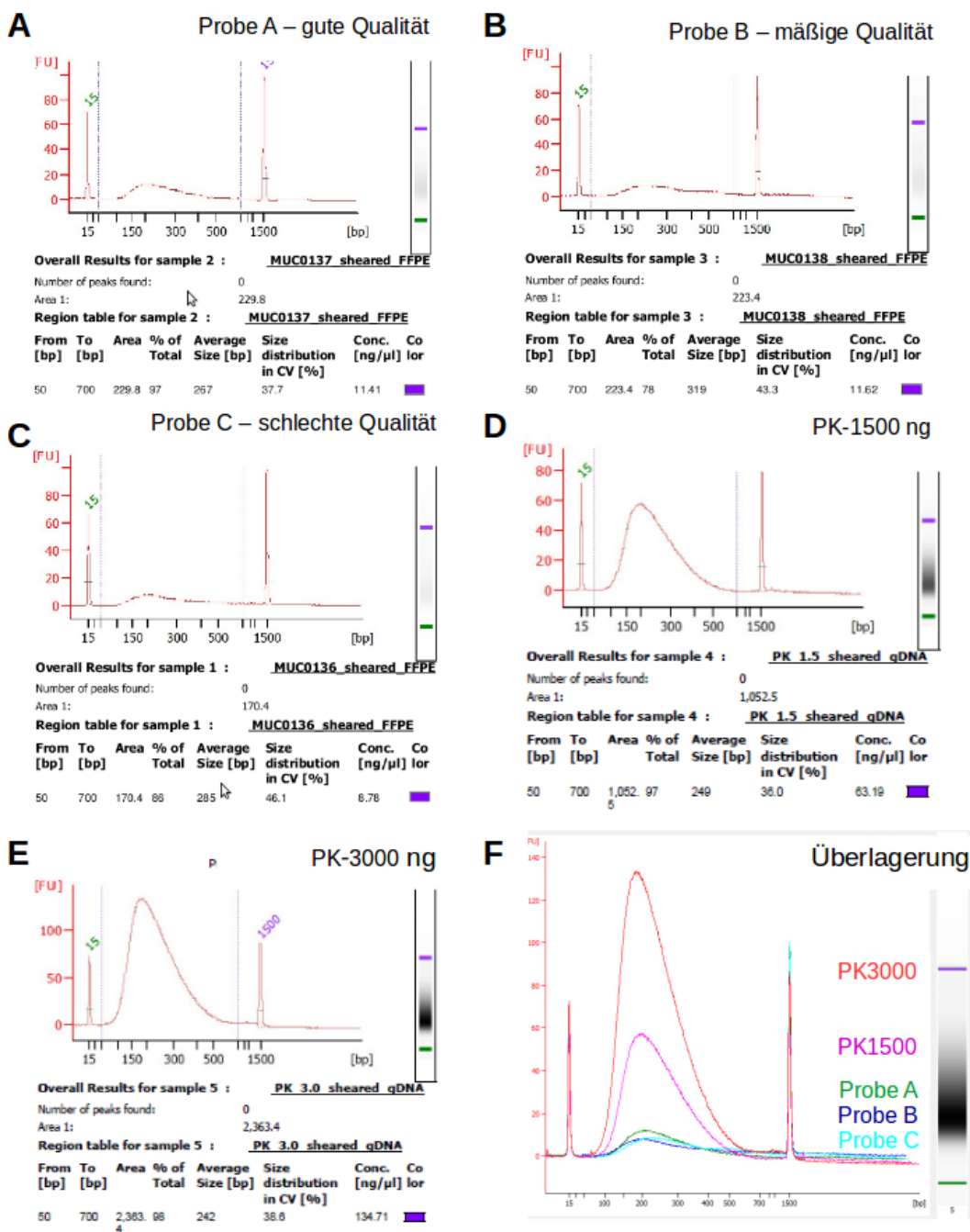


Abbildung 52: Elektropherogramme der Proben nach der Fragmentierung. Die Elektropherogramme A, B und C zeigen die Resultate der FFPE-DNA-Fragmentierung im Vergleich zu den Positivkontrollen in Abbildung D und E. Die Unterschiede der Quantität sind in der Überlagerung der Elektropherogramme in Abbildung F deutlich zu sehen.

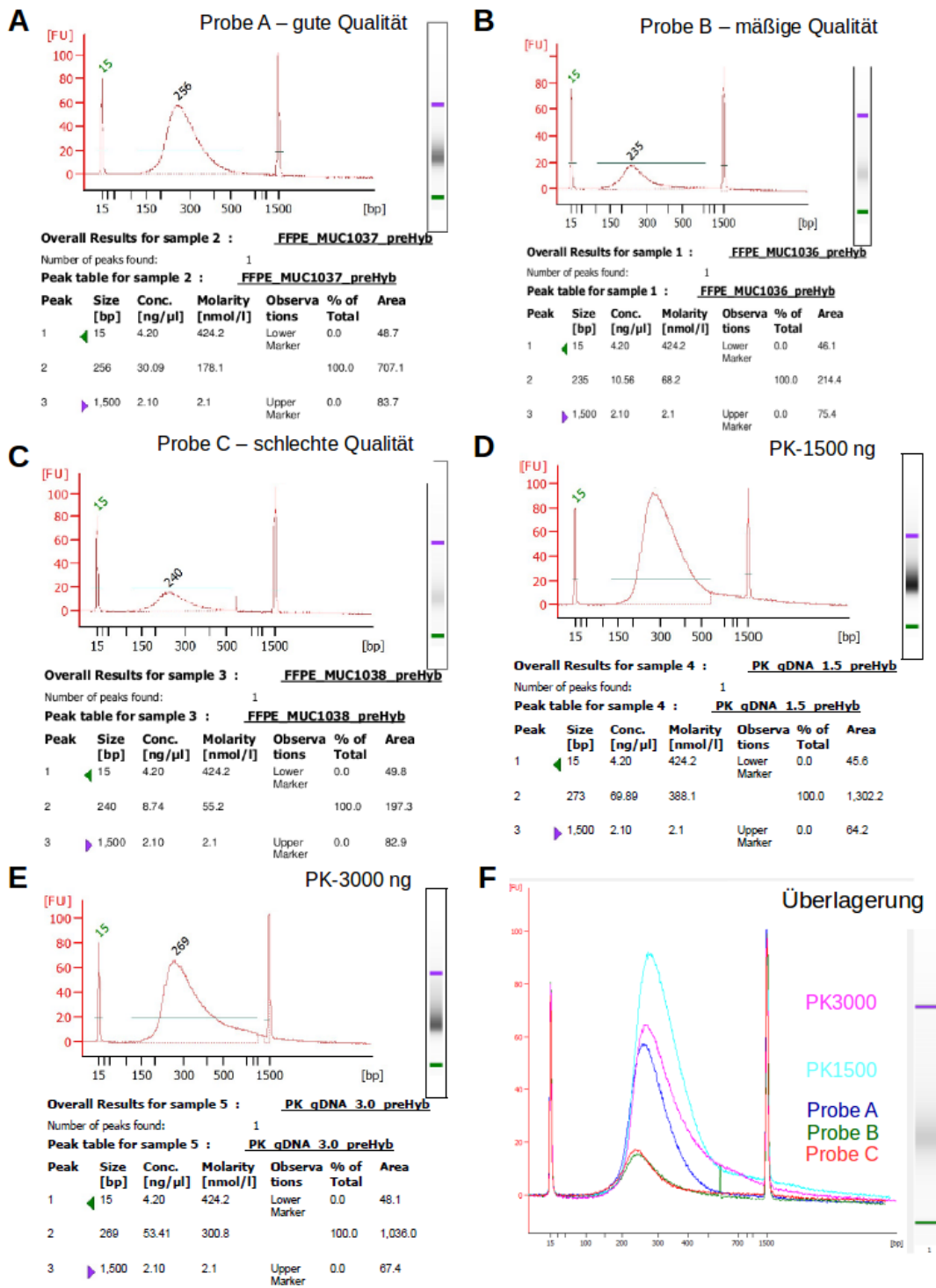


Abbildung 53: Elektropherogramme der *Libraries* vor der Hybridisierung. Die Elektropherogramme **A**, **B** und **C** zeigen die Resultate der *Library*-Präparation im Vergleich zu den Positivkontrollen in Abbildung **D** und **E**. Die Quantitätsunterschiede werden in Abbildung **F** deutlich. Lediglich die *Library* aus qualitativ guter FFPE-DNA erreicht eine ausreichende Menge für die Hybridisierung.

3.3.1.2.2 Sequenzierung von Exomen aus FFPE-DNA

Anhand der aufgestellten Kriterien aus dem ersten Abschnitt des Projektes wurden 61 FFPE-DNAs sequenziert. 25 davon wurden aus FFPE-Normalgewebe isoliert und 36 aus den gepaarten FFPE-Tumoren. Die DNAs wurden qualitativ über den Bioanalyser und quantitativ über den Qubit validiert. Für Material mit mäßiger und schlechter Qualität wurden 3000 ng angefordert, für DNAs mit guter Qualität 1500 ng. Von allen 61 FFPE-DNAs konnte eine exomische *Library* mit dem *SureSelect Human v6 Kit* hergestellt werden. Diese wurden als 100 bp *paired-end* Lauf auf dem HiSeq4000 sequenziert. Die erhaltenen Qualitätsparameter wurden mit zuvor sequenzierten Tumor- und Normalgewebe-Exomdaten verglichen (Tab.: 11). Aufgrund der Vorfragmentierung von FFPE-DNA war eine Erhöhung der Duplikatrate sowie eine verringerte durchschnittliche Fragmentgröße zu erwarten. Dies zeigt sich auch in den Ergebnissen. Daneben gibt es keine stark abweichenden Parameter im Vergleich zu nicht FFPE-DNA. Die Anzahl der zuordenbaren *Reads* auf dem Exom und die durchschnittliche Abdeckung des Exoms sind innerhalb der Gruppen vergleichbar. Lediglich die Anzahl der 20-fach abgedeckten Bereiche ist geringer, aber innerhalb der Spezifikationen (>90 %) für ein qualitativ hochwertiges Exom. Die vorliegenden Ergebnisse zeigen, dass es prinzipiell möglich ist eine exomische *Library* aus FFPE-DNA zu erzeugen, die qualitativ vergleichbar ist mit Exomen aus nicht-FFPE-DNA.

Die eigentliche Analyse der Varianten war nicht Bestandteil der vorliegenden Fragestellung, wird aber derzeit durch die bioinformatische Abteilung noch ausgeführt. Dabei ist vor allem interessant, ob es im Vergleich zu Exomen aus nicht-FFPE-Gewebe, zu mehr falsch-positiven Varianten durch Desaminierungen kommt.

Tabelle 11: Übersicht zur Sequenzier-Statistik von FFPE-DNA versus nicht-FFPE-DNA von Exomen aus Tumor- und Normalgewebe-Proben.

Parameter	FFPE-DNA		FFPE-DNA	
	Normalgewebe n=25	Normalgewebe n=100	Tumor n=36	Tumor n=100
<i>Mapped Reads</i> [%]	99,9	99,7	99,9	99,7
Daten [Gb]	12,7	14,2	13,4	14,9
<i>On-Bait</i> [%]	79,9	77,8	79,2	78,7
Ø Abdeckung	142	161	148	171
20-fache Abdeckung [%]	94,7	98,7	97,6	98,7
Ø Fragmentgröße [bp]	166	206	171	201
Duplikatrate	14	9	14	9
Ø Anzahl der SNVs	77.923	79.799	77.619	78.467

3.3.2 Exom-Sequenzierung zur Identifizierung pathogener genetischer Varianten bei Patienten mit spät beginnendem Morbus Parkinson

Dieser Abschnitt der vorliegenden Arbeit wurde zuvor publiziert als:

A mutation in *VPS35*, encoding a subunit of the retromer complex, causes late-onset Parkinson disease.

Zimprich A*, Benet-Pagès A*, Struhal W*, **Graf E***, Eck SH, Offman MN, Haubenberger D, Spielberger S, Schulte EC, Lichtner P, Rossle SC, Klopp N, Wolf E, Seppi K, Pirker W, Presslauer S, Mollenhauer B, Katzenschlager R, Foki T, Hotzy C, Reinthaler E, Harutyunyan A, Kralovics R, Peters A, Zimprich F, Brücke T, Poewe W, Auff E, Trenkwalder C, Rost B, Ransmayr G, Winkelmann J, Meitinger T, Strom TM.

Am J Hum Genet. 2011 Jul 15;89(1):168-75.

*gleicher Beitrag zur vorliegenden Arbeit

Ich führte bei dieser Studie den Hauptteil der experimentellen Arbeit durch. Hierzu gehörten die Herstellung der Exom-*Libraries*, die Sequenzierung auf dem *Genome Analyzer IIx* sowie die Datenauswertung zusammen mit Tim Strom, Alexander Zimprich und Sebastian Eck. Die identifizierten Varianten bestätigte ich mittels Sanger-Sequenzierung. Für die Mutationsanalyse entwarf ich Primer, führte die Schmelzkurvenanalyse (IDAHO) der Patientengruppe und der Kontrollgruppe durch und wertete sie aus. Die in der Schmelzkurvenanalyse auffälligen Proben sequenzierte ich mittels Sanger-Sequenzierung. Ich bereitete die Patientenproben für eine Genotypisierung vor und trug zur Erstellung des Manuskripts bei. Die Untersuchungen zur Moleküldynamik wurden von den Kooperationspartnern durchgeführt.

3.3.2.1 Zusammenfassung

Zur Identifizierung seltener ursächlicher Genvarianten bei spät-beginnendem Morbus Parkinson (M. Parkinson) wurde eine österreichische Familie mit 16 Betroffenen via Exom-Sequenzierung untersucht. Wir identifizierten eine heterozygote *missense*-Variante c.1858G>A, p.Asp620Asn im Gen *VPS35*. Sieben Betroffenen trugen diese genetische Alteration. Wir identifizierten zwei weitere Familien, bei denen die Punktmutation im autosomal-dominanten Vererbungsmodus mit inkompletter Penetranz kosegregiert. Alle Betroffenen litten unter einer sich spät manifestierenden Form des Morbus Parkinson mit einem mittleren Erkrankungsbeginn von 53 Jahren. Eine Kopplungsanalyse in den drei Familien zeigte einen gemeinsamen Haplotyp von 65 kb, der das *VPS35*-Gen überspannt. Eine Untersuchung von weiteren 860 Betroffenen und 1040 Kontrollen ergab sechs weitere nicht-synonyme *VPS35 missense*-Varianten. Drei dieser Varianten fanden wir nur in Patienten, zwei in Kontrollen und eine sowohl in einem Betroffenen als auch in einer Kontrollperson. Computergestützte sequenzbasierte Analysen, sowie Untersuchungen zur Moleküldynamik, prognostizieren die familiäre Variante p.Asp620Asn und die Variante p.Arg524Trp als schädlich. *VPS35* ist ein Bestandteil des Retromer Komplexes, der den retrograden Transport zwischen Endosomen und dem trans-Golgi Netzwerk bewerkstelligt. Zusätzlich wurde *VPS35* als ein möglicher Faktor bei der Entstehung von Alzheimer diskutiert [132].

3.3.2.2 Hintergrund

Morbus Parkinson (MIM168600) ist neben Morbus Alzheimer die zweit-häufigste neurodegenerative Erkrankung und tritt weltweit mit einer Prävalenz von 1 bis 2 % bei den über 60-Jährigen auf [133]. M. Parkinson ist eine unheilbare, langsam progredierende Erkrankung, die durch eine Störung der motorischen Abläufe charakterisiert ist. Als Kardinalssymptome gelten: Bradykinesie (verlangsamter Bewegungsablauf), Rigor (veränderter Muskeltonus führt zur Versteifung der Gliedmaßen), Ruhetremor und posturale Instabilität (Haltungsinstabilität). Neben der typischen motorischen Dysfunktion können autonome Dysfunktionen (Inkontinenz, orthostatische Intoleranz), neuropsychiatrische Dysfunktionen (Depressionen, Halluzinationen), sowie Schmerzempfindungen und sensorische Störungen auftreten. Ursächlich für die Symptomatik ist die Degeneration dopaminerger Neuronen in der *Substantia nigra pars compacta*, die zu einer Dysfunktion

Tabelle 12: Morbus Parkinson assoziierte Gene und genetische Loci.

Symbol	Gen	MIM	Locus	Erbgang	Klinik	Studie	Referenz
<i>PARK1/4</i>	<i>SCNA</i>	163890	4q22.1	AD	early-onset	KA	[135]
<i>PARK2</i>	<i>Parkin</i>	602544	6q26	AR	early-onset	KA	[136]
<i>PARK3</i>	<i>SZBL</i>	602404	2p13	unbek.	klassischer MP	KA	[137]
<i>PARK5</i>	<i>UCHL1</i>	191342	4p13	AD	klassischer MP	KGA	[138]
<i>PARK6</i>	<i>PINK1</i>	608309	1p36.12	AR	early-onset	KA	[139]
<i>PARK7</i>	<i>DJ1</i>	602533	1p36.23	AR	early-onse	KA	[140]
<i>PARK8</i>	<i>LRRK2</i>	609007	12q12	AD	klassischer MP	KA	[141, 142]
<i>PARK9</i>	<i>ATP13A2</i>	610513	1p36.13	AR	Kufor-Rakeb Syndr.	KA	[143]
<i>PARK10</i>	<i>SZBL</i>	606852	1p32	unbek.	klassischer MP	KA	[144]
<i>PARK11</i>	<i>GIGYF2</i>	612003	2q37.1	AD	klassischer MP	KA	[145]
<i>PARK12</i>	<i>SZBL</i>	300557	Xq21-q25	unb.	klassischer MP	KA	[146]
<i>PARK13</i>	<i>HTRA2</i>	606441	2p13.1	AD	klassischer MP	KGA	[147]
<i>PARK14</i>	<i>PLA2G6</i>	603604	22q13.1	AR	klassischer MP	KA	[148]
<i>PARK15</i>	<i>FBX07</i>	605648	22q12.3	AR	early-onset	KA	[149]
<i>PARK16</i>	<i>SZBL</i>	613164	1q32	unbek.	MP	GWAS	[150]
<i>PARK17</i>	<i>VPS35</i>	601501	16q11.2	AD	late-onset	WES	[69, 151]
<i>PARK18</i>	<i>EIF4G1</i>	600495	3q27.1	AD	late-onset	KA	[152]
<i>PARK19</i>	<i>DNAJC6</i>	608375	1p31.3	AR	juvenile-onset	WES	[153]
<i>PARK20</i>	<i>SYNJ1</i>	604297	21q22.11	AR	early-onset	WES	[154, 155]
<i>PARK21</i>	<i>DNAJC13</i>	616361	3q22.1	AD	klassischer MP	KA	[156, 157]
<i>PARK22</i>	<i>CHCHD2</i>	616244	7p11.2	AD	late-onset	WES, WGS, KA	[150, 158]
<i>PARK23</i>	<i>VPS13C</i>	608879	15q22.2	AR	early-onset	WES	[159]

SZBL, Suszeptibilitätslocus; AD, autosomal dominant; AR, autosomal rezessiv; KA, Kopplungsanalyse; KGA, Kandidatengenanalyse; GWAS, Genomweite Assoziationsstudie; WES, whole exome sequencing; WGS, whole genome sequencing.

der Basalganglien und folgend zu einer Fehlfunktion der Motorik des Patienten führt. Auf Grundlage von mehreren europäischen Studien wird davon ausgegangen, dass sich die Erkrankungsprävalenz bis zum Jahr 2030 in der Population der über 50-jährigen verdoppelt [134]. Die dramatische Steigerung der Prävalenz der Erkrankung, in der immer älter werdenden Population, zeigt die Dringlichkeit nach der Suche zur genetischen Ursache, sowie Therapiemöglichkeiten für Morbus Parkinson. Ein Teil der genetischen Ursachen konnte in verschiedenen Studien aufgedeckt werden. Zu diesem Zeitpunkt sind neben 4 Suszeptibilitätsloci, Veränderungen in neun verschiedene Genen bekannt, die zu einer autosomal-dominante Form der Erkrankung führen, sowie Veränderungen in neun Genen die eine autosomal-rezessive Form des M. Parkinson hervorrufen (Tab.: 12). Hier zeigt sich deutlich die enorme genetische Heterogenität der Erkrankung.

5 bis 10 % der beschriebenen Morbus Parkinson Fälle sind familiär [160]. Die Seltenheit großer Stammbäume mit klarer mendelscher Vererbung erschwert die Identifikation der genetischen Ursache bei familiärem Parkinson.

3.3.2.3 Patienten

In der vorliegenden Studie wurde eine österreichische Familie mit 16 betroffenen Familienmitgliedern via Exom-Sequenzierung untersucht. Der Stammbaum der Familie weist auf eine autosomal-dominante Vererbungsform und eine hohe Krankheitspenetranz hin (Abb. 54A). Sieben der 16 Betroffenen konnten klinisch untersucht werden und stellten ihre DNA zu molekulargenetischen Analysen zur Verfügung. Die klinische Untersuchung durch Neurologen nach den Kriterien der *UK Brain Bank* zeigte, dass bei sechs der sieben Patienten mindestens drei von vier Kardinalsymptomen vorliegen. Zusätzlich zeigten die Betroffenen eine Verbesserung des Gesundheitszustandes nach Behandlung mit Levodopa (L-3,4-Dihydroxyphenylalanin). Ein Patient zeigt einen Bewegungstremor seit seiner Kindheit und entwickelte erst im Alter von 62 Jahren einen Ruhetremor und Akinesie (Bewegungslosigkeit). Der Krankheitsbeginn in der Familie liegt zwischen 40 und 68 Jahren mit einem mittleren Alter von 53 Jahren. Zusätzlich untersuchten wir 486 österreichische Patienten mit einem mittleren Alter von 60 Jahren bei Krankheitsbeginn. Bei 131 von ihnen konnte eine positive Familienanamnese, mit mindestens einem betroffenen Verwandten ersten oder zweiten Grades, festgestellt werden. Weiterhin untersuchten wir 376 deutsche M. Parkinson Patienten aus einer Klinik speziell für Bewegungsstörungen (Paracelsus-Elena Klinik). Als Kontrollpopulation wählten wir 680 unabhängige Probanden der KORA-S4 Studie, 1041 Probanden der KORA-AGE Studie und 554 österreichische gesunde Individuen.

3.3.2.4 Methodik und Ergebnisse

In dieser Studie wurden die Exome von zwei betroffenen Familienmitgliedern untersucht. Die Varianten wurden auf gemeinsam auftretende Veränderungen hin priorisiert. In der Annahme, dass eine gemeinsame seltene Variante krankheitsverursachend ist, reduziert die Analyse von DNA entfernt Verwandter die Anzahl an gemeinsamen Varianten, die sie sich durch ihre Abstammung teilen. Die Voraussetzung für solch eine Analyse ist ein möglichst großer Stammbaum, aus dem zwei entfernt Verwandte analysiert werden. Zur Identifizierung der kausativen genetischen Ursache des M. Parkinson in der Familie A, wählten wir zwei Cousins zweiten Grades (#3017 und #3020) für die Exom-Sequenzierung aus (Abb. 54 A). Die Exom-Library wurde nach Anreicherung mit dem *SureSelect Human All*

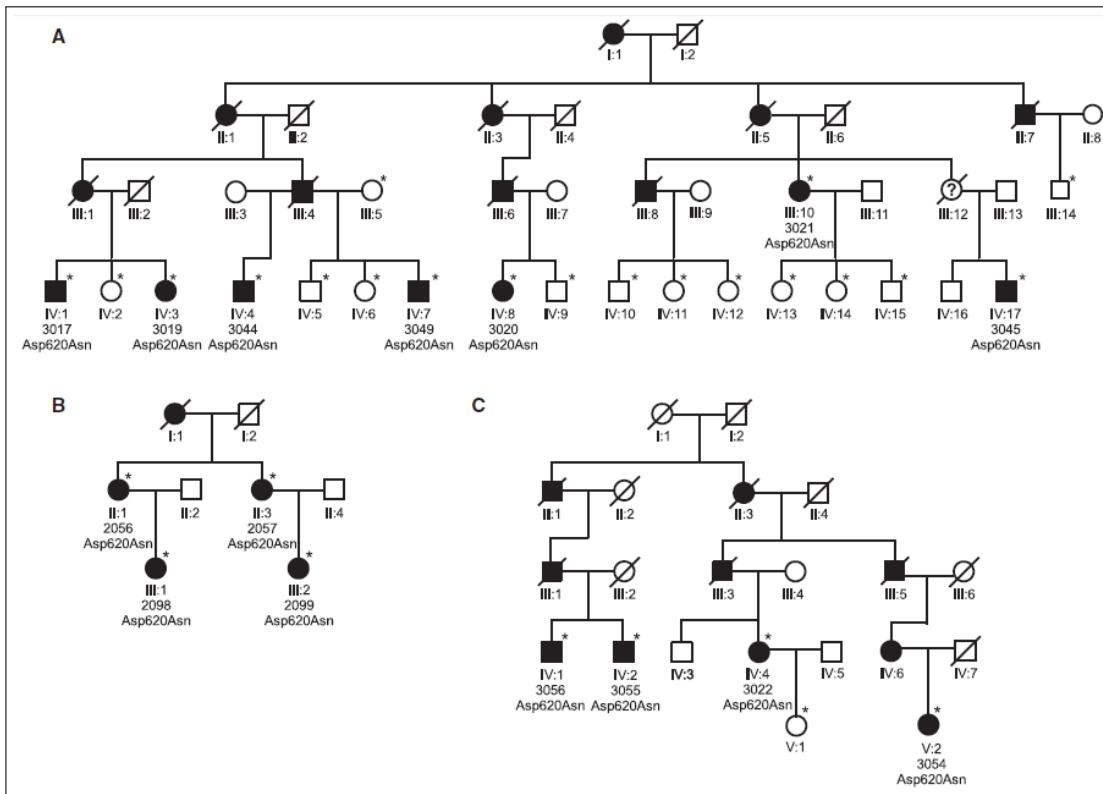


Abbildung 54: Stammbäume der Familien A, B und C [69]. Familie A ist die initial, durch Exom-Sequenzierung, untersuchte Familie. Das Familienmitglied III:12 der Familie A hat einen unbekanntem Phänotyp. Alle drei Familien zeigen eine klare autosomal-dominante Form der Vererbung. * kennzeichnen die p.Asp620Asn genotypisierten Personen. Durchgestrichene Symbole kennzeichnen verstorbene Personen.

Exon 38 Mb Kit, als 54 bp *paired end*-Lauf auf einem *Genome Analyzer Iix* durchgeführt. Wir sequenzierten jeweils 2 *Lanes* einer *Flow Cell* pro Patient (Tab. 13). Das *read alignment* wurde mit BWA (Version 0.5.8) zum Referenzgenom hg19 durchgeführt. Punktmutationen sowie kleine Insertionen und Deletionen wurden mit SAMtools (v 0.1.7) detektiert. Insgesamt wurden initial rund 28.000 Varianten pro Patient identifiziert. Durch die Anwendung weiterer Kriterien, wie der Qualität der Basen (>15) und die Repräsentation der Varianten in mindestens 15 % der Reads, konnten 4.000 Alterationen ausgeschlossen werden. Es fanden sich circa 15.500 nicht kodierende Varianten und circa 7.250 nicht-synonyme Veränderungen pro Patient.

Zur Annotation der Varianten wurden benutzerdefinierte Perl Skripte mit der Integration der Daten von dbSNP (v131) und dem UCSC *Genome Brow-*

Tabelle 13: Statistik der Sequenzierung

Patient	Mapped Reads	[Gb]	On-Bait- Rate [%]	Abdeckung >1X [%]	Abdeckung >8X [%]	Abdeckung >20X [%]	Ø Abdeckung
#3017	119.137.369	6.4	46.19	99.11	92.48	80.02	65.61
#3020	112.524.559	6.1	44.48	98.84	91.94	78	59.31

Ergebnisse der Sequenzierung auf dem *Genome Analyzer IIx*.

ser known Gene Track genutzt. Die Annotation der genetischen Veränderungen umfasste die Präsenz in dbSNP, die Art der Mutation und, wenn vorhanden, der vorhergesagte Aminosäure-Austausch im jeweiligen Protein. Unter der Annahme, dass eine seltene Variante krankheitsverursachend ist, schlossen wir in einer zweiten Filterung des Datensets alle Varianten, die auch in 72 Kontroll-Exomen, bei Patienten mit unterschiedlichen Erkrankungen, auftraten aus. Zusätzlich wurden alle Varianten ausgeschlossen, die im Datensatz von dbSNP 131 mit einem durchschnittlichen Heterozygotiegrad von mehr als 0.02 auftraten. Die Annotation der verbliebenen Varianten wurde mit eigenen Skripten durchgeführt. Final identifizierten wir 10 seltene, heterozygote, nicht-synonyme Varianten, die in beiden Patienten gemeinsam auftreten (Tab.: 14).

Zur weiteren Priorisierung wurde die Häufigkeit der zehn Varianten in 670 KORA-S4 Probanden mittels MALDI-TOF Massenspektrometrie untersucht. Nur die heterozygote Variante c.1858G>A (p.Asp620Asn) in *VPS35* konnte in keinem der KORA-S4 Probanden nachgewiesen werden und kosegregiert mit der Erkrankung in allen betroffenen Familienmitgliedern.

Wir untersuchten zusätzlich 486 österreichische M. Parkinson Patienten auf die p.Asp620Asn Variante und identifizierten zwei zusätzliche Indexpatienten, der Familien B und C, mit dieser Mutation (Abb. 54 B und C). In beiden Familien wurde die Variante in allen untersuchten Betroffenen gefunden (n=8). Die *VPS35*-Variante c.1858G>A konnte weder in 554 österreichischen noch in 1014 zusätzlichen KORA-AGE Probanden gefunden werden. Drei, bisher nicht betroffene Familienmitglieder der Familien A, B und C, tragen ebenfalls die genetische Veränderung. Allerdings sind diese Angehörigen zum Zeitpunkt der Studie jünger als 60 Jahre, so dass entweder von einem präsymptomatischen Zustand auszugehen ist, oder die Variante eine nicht vollständige Penetranz zeigt.

Die Aminosäure Asp620 in *VPS35*, liegt innerhalb der Spezies der Pflanzen, Pilze, Invertebraten und Vertebraten hoch konserviert vor. Die Mutation wird

Tabelle 14: Identifizierte Varianten nach der ersten Priorisierung des Datensets [69].

Gen	Position (hg19)	Transkript NM_	Nukleotid Varianten	Aminosäure	Kontroll- Genotypen			Ko- segreg.
					1/1	1/2	2/2	
<i>PLK3</i>	chr1:45270359	004073.2	c.1543T>A	p.Ser515Thr	669	0	0	4 von 7
<i>C8A</i>	chr1:57383357	000562.2	c.1723C>T	p.Pro575Ser				5 von 7
<i>ADCY10</i>	chr1:167787479	018417.4	c.4313A>G	p.Asn1438Ser				2 von 7
<i>LAMB2</i>	chr3:49166460	002292.3	c.1724C>T	p.Arg575Gln	647	28	0	5 von 7
<i>NOM1</i>	chr7:156762317	138400.1	c.2503G>A	p.Ala835Thr	670	0	0	3 von 7
<i>KIF22</i>	chr16:29816237	007317.1	c.1780G>A	p.Asp594Asn	665	6	0	6 von 7
<i>SEZ6L2</i>	chr16:29899021	012410.2	c.947C>T	p.Arg316His	660	4	0	7 von 7
<i>VPS35</i>	chr16:46696364	018206.4	c.1858G>A	p.Asp620Asn	1069*	0	0	7 von 7
<i>NLRP1</i>	chr17:5421150	001033053.2	c.3985G>A	p.Val1329Ile	666	4	0	3 von 7
<i>NEURL4</i>	chr17:7221197	001005408.1	c.4109C>T	p.Arg1370Gln				3 von 7

Alle identifizierten, seltenen genomischen Varianten, wurden in 670 KORA-S4 Kontrollen mittels MALDI-TOF Analyse untersucht. Das variante Allel ist als 2 gekennzeichnet, das Referenzallel als 1. * Die angegebene Anzahl an Kontrollgenotypen beinhaltet zusätzlich 554 österreichische gesunde Individuen, die mittels TaqMan Analyse genotypisiert wurden. Die Kosegregation zeigt die Anzahl der Betroffenen der Familie A, welche die p.Asp620Asn Veränderung tragen. Die Varianten der Gene *C8A* und *ADCY10*, sind als Polymorphismen in dbSNP hinterlegt (rs41285938 und rs41270737).

durch PolyPhen2, SIFT und SNAP als schädlich eingestuft. Durch die Summe dieser Ergebnisse kann geschlussfolgert werden, dass die Mutation c.1858G>A in *VPS35* ursächlich für M. Parkinson in den Familien A, B und C ist.

Zur Detektion eines potentiell gemeinsamen Haplotyps, wurden 20 Individuen der Familien A, B und C genotypisiert. Es zeigt sich, dass sich die Familien A und B eine gemeinsame Region von 21 Mb teilen, während die Familien A, B und C nur 65 kb, überlappend mit *VPS35*, gemeinsam haben. Es konnte keine Aussage darüber getroffen werden, ob die drei Familien einen gemeinsamen alten Haplotyp teilen und die Mutation auf einem Gründereffekt beruht, oder ob sie kürzlich durch zwei unabhängige Vorgänge entstanden ist.

Um die Prävalenz von *VPS35*-Varianten in der generellen Population zu bestimmen, untersuchten wir mittels hochauflösender Schmelzkurvenanalyse (IDAHO) alle 17 kodierenden Exons auf genetische Veränderungen bei 484 österreichischen sowie 376 deutschen M. Parkinson Patienten. Als Kontrollpopulation wurden 1014 Probanden der KORA-AGE Studie mit einem mittleren Alter von 76 Jahren gewählt. Die Untersuchung ergab neben sieben synonymen Punktmutationen und 18 Varianten im nicht-kodierenden Bereich, zusätzlich zu p.Asp620Asn weitere 6 seltene kodierende Varianten (Tab.: 15). Zusammenfassend identifizierten wurden vier nicht-synonyme missense-Varianten, die ausschließlich in Patienten zu finden sind, zwei Varianten, die nur in Kontrollen auftreten und eine

Tabelle 15: Seltene, nicht-synonyme *VPS35*-Varianten in Patienten und Kontrollen [69].

Patienten	KORA AGE	Heterozygoter Nukleotid-austausch	Aminosäure-austausch	Einfluss auf das Protein			Exon	Position hg19 Chr.:16	KORA-S4		
				(a)	(b)	(c)			1/1	1/2	2/2
-	1	c.151G>A	p.Gly51Ser	+	+	+	3	46,716,039			
90/05	-	c.171G>A	p.Met57Ile	+	+	+	3	46,716,019	670	0	0
-	1	c.245C>G	p.Thr82Arg	+	+	+	4	46,715,367			
806	-	c.723T>G	p.Ile241Met	+/-	+	+	7	46,711,308	667	0	0
(211)	-	c.1570C>T	p.Arg524Trp	-	-	-	13	46,702,919	671	0	0
(Fam. A-C)	-	c.1858G>A	p.Asp620Asn	-	-	-	15	46,696,364	669	0	0
243, 524	2	c.2320C>A	p.Leu774Met	+	+	+	7	46,694,455			

Die Varianten für 863 Patienten und 1014 KORA-AGE Kontrollen wurden durch hochauflösende DNA-Schmelzkurvenanalyse (IDAH) bestimmt und mittels Sanger-Sequenzierung bestätigt. Die identifizierten Varianten untersuchten wir zusätzlich in 680 KORA-S4 Kontrollen mittels MALDI-TOF. Für die KORA-S4 Kontrollen ist das variante Allel mit 2 gekennzeichnet, während das Referenzallel als 1 bezeichnet wird. Die familiären Fälle sind mit Klammern gekennzeichnet. Die Prediktionsprogramme (a) PolyPhen2, (b) SNAP und (c) SIFT prognostizieren den Einfluss einer genetischen Variante als: + für benign, +/- für möglicherweise schädlich und - für wahrscheinlich schädlich.

Variante, die sowohl in der Kontrollpopulation als auch in M. Parkinson Patienten auftrat, identifiziert. Zusätzlich zu p.Asp620Asn wurde eine zweite Variante, p.Arg524Trp, als schädlich durch PolyPhen2, SIFT und SNAP eingestuft. Die Veränderung p.Ile241Met wurde lediglich durch PolyPhen2 als schädlich eingestuft, während alle weiteren Varianten als benign klassifiziert werden.

Klinische Informationen der Familienmitglieder gab es nur bei dem Patienten 211 (p.Arg524Trp). Die 74 jährige Mutter, ebenfalls Trägerin der Variante, zeigt schwache extrapyramidale Symptome wie leichter, posturaler Tremor der oberen Extremitäten. Sie zeigt keine Bradykinesie und eine DAT-SPECT Untersuchung zeigte keine Reduktion der striatalen Bindung, was die Möglichkeit eines frühes Stadiums des M. Parkinson ausschließt.

Überdies hinaus konnten bei der Untersuchung der Patienten und Probanden keine häufigen, nicht-synonymen Varianten in *VPS35* identifiziert werden. Weiterhin waren zum Zeitpunkt der Studie keine häufigen, nicht-synonymen Varianten in 72 Kontroll-Exomen, sowie in dbSNP131 gelistet.

Die Kristallstruktur des C-terminalen Bereichs von *VPS35* wurde durch Hierro *et al.* aufgeklärt [161]. Die drei nicht-synonymen Varianten p.Asp620Asn, p.Arg524Trp und p.Leu774Met, sind alle in diesem Bereich lokalisiert. Mit Hilfe computergestützter Analysen zur Moleküldynamik, wurde der Einfluss der Varianten auf die Proteinstabilität untersucht.

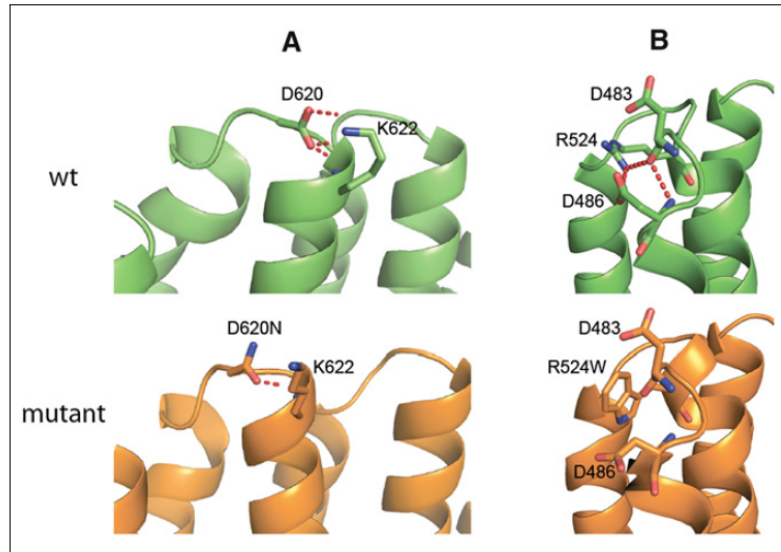


Abbildung 55: Analyse der Moleküldynamik im wildtyp- und mutanten-VPS35-C-terminalen Bereich [69]. Nach dem manuellen Einbau der Alterationen in die Kristallstruktur nutzten wir die Software scwrl 4.0 [162] zur Modellierung der Seitenketten. Alle Simulationen zur Moleküldynamik führten wir mit GROMACS 4.5 [163] und dem Atomkraftfeld AMBER03 [164] sowie dem Wassermodell TIP3P [165] durch. Die Wasserstoffbrückenbindungen, die sich zwischen den einzelnen Aminosäuren ausbilden sind als rote gestrichelte Linien dargestellt. (A) Asp620 bildet eine Wasserstoffbrücke zu Lys622 zusammen mit einer zusätzlichen Ionenbindung (Salzbrücke), die durch den Austausch von Asp 620 zu Asn verloren gehen. (B) Arg524 bildet mehrere Wasserstoffbrückenbindungen zu den Aminosäuren Asp483 und Asp486. Wird Arg524 gegen Trp ausgetauscht gehen diese Bindungen verloren.

Die drei betroffenen Aminosäuren befinden sich am Ende einer Helix, die mit VPS29 interagiert. Der Asp620 Wildtyp bildet regelmäßige Wasserstoffbrückenbindungen zu Lys622 aus, die weniger stark frequentiert im varianten Protein vorliegen (Abb. 55 A). Ähnlich dazu verhält sich der Arg524-Wildtyp, der drei Wasserstoffbrückenbindungen zu den Aminosäuren Asp483 und Asp486 bildet. Diese Verbindung entsteht nicht im mutanten Protein (Abb. 55 B).

Die Veränderungen p.Asp620Asn und p.Arg524Lys führen zum Verlust von Salzbrücken (Ionenbindungen) und bewirken eine lokale Instabilität von VPS35, wie Bewegungsanalysen (*Root Mean Square Fluctuation Profiling*) des Moleküls zeigen (Abb. 56). Im Gegensatz dazu hat die p.Leu774Met Variante kaum Einfluss auf die Stabilität von VPS35.

In dieser Studie identifizierten wir die Variante c.1858G>A (p.Asp620Asn) im Gen *VPS35*. Sie erfüllt die Kriterien für eine mögliche Pathogenität: sie kosegregiert mit der Erkrankung in drei Familien und war nicht detektierbar in 1694

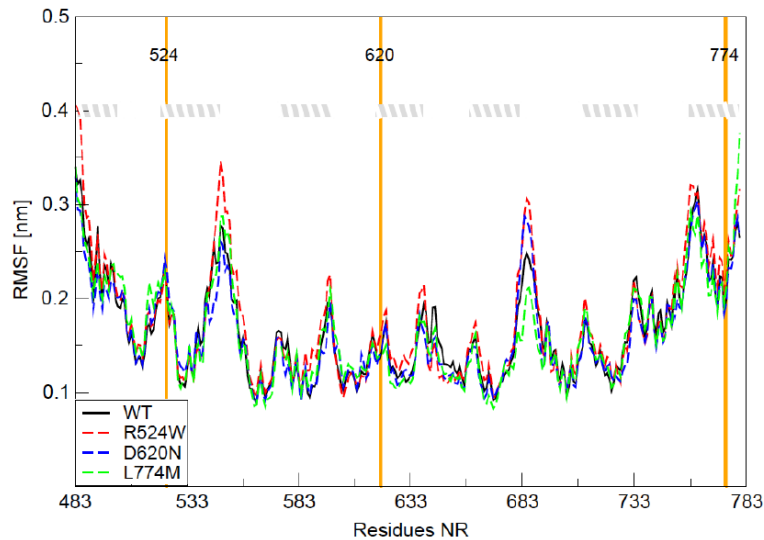


Abbildung 56: *Root Mean Square Fluctuation* (RMSF) Profil [69]. Vergleichende Untersuchungen zur Moleküldynamik im wildtyp-Protein sowie in den aberranten Proteinen. Die Molekülbewegung im wildtyp Protein ist als schwarze Linie dargestellt, während die Varianten p.Asp620Asn in blau, Leu774Met in grün und p.Arg542Trp in rot zu sehen ist. An den gelben vertikalen Linien befinden sich die Positionen der Varianten im Protein. Die graue Schraffierung kennzeichnet die Interaktionsdomänen zu VPS29. Abweichende Flexibilitäts-Peaks für beide Aminosäureveränderungen finden sich an den Positionen 483, 548, 639 und 685.

Kontrollen. Die Aminosäure Asp620 in VPS35 ist innerhalb verschiedener Spezies evolutionär hoch konserviert und der Austausch der Aminosäure wird durch die Prediktionsprogramme PolyPhen2, SNAP und SIFT als schädlich eingestuft. Der Aminosäureaustausch führt zu Instabilität des Proteins. Zusammenfassend kann die Variante c.1858G>A (p.Asp620Asn) im Gen *VPS35*, als ursächlich für Morbus Parkinson in drei Familien eingestuft werden. Diese Studie zeigt exemplarisch die erfolgreiche Anwendung der Exom-Sequenzierung zur Erkennung von ursächlichen genetischen Varianten bei monogenen Erkrankungen.

3.3.3 Exom- und Genom-Sequenzierung zur Identifizierung pathogener Mutationen bei Patienten mit Burn-McKeown-Syndrom

Dieser Abschnitt der vorliegenden Arbeit wurde zuvor publiziert als:

Compound Heterozygosity of Low-Frequency Promoter Deletions and Rare loss-of-function Mutations in *TXNL4A* Causes Burn-McKeown Syndrome

Wieczorek, D., Newman, W.G., Wieland, T., Berulava, T., Kaffe, M., Falkenstein, D., Beetz, C., **Graf, E.**, Schwarzmayr, T., Douzgou, S., Clayton-Smith, J., Daly, S.B., Williams, S.G., Bhaskar, S.S., Urquhart, J.E., Anderson, B., O'Sullivan, J., Boute, O., Gundlach, J., Czeschik, J.C., Van Essen, A.J., Hazan, F., Park, S., Hing, A., Kuechler, A., Lohmann, D.R., Ludwig, K.U., Mangold, E., Steenpaß, L., Zeschmick, M., Lemke, J.R., Lourenco, C.M., Hehr, U., Prott, E.-C., Waldenberger, M., Böhmer, A.C., Horsthemke, B., O'Keefe, R.T., Meitinger, T., Burn, J., Lüdecke, H.-J., Strom, T.M.

American Journal of Human Genetics, 2014 95 (6), pp. 698-707.

In dieser Studie war ich verantwortlich für die Herstellung der Exom- und Genom-*Libraries*, sowie deren Sequenzierung. Gemeinsam mit den anderen Autoren wertete ich die Ergebnisse der Sequenzierungen aus und unterstützte bei der Erstellung des Manuskripts. *Mikroarray*-Analysen sowie funktionelle Untersuchungen wurden von den Kooperationspartnern durchgeführt.

3.3.3.1 Zusammenfassung

Durch eine Kombination Exom- und Genom-Sequenzierung, sowie *Mikroarray*-Analysen, wurden biallele Mutationen im Gen *TXNL4A*, bei Patienten mit Burn-McKeown-Syndrom, einer seltenen Erkrankung mit bilateralem Verschluss der Choanen (Choanalatresie) und kraniofazialen Fehlbildungen identifiziert. In neun von elf untersuchten Familien fanden sich Veränderungen in *TXNL4A*. In acht Familien wurden seltene *loss-of-function* Mutationen (*nonsense*, *frameshift* oder Mikrodeletionen) auf dem einen Allel, in Kombination mit einer 34 bp Deletion in der Kern-Promotorregion auf dem zweiten Allel identifiziert [94]. In einer konsanguinen Familie sind alle vier Betroffene homozygot für eine abweichende 34 bp Promoterdeletion in *TXNL4A*. Funktionelle Untersuchungen zeigen, dass die Promoterdeletionen zu einer verminderten *TXNL4* Expression führen [94]. *TXNL4A* ist eine essentielle Untereinheit des U5-Spleißkomplex. In verschiedenen Erkrankungen mit kraniofazialen Fehlbildungen, wie dem Nager-Syndrom oder

der Mandibulofazialen Dysostose Typ Guion-Almeida, wurden Mutationen in Genen beschrieben, die für Untereinheiten des Spleißosoms kodieren [94]. Die U5-spleißosomale-Untereinheit besteht aus acht hoch konservierten Proteinen und ist essentiell für das Spleißen von prä-mRNAs [166]. Die Ergebnisse zeigen, dass das Burn-McKeown-Syndrom eine autosomal-rezessive Erkrankung ist, die primär durch *compound*-heterozygote Promotordeletionen in Kombination mit sehr seltenen *loss-of-function* Mutationen hervorgerufen wird [94].

3.3.3.2 Hintergrund

Das Burn-McKeown-Syndrom (BMKS, MIM%608572) ist eine seltene Erkrankung, bei der sich die Patienten psychomotorisch normal entwickeln. Betroffene zeigen eine klassische Kombination aus Choanalatresie (Verschluss der hinteren Nasenöffnungen), sensorineuraler Schwerhörigkeit, Herzfehler und typischen kraniofazialen Dysmorphien [167, 168, 169]. Die Erkrankung wurde erstmals 1992 von Burn *et al.* [169] in drei Familien mit fünf Betroffenen beschrieben. Man ging von einer X-chromosomalen oder autosomal-rezessiven Vererbung aus [169]. Ein betroffenes Mädchen dieser Studie hatte ein Ringchromosom 18 (46,XX,r(18)(p14q23)), was zu der Hypothese führte, dass die genetische Ursache der Erkrankung am terminalen Ende des kurzen oder langen Arms von Chromosom 18 lokalisiert ist [169, 170].

Innerhalb dieser Studie wurden Mutationen im Gen *TXNL4* identifiziert, welches für eine Komponente der U5-spleißosomalen Untereinheit kodiert [94].

Das Spleißosom besteht aus fünf snRNAs (*small nuclear ribonucleic acid*) und mehr als 50 Proteinen und wird an prä-mRNAs rekrutiert um dort die Introns zu spleißen [171, 172] (Abb.: 57).

Bisher sind vor allem humane Erkrankungen bekannt, die durch genetische Veränderungen der Zielsequenzen des Spleißosoms entstehen [173, 174]. Aber auch Mutationen in Komponenten des Spleißosoms, die zu distinkten Phänotypen, wie Retinitis Pigmentosa, hereditärer Hypotrichosis simplex oder Kleinwüchsigkeit führen, wurden in den letzten Jahren beschrieben [175, 176, 177, 178, 172]. Diese Erkrankungen zählen zu den Spleißosomopathien.

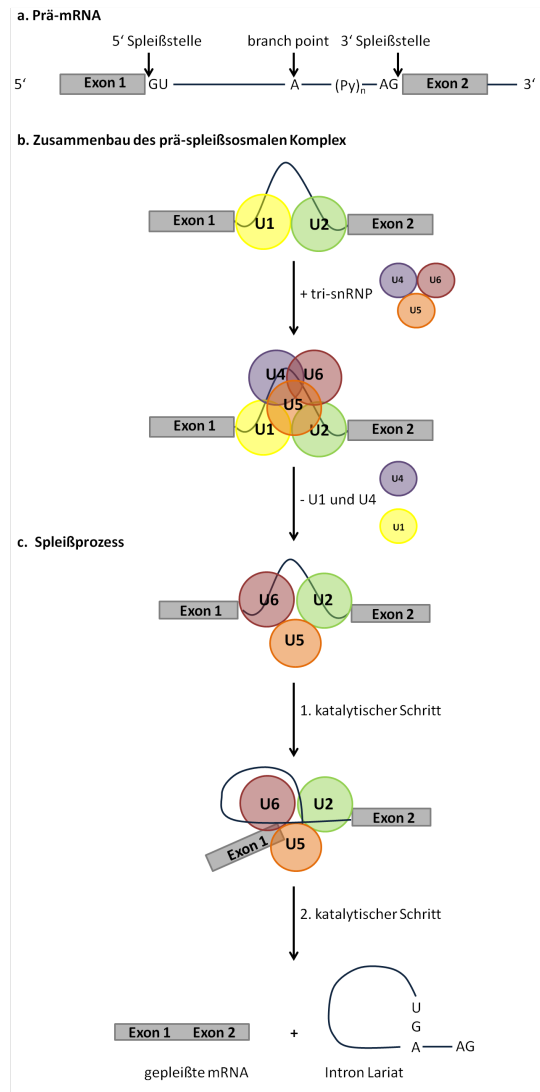


Abbildung 57: Übersicht zum Zusammenbau des Spleißosoms und dem anschließenden Spleißprozess. **(a)** Signalsequenzen an den Exon-Intron Übergängen, die zur Erkennung einer prä-mRNA durch spleißosomale Untereinheiten führen. Die kanonische Spleißstelle wird durch ein 5' GU Dinukleotid und ein 3' AG Dinukleotid gekennzeichnet. 10 bis 40 Nukleotide strangaufwärts des 3' AG befindet sich ein Adenin, auch als Verzweigungsstelle (*branch point*) bezeichnet. Zwischen der 3' Spleißstelle und dem *branch point* befindet sich zudem der Polypyrimidin-Trakt (Py_n). **(b)** Die U1 spleißosomale Untereinheit erkennt und bindet die 5' Spleißstelle, während die U2 spleißosomale Untereinheit zur Verzweigungsstelle rekrutiert wird. Im Anschluss erfolgt die Rekrutierung des zuvor assemblierten tri-snRNP Komplex zu U1 und U2. Folgend werden durch strukturelle Veränderungen die U1 und U4 Untereinheiten abgespalten und es entsteht das katalytisch aktive Spleißosom. **(c)** In einer Transesterifizierungsreaktion wird das 5' Exon abgelöst und die 5' Seite des Introns an den *branch point* gebunden. Die Lariatform des Introns wird in einer zweiten katalytischen Reaktion vom 3' Exon gelöst und abgebaut. Durch ATP-Hydrolyse werden die beiden Exons verbunden und die spleißosomalen Untereinheiten stehen für einen nächsten Zusammenbau des Spleißosoms zur Verfügung [171, 172].

3.3.3.3 Probandenauswahl

Die Patienten wurden anhand ihres klinisches Bildes für diese Studie ausgewählt. Die initiale klinische Untersuchung erfolgte durch mindestens einen Humangenetiker. In dieser Studie untersuchten wir elf Familien mit insgesamt 14 Betroffenen (BMKS001-BMKS011), 6 weibliche und 8 männliche Patienten (Abb.: 60) [94]. Alle Patienten zeigten eine Verkürzung der Lidspaltenbreite und eine bilaterale Choanalatresie/Choanalstenose (Verschluss der hinteren Öffnungen der Nasenhöhle). Mindestens 12 der 14 Betroffenen zeigen einen okulären Hypertelorismus, einen Defekt des unteren Augenlids, einen prominenten Nasenrücken und/oder ein verkürztes Philtrum. Mindestens acht der Patienten leiden unter Taubheit, haben dünne Lippen, eine Lippen-Kiefer-Gaumenspalte, prominente Ohren, präaurikuläre Anhängsel und/oder eine Mikrognathie [94]. Sporadisch treten Kleinwuchs, Herzfehler oder Kieferspalten auf. Alle Patienten zeigen eine normale psychomotorische Entwicklung. Fünf Familien zeigten eine positive Familienanamnese. Die Patientin der Familie BMKS009 ist in der initialen BMKS Publikation beschrieben und trägt ein Ringchromosom 18 [169, 94]. Als Kontrollpopulation wurden gesunde Probanden aus Manchester, Essen und der KORA-Studie ausgewählt. Insgesamt untersuchten wir die DNA von 3343 Probanden [94].

3.3.3.4 Exom-Sequenzierung

Wir untersuchten die fünf Betroffenen der Familien BMKS001, BMKS002, BMKS003 und BMKS005 mittels Exom-Sequenzierung. Nach vorheriger Anreicherung des Exoms mit dem *SureSelect XT Human All Exon 50 Mb v4 kit* der Firma Agilent wurde ein 100 bp *paired-end* Lauf auf dem HiSeq2500 System (Illumina) mit einem *Library-Pool* aus 12 Proben pro 4 *Lanes* durchgeführt (Tab.: 16).

Um mögliche pathogene genetische Alterationen zu finden, wurde das Datenset auf gemeinsame genetische Varianten oder auf Mutationen in den gleichen Genen der BMKS-Patienten hin untersucht. Dabei wurden in drei der vier Familien heterozygote *loss-of-function*-Varianten im *TXNL4A*-Gen ermittelt: die Brüder der Familie BMKS001 und der Betroffene der Familie BMKS002 tragen die *nonsense*-Variante c.349G>T; p.Glu117* bzw. c.37C>T; p.Gln13*. Die frameshift-Mutation c.131delT; p.Val44Alafs*48 in *TXNL4A*, liegt bei der Betroffenen der Familie BMKS005 vor [94]. Alle Mutationen der Indexpatienten koregrieren mit der Erkrankung innerhalb der Familie (Abb.: 58). Wir fanden

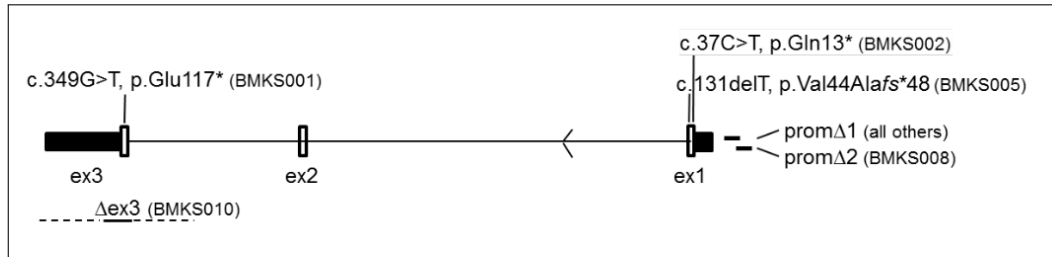


Abbildung 58: Genstruktur von *TXNL4A* mit den durch MLPA sowie Genom- und Exom-Sequenzierung identifizierten Mutationen [94]. Die Abbildung zeigt die drei Exons von *TXNL4A* und die Lokalisation der *loss-of-function*-Varianten, die durch eine Exom-Sequenzierung identifiziert wurden. Die durch MLPA-Analyse gefundene Exon 3 Deletion des Indexpatienten der Familie BMKS010, ist darunter dargestellt.

Tabelle 16: Statistik der Exom-Sequenzierung bei Patienten mit BMKS.

Patient	Reads	[Gb]	Ø Abdeckung [-fach]	20x Abdeckung [%]
BMKS001_1	98653970	9,96	127,68	95,69
BMKS001_2	82464779	8,33	108,23	94,72
BMKS002	85408646	8,63	110,11	93,74
BMKS003	99780314	10,08	129,18	91,14
BMKS005*				

*Das Exom der Betroffenen der Familie BMKS005 wurde extern sequenziert. Die Daten wurden uns zur Auswertung zur Verfügung gestellt.

keine dieser Varianten in den Daten des 1000-Genom-Projektes, in dbSNP und in 3000 internen Kontroll-Exomen. Bei Untersuchungen der Eltern zeigte sich, dass alle intragenischen *TXNL4A*-Varianten, auch in den gesunden Eltern, Großeltern oder Geschwistern [94] auftraten.

Die DNA der Betroffenen aus den 11 Familien wurde zusätzlich mittels *Mikroarrays* durch die Projektpartner analysiert [94]. Dies führte zur Identifikation von fünf heterozygoten Mikrodeletionen in 18q23, die alle *TXNL4A* (NM_006701) überspannten (Abb.: 59). Die Patientin der Familie BMKS004 trägt eine terminale Deletion (1,164 Mb), die fast identisch ist, mit der terminalen Deletion der beiden Brüdern der Familie BMKS007 (1,14 Mb). Die Betroffene der Familie BMKS006 trägt die kleinste terminale Deletion (0,484 Mb) und die Patientin des Stammbaums BMKS009 (Ringchromosom 18) trägt eine terminale Deletion von 4,7 Mb [94].

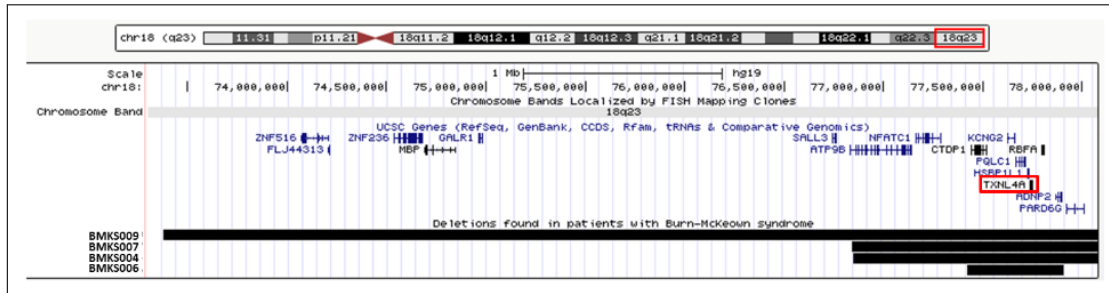


Abbildung 59: Identifizierte Mikrodeletionen nach der Untersuchung der DNA mittels *Mikroarray* [94]. Die heterozygoten Deletionen überspannen das *TXNL4A*-Gen auf Chromosom 18q23.

Die durch die Projektpartner durchgeführten qPCR Untersuchungen der parental DNA der Familien BMKS006 und BMKS007, ergaben, dass die gefundenen Deletionen von jeweils einem gesunden Elternteil ererbt wurden (BMKS006 von der Mutter, BMKS007 vom Vater) (Tab.: 18). Um auch kleinere Deletionen in Patienten mit BMKS zu finden wurde eine MLPA-Analyse für die drei Exons des *TXNL4A*-Gens durchgeführt. Es wurde zusätzlich eine heterozygote Deletion von Exon 3 in *TXNL4A* im Betroffenen der Familie BMKS010 identifiziert [94].

3.3.3.5 Genom-Sequenzierung

Das wiederholte Auftreten von heterozygoten *loss-of-function* Mutationen in *TXNL4A* in gesunden Familienmitgliedern, macht eine mögliche Haploinsuffizienz als Krankheitsursache unwahrscheinlich, so dass eine bisher unentdeckte zweite Variante innerhalb dieses Abschnitts vermutet wurde [94]. Basierend auf dieser Hypothese wurde zur Identifikation nicht kodierender Varianten, eine Genom-Sequenzierung in sechs betroffenen Individuen aus fünf Familien durchgeführt (BMKS001, BMKS002, BMKS004, BMKS005 und BMKS008 [179]). Die DNA-Libraries wurden mit dem *TruSeq DNA PCR-free Kit* (Illumina) nach Angaben des Hersteller präpariert und als 100 bp *paired-end*-Lauf auf dem HiSeq2500 sequenziert (Tab.: 17).

In den betroffenen Mitgliedern der Familie BMKS001, BMKS002, BMKS004 und BMKS005 wurde die gemeinsame heterozygote 34 bp-Deletion, chr18:g.77,748,581_77,748,614 (1 Δ) identifiziert. Die Deletion liegt innerhalb der angenommenen Promotorregion von *TXNL4A* (Tab.: 18). Sie besteht aus zwei gleichen 22 bp Motiven (CATGCCGTCAGCACGCACGGCG) die durch 12 bp (CTAGCGCGCGCG) geteilt wird. Diese 56 bp Region (hg19 chr18:g.77,748,581-

Tabelle 17: Statistik der Genom-Sequenzierung bei Patienten mit BMKS.

Patient	Reads	[Gb]	Ø Abdeckung [-fach]	20x Abdeckung [%]
BMKS001_1	1495359669	151,03	48,19	99,12
BMKS001_2	1480884019	149,57	45,71	98,79
BMKS002	1629589933	164,59	50,92	99,25
BMKS004	1525217593	106,8	49,75	99,57
BMKS005	1811071914	182,92	54,04	99,61
BMKS008	1386153622	140	42,23	98,11

77,748,636) befindet sich 49 bis 105 bp aufwärts der angenommenen Transkriptions-Startseite (GenBank acc. NM_06701.2) und enthält mehrere potentielle Bindestellen für Transkriptionsfaktoren. Wir identifizierten die 34 bp große Deletion 1 Δ , in der Promoterregion von *TXNL4A*, ebenfalls in den Betroffenen der restlichen vier Familien mit heterozygoten Mikrodeletionen. Parentale DNA war von neun Betroffenen verfügbar. Segregationsstudien zeigten, dass die gemeinsame 34 bp-Deletion 1 Δ von dem gesunden Elternteil, welches nicht die *loss-of-function* Mutation trägt, ererbt wurden (Abb.: 60).

Interessanterweise trägt die Patientin der Familie BMKS008 [180] eine sehr ähnliche homozygote Deletion in der angenommenen Promoterregion von *TXNL4A*, die in 11 bp mit 1 Δ überlappt (Abb.: 58): chr18:g.77,748,604_77,748,637del (Δ 2). Die Segregationsanalyse der Familie BMKS008 zeigte eine Kosegregation der homozygoten Deletion mit der Erkrankung, während alle heterozygoten Träger gesund erscheinen [94].

Die Deletionen 1 Δ und 2 Δ betreffen die 22 bp Wiederholung in der Promoterregion. Deletionen in der gleichen Region kommen nicht im Datenset des 1000-Genome-Projekts vor. Dieser Umstand ist womöglich bedingt durch die nicht ausreichende Repräsentation dieser GC-reichen Region in genomischen DNA-Libraries, die mit einem Protokoll hergestellt wurden, welches einen Amplifikationsschritt beinhaltet. Diese Annahme wird unterstützt durch die Tatsache, dass die Region in internen *low-coverage* Genomen, die ebenfalls mittels PCR amplifiziert wurden, nicht abgedeckt ist [94].

Daher wurde die DNA von 3165 deutschen Probanden und 178 süd-asiatischen Probanden klassisch über PCR untersucht und die Produkte auf einem Agarosegel analysiert. Wir identifizierten 45 heterozygote und eine homozygote 1 Δ -

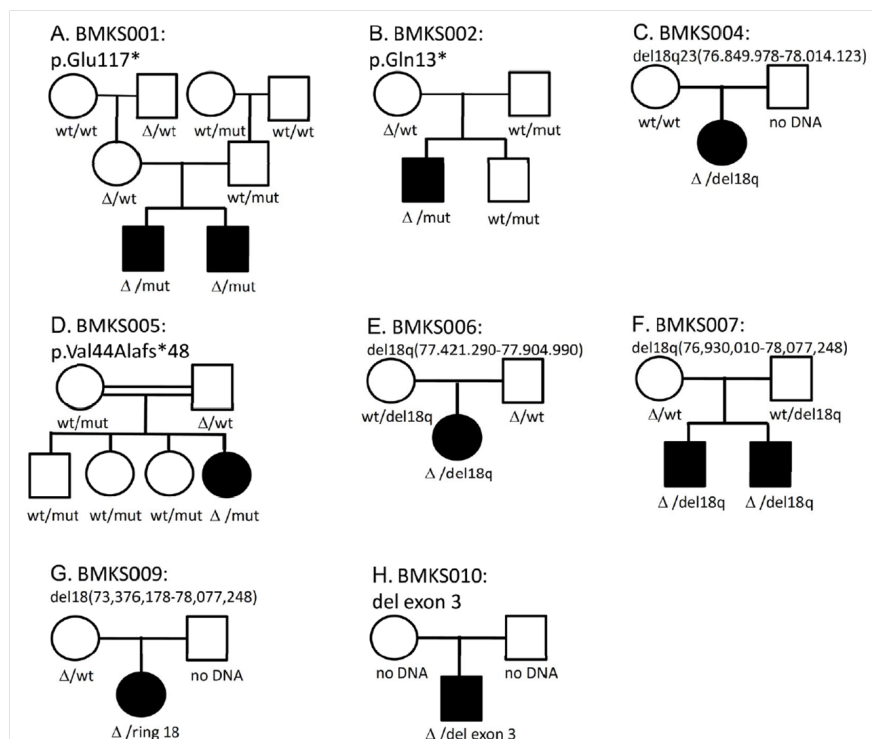


Abbildung 60: Stammbäume der Indexpatienten mit BMKS und biallelen Mutationen in *TXNL4A* [94].

Deletion und eine heterozygote 2 Δ -Deletion. Daraus ergibt sich eine deutsche Allel-Trägerfrequenz von circa 0,76 %. Es standen keine Daten zum kraniofazialen Phänotyp der betroffenen Probanden zur Verfügung [94].

Untersuchungen des Haplotyps aus dem Datensatz der Genom-Sequenzierung zeigten, dass die Promoterdeletionen auf verschiedenen Haplotypen liegen, so dass eher von einem wiederkehrenden Mutationsereignis, als von einem Gründereffekt, auszugehen ist [94].

Zusätzlich fanden wir in elf Probanden (Allel-Frequenz 0,17 %) eine heterozygote 34 bp-Insertion (hg19 chr18:g.77,748,580_77,748,581 ins chr18:g.77,748,581-77,748,614) in die Promoterregion. Die inserierte Sequenz ist identisch zu der 34 bp 1 Δ -Deletion und muss als Kopf-zu-Schwanz Tandemduplikation angesehen werden. Das ergibt eine 90 bp Sequenz mit drei 22 bp Motiven, die durch zwei 12 bp-Motive getrennt werden. Dieses Ergebnis unterstützt die Annahme, dass diese Region in *TXNL4A* anfällig für wiederkehrende Mutationsereignisse ist [94].

Funktionelle Untersuchungen durch die Projektpartner zeigten die Reduktion der Expression von *TXNL4A* um 59 % bei Vorhandensein der 1 Δ -Deletion und 72 % bei der 2 Δ -Deletion im Vergleich zum Wildtyp. Die verminderte Expression weist darauf hin, dass die Region tatsächlich relevant für die Expression des Gens ist. Dennoch scheint die Expression von *TXNL4A* bei einer homozygoten 2 Δ -Deletion ausreichend zu sein, um den Phänotyp hervorzurufen [94].

Das Spleißosom ist ein großer Komplex aus sechs Untereinheiten (U1 bis U6), der aus mehr als 50 Proteinen und fünf snRNAs aufgebaut ist. Die U5 spleißosomale Einheit besteht aus acht hoch konservierten Proteinen und ist kritisch für das Spleißen von prä-mRNAs [166]. In Hefe ist das orthologe Gen zu *TXNL4A* bekannt als *DIB1/SNU16*. Es kodiert ein hoch konserviertes 138 Aminosäuren langes Protein, welches essentieller Bestandteil des U4/U6.U5 tri-snRNP Komplexes ist und notwendig für das Spleißen von prä-mRNAs in Hefe [181]. In Hefe führen Nullmutationen in *DIB1* zu Letalität. Daher könnten auch homozygote Nullmutationen im Menschen nicht vereinbar mit dem Leben sein. Die Auswirkungen der reduzierten Expression von *DIB1* in Hefe auf den Zusammenbau des Spleißosoms wurden final untersucht. Es zeigte sich, dass eine Reduktion der Expression in einer verminderten Assemblierung des U4/U6.U5 tri-snRNP Komplexes resultierte [94].

Es ist anzunehmen, dass durch die hochgradige Homologie zwischen *DIB1*

und TXNL4A ebenso wie in der Hefe, auch beim Menschen die Assemblierung des U4/U6-U5 tri-snRNP Komplexes herabsetzt wird [94]. Ein verminderter Zusammenbau des U4/U6.U5 tri-snRNP Komplexes könnte das Spleißen spezifischer prä-mRNA Untergruppen beeinflussen und so zum Phänotyp des Burn-McKewon-Syndroms führen [94].

Diese zweite Studie zeigt exemplarisch, welche Grenzen der Exom-Sequenzierung gesetzt sind. Als aussichtsreiche Methode zur Identifikation von genetischen Alterationen in den kodierenden Abschnitten des Genoms, bleiben mögliche kausative Varianten in den nicht-kodierenden Bereichen unerkannt. Zusätzlich war es zu Beginn der *NG*-Sequenzierung bioinformatisch erschwert, größere Deletionen und Insertionen in Exom- und Genom-Daten zu identifizieren. Dazu waren zusätzliche Untersuchungen wie *DNA-Mikroarrays* oder *MLPA*-Analysen notwendig. Erst in den letzten Jahren konnte durch die Entwicklung bioinformatischer Anwendungen auch die Untersuchung auf Insertionen und Deletionen innerhalb von *NGS*-Daten erfolgen.

Tabelle 18: Übersicht der in *TXNL4A* identifizierten Mutationen und Deletionen [94].

Stammbaum/ Individuen	1. <i>TXNL4A</i> - Mutation hg19; chr18	Parentale Herkunft	2. <i>TXNL4A</i> - Mutation hg19; chr 18	Parentale Herkunft
BMKS001 III/1 und III/2	g.77,733,765C>A, c.349G>T, p.(Glu117*), Exon 3	paternal	Promotor Δ 1	maternal
BMKS002	g.g.77,748,356G>A, c.37C>T, p.(Gln13*), Exon 1	paternal	Promotor Δ 1	maternal
BMKS003	negativ		negativ	
BMKS004	18q23(76,854,774-78,077,248)x1 Δ 1.222 Mb, del(18)(q23>qter)	?	Promotor Δ 1	?
BMKS005	g.77,748,262delA, c.131delT, p.(Val44Alafs*48), Exon 1	maternal	Promotor Δ 1	paternal
BMKS006	18q23(77,421,290-77,904,990)x1, Δ 0.484 Mb,	maternal	Promotor Δ 1	paternal
BMKS007 II/1 und II/2	18q23(76,841,645-78,077,248)x1 Δ 1.235 Mb, del(18)(q23>qter)	paternal	Promotor Δ 1	maternal
BMKS008	Promotor Δ 2	paternal maternal	Promotor Δ 2	paternal maternal
BMKS009	18q23(73,376,178-78,077,248)x1 Δ 4.701 Mb, del(18)(q23>qter)	<i>de novo</i>	Promotor Δ 1	maternal
BMKS010	Δ von Exon 3	?	Promotor Δ 1	?
BMKS011	negativ		negativ	

Zusammenfassung der gefundenen genetischen Varianten in 11 Familien mit BMKS. Die zugrunde liegende *TXNL4A*-Referenzsequenz ist NM_006701.2. Die Sequenz aus der Genom-Sequenzierung erstreckt sich bis zu g.78,017,154, wo die terminale *Repeat*-Region beginnt. Die Positionen der Endpunkte der terminalen Deletionen sind auf 78,077,248 anhand der Länge des Chromosom 18 gesetzt. Promotor Δ 1: WGS[hg19] chr18:g.77,748,581_77,748,614delCGCCGTGCGTGCTGACGGCATGCGCGCGCGCTAG. Promotor Δ 2: WGS[hg19] chr18:g.77,748,604_77,748,637delGCGCGCGCTAGCGCCGTGCGTGCTGACGGCATGT. Die Überlappung der Promotor Δ 1 und der Promotor Δ 2 sind unterstrichen.

4 Diskussion

Die vorliegende Arbeit beschreibt die Einführung und Entwicklung der Exom-Sequenzierung am Institut für Humangenetik. Die klassischen Strategien zur Identifikation von krankheitsverursachenden genetischen Veränderungen wurden seit dem Jahr 2009 schrittweise durch die *Next-Generation*-Sequenzierung abgelöst. Der Fokus liegt dabei auf der Untersuchung des Exoms von betroffenen Patienten. Da die Hochdurchsatzsequenzierung ihre erfolgreiche Anwendung in vielen Publikationen bewiesen hat, ist das Probenaufkommen entsprechend hoch. Um den stetig wachsenden Nachfragen nachkommen zu können, wurde zum Einen ein *NGS-LIMS* und zum Anderen eine Automation der *Library*-Präparation entwickelt. Neben der Optimierung und Standardisierung der Laborabläufe wurden Standards an das Ausgangsmaterial für eine *Library*-Herstellung abgeleitet.

Die Evolution der Laborfertigkeiten konnte genutzt werden um neue Protokolle zu entwickeln, wie die Exom-*Library*-Präparation aus weniger Startmaterial mit erhöhter *Insert*-Größe oder aber der *Library*-Präparation aus FFPE-Material. Die erfolgreiche Anwendung der automatisierten *Library*-Präparation und des LIMS, zur Erkennung der ursächlichen Varianten bei monogenen Erkrankungen, wird exemplarisch durch zwei publizierte Studien gezeigt. Hierbei wurden durch Exom-Sequenzierung, im Fall der Patienten mit Morbus Parkinson und einer Kombination aus Exom- und Genom-Sequenzierung bei Patienten mit Burn-McKeown-Syndrom, die kausativen genetischen Varianten identifiziert.

4.1 Entwicklung eines LIMS für die *Next-Generation-Sequenzierung*

Labor-Informations- und Managementsysteme wurden entwickelt um automatisierte Laborprozesse zu unterstützen. Die ersten Beschreibungen zu LIM-Systemen stammen aus dem Jahr 1973, die ersten kommerziellen LIM-Systeme sind seit 1982 verfügbar [46]. Mit dem Einzug der Hochdurchsatz-Sequenzierung war es notwendig auch für diesen Bereich eine Anwendung zu entwickeln, die mit dem Einzug der *NGS*-Technologie am Institut für Humangenetik im Jahr 2008 nicht zur Verfügung stand. Kommerziell erhältliche Systeme [182, 107] sind oft teuer, wenig flexibel und undurchsichtig. Quelloffene Anwendungen [183, 184, 185] wie GnomEx [186] sind erst seit 2010 verfügbar. Zudem bedürfen externe Anwendungen typischerweise vieler Anpassungen, um den individuellen Ansprüchen innerhalb der *NGS*-Labore gerecht zu werden [187].

Die vorliegende Arbeit beschreibt die Funktionsweise eines Labor-Informations- und Management-Systems zur Verwaltung von *NGS*-Abläufen und -Daten und beleuchtet die laborrelevanten Anwendungen. Die Registrierung von Proben und die Steuerung von Laborprozessen im Bereich der Hochdurchsatz-Sequenzierung wird durch die Nutzung eines Labor-Informations- und Management-Systems vereinfacht und standardisiert.

Die Aufgaben, mit denen ein *NGS*-Labor konfrontiert ist, sind die Koordination und Organisation von Sequenzieranfragen, Probeneingängen, Prozessen und der Mitteilung und Darstellung von Analyseergebnissen [188]. Die Koordination umfasst Fragestellungen des Probenverbleibs und Probenstatus sowie der Analyse von Sequenzierdaten über einen Zeitraum von mehreren Wochen pro Probe. Zeitgleich soll jeder Mitarbeiter auf die gleichen Daten Zugriff haben.

Unser Anspruch an ein LIMS erstreckte sich von der Probenregistrierung, Probenverwaltung, Probenverfolgung bis hin zu mathematischen Kalkulationen und Qualitätskontrollen innerhalb der *NGS*-Arbeitsabläufe. Daneben sollten administrative Aufgaben wie Rechnungsstellung und Bestellungen dokumentiert werden. Wir entwickelten ein web-basiertes LIM-System für den *NGS*-Bereich, das unsere Anforderungen abdeckt und auch von Anwendern ohne informatische Grundkenntnisse genutzt werden kann. Das LIMS erstreckt sich über drei Ebenen: von der Probenebene, über die Prozessebene bis hin zur Analyseebene. Während die Probenebene zur Registrierung von Probandaten dient, deckt die Prozessebene

alle relevanten Abläufe der *Library*-Präparation und der Sequenzierung ab. Innerhalb der Analyseebene können Exom-, Genom-, ChIP- und Transkriptom-Daten ausgewertet werden. Alle relevanten Patienten-/Probanddaten, Prozessinformationen und Analyseergebnisse können zu jederzeit vom Anwender abgerufen werden. Die Anwendung ist einfach und intuitiv gehalten.

4.1.1 Weiterentwicklung des *NGS-LIMS*

Die beschriebenen Grundfunktionen des LIMS wurden im Verlauf der Entwicklung um spezielle Applikationen erweitert. Die bisherige Anwendung zeigte, dass vor allem manuelle Einträge zu Fehlern bei der Probenregistrierung und Probenverarbeitung führten. Um dem entgegen zu wirken wurde ein *Sample Manifest* für den Import von Probanddaten entwickelt.

Für den Prozess der *Library*-Präparation wurde ein *Sample Tracking Sheet* generiert, welches sukzessive über den *Library*-Präparationsprozess ausgefüllt und im Anschluss mit den notwendigen Daten wieder in das LIMS reimportiert wird. Eine weitere Entwicklung war die Einführung eines 2D-Barcode Systems von Proben-, *Library*- und *Pool*-Röhrchen. Dadurch wurde ein manuelles Eintippen von Identifikationsnummern gänzlich umgangen.

Zusätzlich dazu wurde die Rekalkulation von genomischen *Library-Pools* eingeführt. Dadurch wurde die Sequenziertiefe einzelner *Libraries* vereinheitlicht, sowie eine Über- oder Untersequenzierung einzelner Proben vermieden. Die Kapazität der Sequenziermaschinen kann so gänzlich ausgenutzt werden, bei gleichzeitigem Erhalt von vergleichbaren Abdeckungen einzelner Proben. Die Anwendung der Rekalkulation von *Library-Pools* hat sich auch in anderen Laboren etabliert [110].

Eine weitere Implementierung war die Einführung der Kontaminationskontrolle [58] zur Aufrechterhaltung der Genotyp-Qualität und Identität exomischer und genomischer Proben. Besonders bei der Suche nach Varianten in heterogenen Tumoren oder bei der Identifikation von genetischen Mosaiken, bei denen man kein heterozygoten Varianten-Verhältnis von 50:50 erwartet, muss eine hohe Probenidentität bestehen. Bei RNA-Sequenzierstudien ist eine hohe Probenidentität vor allem bei der Suche nach sehr seltenen Transkripten essentiell. Mit der Bestimmung der Kontamination einer Probe, kann zum Einen eine stetige Qualitätskontrolle erfolgen, die Anwendung kann aber auch, wie sich in der Vergan-

genheit zeigte, systematische Fehler aufdecken.

Um die Kapazitäten der Sequenziermaschinen auslasten zu können und *Lane*-abhängige Sequenzierartefakte bioinformatisch besser normalisieren zu können, werden *Libraries* für die Sequenzierung zu einem *Library-Pool* gemischt. Die einzelnen *Libraries* besitzen einen patientenspezifischen Index, der während der Sequenzierung mitbestimmt wird. Innerhalb der bioinformatischen Analyse können die einzelnen *Reads* über den Index einem Patienten zugeordnet werden.

Über die eingeführte Kontaminationsanalyse konnte festgestellt werden, dass mit dem Einzug der *Patterned Flow Cell*-Technologie des HiSeq4000, die Kontaminationsrate bei exomischen *Libraries* stieg. Der Effekt war auf Grund des geringen Durchsatzes und der damit verbundenen geringen Multiplexierung von genomischen *Libraries*, bei Genom-Sequenzierungen nicht sichtbar. Untersuchungen [19, 111] zeigten, dass die Erhöhung der Kontaminationsrate durch den Prozess des *Adapter Hoppings* ausgelöst wird. Dabei kommt es zum Austausch patientenspezifischer Indizes innerhalb eines *Library-Pools* und im Folgenden zu einer falschen Zuordnung von genomischen Fragmenten zu den einzelnen Patienten innerhalb der bioinformatischen Analyse. Die Anwendung einer Index-Strategie für genomische *Libraries*, bei der nur individuelle duale Indizes genutzt werden [19, 111], führte zu einer eindeutigen Reduktion der Kontamination. Für exomische *Libraries* standen keine individuellen dualen Indizes zur Verfügung. Hier wurde ein *Blocking*-Reagenz (Illumina) genutzt, um die freien 3'OH-Enden der im Pool verbliebenen Indizes zu blockieren und die Kontaminationsrate zu reduzieren. Dabei war vor allem interessant, dass durch die beiden Strategien, die Kontaminationsrate auf dem HiSeq4000 unter die des HiSeq2500 gefallen ist. Daraus lässt sich schließen, dass es auch bei der *Non-Patterned Flow Cell*-Technologie zum Austausch von Indizes innerhalb der *Libraries* eines *Pools* kommen kann, was auch vergangene Studien bestätigen [19, 189, 190]. Je mehr exomische oder genomische *Libraries* miteinander gemischt werden, desto höher ist die Kontaminationsrate einzelner Proben [111]. Das erklärt auch, warum bei der Genom-Sequenzierung keine Kontaminations-Erhöhung im Vergleich zur Sequenzierung auf dem HiSeq2500 sichtbar wurde. Der Durchsatz der Genom-Sequenzierung ist bisher im Vergleich zur Exom-Sequenzierung gering, wodurch eine hohe Multiplexierung der Proben vermieden wird. Maximal 6 Proben pro *Flow Cell* werden derzeit für einen Genom-Pool gemischt, im Vergleich zu 40 Proben für einen Exom-*Pool*.

In der bestehenden Literatur zum *Adapter Hopping* auf *Patterned Flow Cells* wird ein Kontaminationsgrad von bis zu 10 % [19] angegeben. Eine so hohe Kontamination konnte in der vorliegenden Arbeit nicht detektiert werden. Es zeigte sich, dass die Kontamination bei unterschiedlichen *Library*-Präparationsprotokollen unterschiedlich hoch ausfällt [111]. Dabei ergaben PCR-freie Protokolle im Vergleich zu Protokollen mit PCR-Schritt eine höhere Kontaminationsrate. Je besser die Aufreinigung der finalen *Library* vor der Multiplexierung der Proben ist, desto weniger freie Adapter befinden sich im *Library-Pool* und desto geringer ist der Grad der Kontamination pro Probe [111]. Die Kontamination durch die Anwendung des HiSeq4000 stieg bei exomischen *Pools* um 0,4 %. Freie Adapter konnten bei der Qualitätsanalyse am Ende der *Library*-Präparation nicht detektiert werden, wodurch sich der geringe Anstieg der Kontamination erklären lässt. Neben dem HiSeq4000 arbeitet auch der HiSeqX und die neu eingeführte Sequenziermaschine, Illumina NovaSeq, mit der *Patterned Flow Cell*-Technologie [191]. Bei der Projektplanung und der Anwendung dieser Sequenzier-technologie ist somit zwingend auf die Problematik zu achten und entsprechende Strategien zur Vermeidung des *Adapter Hopping* müssen mit einbezogen werden. Gleichzeitig konnte gezeigt werden, dass nicht nur die Menge freier Adapter innerhalb eines *Library-Pools* die Rate der Kontamination beeinflusst, sondern dass es auch zu Unterschieden bei individuellen Läufen kommt. Das kann auf das Anmischen der Sequenzier-Polymerase, unterschiedliche Hersteller-Lots der Sequenzierchemie oder durch die *Flow Cell* selbst entstehen [111].

Eine generelle Reduktion der Kontaminationsrate auf 0 % war durch die Anwendung der verschiedenen Strategien nicht möglich. Für genomische *Libraries* wird eine minimale Verunreinigung von 0,2 % und für exomische *Libraries* von 0,6 % erreicht. Andere Studien zeigen ebenfalls eine Grundkontamination von unter 1 % in Abhängigkeit von der Art und Herstellung der *Library* [111]. Es kann sich hierbei um echte Kontaminationen, PCR-Artefakte oder Sequenzierartefakte handeln. Es kann aber auch, bedingt durch die Art der Kontaminationsdetektion, die auf einem statistischen Verfahren beruht, zu einem Hintergrundrauschen kommen, welches sich durch eine marginale Pseudokontamination ausdrückt. Generell gilt die Empfehlung die Grundkontamination bei der Datenauswertung zu beachten und vor allem bei der Suche nach Varianten, die nicht im homo- oder heterozygoten-Status zu erwarten sind, mit einzubeziehen.

Das in dieser Arbeit entwickelte LIM-System bietet alle notwendigen Grundfunktionen der Probenerfassung, Proben- und Prozessverfolgung sowie die Qualitätskontrolle innerhalb des *NGS*-Labors an, um tausende von Proben und Probanden zu organisieren. Die Grundfunktionen wurden durch weitere qualitätsverbessernde Anwendungen erweitert, wodurch ein umfangreiches Konstrukt entstand. Das LIMS kann zu jederzeit an aktuelle Prozesse, neue Methodiken oder geänderte Abläufe angepasst werden und ist eine wichtige Grundlage zur Organisation und Koordination von *NGS*-relevanten Daten und Prozessen.

4.2 Einführung einer automatisierten Exom-*Library*-Präparation

Mit der Einführung von Exom-Anreicherungstechniken im Jahr 2009 stieg die Anfrage nach *NG*-Sequenzierungen enorm. Im Vergleich zur Genom-Sequenzierung, ist die Exom-Sequenzierung ein kostengünstiger und erfolversprechender Ansatz zur Identifikation pathogener Varianten in kodierenden Bereichen des Genoms [23, 113, 114, 115, 78]. Um der steigenden Nachfrage nachkommen und reproduzierbar im größeren Maßstab arbeiten zu können, wurde in der vorliegenden Arbeit eine Automation der *Library*-Präparation entwickelt. Diese wurde in zwei Schritten eingeführt: (a) durch die Entwicklung von Protokollen für eine Semi-Automation und (b) durch die Erweiterung der bestehenden Automation zur weitgehenden Automation. Zum Einzug der Exom-Sequenzierung am Institut für Humangenetik Ende 2009 stand keine Automation der exomischen *Library*-Präparation zur Verfügung. Mit der Anschaffung der *Bravo Automated Liquid Handling Platform* (Agilent) und der Weiterentwicklung des manuellen *Library*-Präparationsprotokolls *SureSelect Target Enrichment System for Illumina Paired-End Sequencing Library, Version 2.0.1, May 2010* wurden die Grundlagen für eine Automation geschaffen. Die erste Publikation zur automatisierten *Library*-Präparation wurden ein Jahr später veröffentlicht [112]. Kurze Zeit danach stellte auch Agilent ein Automationsprotokoll für die Anwendung der *SureSelect* Reagenzien in Verbindung mit der *Bravo Automated Liquid Handling Platform* zur Verfügung.

Das bisher genutzte Protokoll zur manuellen Herstellung von exomischen *Libraries* konnte nicht automatisiert werden. Nach jedem enzymatischen Schritt innerhalb der Präparation, war eine säulchenbasierte Aufreinigung der Lösung über das *QIAquick PCR Purification Kit* bzw. *Qiagen MinElute Kit* (insgesamt 5 mal im Prozess) notwendig. Eine Automation dieser Schritte war durch die notwendigen Zentrifugationsschritte nicht möglich. Im darauf folgenden Präparationsprotokoll (Mai 2010) wurden die säulchenbasierten Aufreinigungsschritte, so wie sie zur Einführung der Exom-*Library*-Präparation genutzt wurden, durch eine Aufreinigung mit AMPure XP *Beads* ersetzt. Mit der Einführung der magnetischen AMPure XP *Beads* anstelle von Säulchen zur Aufreinigung der DNA-Enzym-Lösungen, konnten diese Prozessschritte automatisiert werden.

Das Automationsprotokoll sollte möglichst viele Schritte der *Library*-Präpara-

tion abdecken bei gleichzeitig geringer Bearbeitungszeit durch den Anwender. Es wurde ein Protokoll zur automatisierten Aufreinigung über AMPureXP *Beads* entwickelt, das es erlaubte bis zu 48 Proben parallel zu reinigen. Für jeden Aufreinigungsschritt musste die Automationsplattform nur ein mal bestückt werden. Dies gelang durch die Anwendung einer Platten-Stapelfunktion für alle notwendigen Reagenzienplatten. Die aufzureinigenden Proben wurden zusammen mit allen Reagenzien auf die Plattform geladen, das Protokoll gestartet und die aufgereinigten Proben am Ende des Protokolls wieder entnommen um sie dem nächsten enzymatischen Schritt zuzuführen. Zusätzlich wurden zwei weitere Automationsprotokolle entwickelt: zur Anreicherung der exomischen Regionen sowie der finalen Elution nach der Anreicherung. Neben der Einführung der automatisierten *Library*-Präparation konnten weitere Engpässe im Labor identifiziert und optimiert werden. Geräte mit geringem Durchsatz (Covaris S2, SavantSpeed Vac, Bioanalyzer 2100) wurden gegen Geräte mit hohem Durchsatz (Covaris E220, Eppendorf Concentrator Plus, Caliper LabChip GX) getauscht. Die Automation zusammen mit der Optimierung des Laborablaufs durch den Austausch von Geräten konnten 48 *Libraries* in 7 Werktagen hergestellt werden. Die Semi-Automation bot zusätzlich eine Skalierbarkeit von 12 bis 48 Proben, je nach Probenaufkommen.

Neben der Einrichtung der Semi-Automation und der Optimierung der Laborausstattung wurden zusätzliche Maßnahmen zur Qualitätssicherung ergriffen: Der automatisierte Prozess findet, im Gegensatz zum händischen Protokoll, in Mikrotiterplatten statt. Das verringerte die Gefahr von Verwechslungen und Kontaminationen innerhalb des prozessierten Probensatzes. Als Teil der Automation entwickelte ich eine Automations-Anleitung und ein *Lab Tracking*-Formular zur Unterstützung und Protokollierung des Laborablaufs. Innerhalb der Automation wurden diverse Kontrollpunkte eingerichtet. Vor jedem Automationsstart wird über die Software noch einmal die Oberflächen-Konfiguration des Bravo visualisiert und abgefragt. Die eingegebenen Volumen-Parameter innerhalb einer *Bead*-basierten Aufreinigung müssen am Ende noch einmal bestätigt werden. Zusammen mit dem Automations-Protokoll entstand eine umfassende Standardisierung und Kontrolle der Prozesse.

Ein Vergleich der automatisierten und manuellen AMPure XP *Bead*-Aufreinigung zeigte, dass die Ausbeute des eluierten Materials bei der automatisierten Reinigung um bis zu 29 % geringer war. Dieser Effekt lässt sich auf die nicht

vorhandene Flexibilität der Automation zurückführen. So ist ein sensibler Schritt während der Aufreinigung die Trocknung der *Beads* vor der DNA-Elution, um Reste von Ethanol aus dem Waschprozess zu entfernen. Werden die *Beads* dabei zu lang getrocknet, verringert sich die Elutionsrate der gebundenen DNA. Die Trocknungszeit ist zum Einen abhängig vom fixen Faktor der *Bead*-Menge, aber auch vom flexiblen Faktor der Raumtemperatur. Auf den flexiblen Faktor kann die Automation nicht eingehen. Der manuelle Bearbeiter kann hingegen visuell den Trocknungsgrad des *Bead*-Pellets einsehen und entsprechend reagieren. Daher ergibt sich für den Einzelprozess der Aufreinigung ein Nachteil hinsichtlich der Automation. Im Gesamtprozess zeigte sich hingegen, dass die Semi-Automation um einiges besser arbeitet als der manuelle Anwender. Während die durchschnittliche exomische Abdeckung annähernd gleich ist, zeigt sich eine deutliche Erhöhung der 20-fachen Abdeckung, sowie eine Erhöhung der Hybridisierungseffizienz (*On-Bait-Rate*). Die Duplikatrate hingegen verdoppelte sich mit der Einführung der Automation, was durch den Verlust von Material und somit ein Verlust der Komplexität, durch die Aufreinigung zu erklären ist.

Im Jahr 2011 publizierten Fisher *et al.* [112] ein Automationsprotokoll für die Bravo-Plattform und der *SureSelect*-Anreicherungs-methode. Sie entwickelten eine *On-Bead*-Methode, bei der die DNA nach der Aufreinigung mit AMPure XP *Beads* direkt in den Master Mix eluiert wird. Die *Beads* verbleiben während der enzymatischen Schritte in der Lösung und werden durch die Zugabe eines Puffers aus Polyethylenglycol (PEG) und Salzen in ihrer Funktionsweise reaktiviert, sobald die enzymatische Reaktion abgelaufen ist. Damit erreichten die Forscher eine durchschnittliche Duplikatrate von 4 % bei einer mittleren exomischen Abdeckung von 130-fach, mit einer 20-fach Abdeckung von 87,6 % der Basen. Mit dem am Institut für Humangenetik entwickelten Semi-Automationsprotokoll lag die Duplikatrate bei durchschnittlich 16 %, wobei die 20-fache Abdeckung bei 95 % lag.

Die hier dargestellte Entwicklung der Semi-Automation fand ebenso Anwendung bei der Herstellung von Genom-*Libraries* und wurde dazu übernommen. Die Automations-Anleitung und das *Lab Tracking*-Formular wurden entsprechend angepasst. Die Genom-*Library*-Präparation konnte dadurch von 12 auf 48 Proben pro Durchlauf erhöht werden.

Der Nachteil der etablierten Semi-Automation besteht in der maximalen Pro-

zessierung von 48 Proben. Dieser Umstand ist bedingt durch die wenigen Positionen, die auf dem Bravo für Platten und Spitzen zur Verfügung stehen. In einem zweiten Entwicklungsschritt wurde die Semi-Automation durch Erweiterung des *Bravo* um den Agilent *BenchCel 4R Microplate Handler* und den Agilent *Labware MiniHub* von einer weitgehenden Automation abgelöst. Die Protokolle für die weitgehende Automation werden von Agilent zur Verfügung gestellt und bedürfen keiner weiteren Modifikation. Bis auf die initiale Verdünnung der DNA, der Qualitätskontrolle auf dem *Caliper LabChip GX* vor der Hybridisierung und am Ende der *Library*-Präparation, sowie dem Anmischen der notwendigen Master Mixe, finden alle Schritte automatisiert auf der *Bravo Liquid Handling Station NGS Option B* statt. Die 20-fache exomische Abdeckung, die *On-Bait*-Rate und die Duplikatrate wurde dadurch noch einmal ein wenig erhöht. Mit der weitgehenden Automation können innerhalb von 7 Werktagen 96 exomische *Libraries* präpariert werden.

Die weitgehende Automation kann ebenso für die Genom-*Library*-Präparation eingesetzt werden. Hier können 96 Genom-Proben in 1,5 Werktagen fertig gestellt werden. Allerdings müssen die PCR-freien Genom-*Libraries* im Kontrast zu exomischen *Libraries* mittels qPCR quantifiziert werden, bevor sie auf einer *Flow Cell* geclustert werden können. Der Schritt der qPCR ist sehr zeit- und arbeitsaufwendig und gleichzeitig anfällig für Fehler. In der Zukunft soll die Automation auch dazu genutzt werden, das qPCR-Protokoll auszuführen, um reproduzierbarer und effizienter arbeiten zu können. Zusätzlich sollen die Schritte der *Library*-Präparation, die bisher nicht automatisiert sind, wie die initiale Verdünnung der DNA und das *Poolen* der *Libraries* automatisiert werden.

Mittlerweile stehen dem *NGS*-Anwender neben der Automationsplattform von Agilent mehrere kommerzielle Plattformen zur Verfügung (Perkin Elmer, Beckman Coulter, Eppendorf, Hamilton, Tecan). Ein systematischer Vergleich zwischen den Automationsanbietern und deren Protokollen zur Effizienz sowie den Kosten einer *Library*-Präparation steht nicht zur Verfügung.

4.2.1 Qualifikationsablauf und Ableitung der Anforderungen an DNA-Material für die NGS

Neben der Entwicklung einer Automation, etablierte ich einen Ablauf zur Bestimmung der Qualifikation einer DNA für eine *Library*-Präparation. Das Gelingen der zeit- und kostenintensiven *Library*-Präparation sowie die erfolgreiche Sequenzierung einer DNA-Probe sind abhängig von der Qualität des Ausgangsmaterials. Eine hohe Qualität ist essentiell um eine hohe Sequenziergenauigkeit zu erreichen, sowie die notwendige Sequenziertiefe. Die Integrität, die Reinheit sowie die Konzentration der Nukleinsäure sind die qualitätsbeeinflussenden Faktoren. Eine Bestimmung der einzelnen Parameter zu Beginn einer *Library*-Präparation, ist ein kritischer Schritt der standardisierte Abläufe und Methoden erfordert. Die Kontrolle der Integrität über ein Agarosegel, die Bestimmung der Reinheit mittels NanoDrop, sowie die Messung der Quantität mittels Qubit reichen aus, um eine Aussage über die Qualifizierung einer DNA für die *NGS-Library*-Präparation zu treffen. Die Werte zur Konzentration, Integrität und Reinheit werden als Qualitätsmerkmale im LIMS zu der jeweiligen Probe protokolliert und können bei auffälligen Sequenzierqualitäten zur Fehlersuche herangezogen werden.

Es zeigte sich bei einer Studie an Patienten mit amyotropher Lateralsklerose, das es DNAs gibt, von denen keine qualitativ hochwertige *Library* in Bezug auf die exomische Abdeckung hergestellt werden kann, obwohl sie im Qualifikationsdurchlauf als gut beurteilt wurden. Die DNAs aus dieser Studie wurden uns zugesandt und beim Einsender über ein eigenes DNA-Isolationsprotokoll isoliert. Die 20-fache Abdeckung der Exome lag bei unter 90 %, während bis zu 4,7 % der Zielregion nicht abgedeckt waren. Die DNA-Isolation von DNA aus Blut der Patienten am Institut für Humangenetik mit dem *Chemagic360* und eine erneute Sequenzierung der Proben zeigte, dass aus der am Institut isolierten DNA qualitativ hochwertige Exome erzeugt werden können. Es ist anzunehmen, dass in der DNA, die uns der Einsender zur Verfügung gestellt hat, Inhibitoren waren, die nicht über den Qualifikationslauf detektiert werden konnten, aber die *Library*-Präparation negativ beeinflusst haben.

Die meisten DNAs, die innerhalb der Automation bearbeitet werden, werden über den *Chemagic360* hier am Institut für Humangenetik isoliert. Erfahrungsgemäß zeigen sich hierbei keine Qualitätseinbußen. Bei eingesandter DNA wurde in der Vergangenheit in seltenen Fällen das Phänomen der schlechten Abdeckung,

ähnlich wie bei dem beschriebenen ALS-Projekt, beobachtet. Daher wurde getestet, ob es möglich ist das DNA-Isolat noch einmal säulchen-basiert aufzureinigen, um mögliche Inhibitoren zu entfernen. Es zeigte sich bei der Testung an zwei Proben, dass eine erneute Aufreinigung, *Library*-Präparation und Sequenzierung nicht zu einer Steigerung der exomischen Abdeckung führt. Die beiden Proben haben sehr ähnliche Ergebnisse erbracht.

Abgeleitet durch die Testungen und Erfahrungswerte der vergangenen Jahre konnten Anforderungen abgeleitet werden, die eingesandte DNA erfüllen muss, um daraus eine qualitativ hochwertige *Library* zu erzeugen. Wenn möglich sollte vom Einsender EDTA-Blut zur Verfügung gestellt werden. Ist dies nicht möglich, kann hochmolekulare DNA eingeschickt werden, die keine Anzeichen von Verunreinigungen und Degradation aufweist. Der Qualifikationsablauf kann vom Institut für Humangenetik zur Verfügung gestellt und die Ergebnisse mit dem Einsender ausgetauscht werden. In den seltenen Fällen, bei denen aus eingesandtem DNA-Material keine *Library* mit den entsprechenden Qualitäten hergestellt werden kann, wird ebenfalls das Einsenden von EDTA-Blut empfohlen. Weiterhin sollte bei solchen Fällen getestet werden, ob eine Aufreinigung des DNA-Isolats eine Verbesserung der Sequenzierdaten bewirkt. Die bisherige Testung bei zwei Proben ist nicht aussagekräftig und kann nur zufällig negativ ausgefallen sein.

4.2.2 Methodiken zum Ausschluss von Probenverwechslungen und Kontaminationen bei der *NG*-Sequenzierung

Beim Durchsatz großer Probenmengen ist besonders auf die Identität der Proben zu achten, um die erhaltenen Sequenzierdaten nicht zu verfälschen. Zum Ausschluss von Probenverwechslungen oder Kontaminationen entwickelten wir verschiedene Methoden. Die DNA/RNA-Proben werden in 2D-Barcode Röhrchen auf definierten Stellplätzen gelagert. Die Barcodes werden vor Beginn der *Library*-Präparation verglichen um die Bearbeitung der richtigen Probe sicher zu stellen. Beim Probeneingang findet eine Geschlechtsstimmung über eine Amelogenin-PCR statt und das Ergebnis wird mit den Angaben des Einsenders verglichen. Durch das Arbeiten in 96-well PCR-Platten und dem Einsatz von 8- oder 12-Kanal Pipetten, wird ein Vertauschen der Probe während der *Library*-Präparation minimiert oder kann nachträglich systematisch zugeordnet werden. Die Proben bleiben von Beginn der Präparation bis zum Ende an der gleichen Position einer PCR-Platte. Eine Bestimmung von Probenverwechslungen, vor allem aber Kontaminationen, kann anhand der Sequenzdaten im Anschluss an die Sequenzierung erfolgen. Nach einer statistischen Methode, entwickelt von Jun *et al.* [58], werden die Exom- und Genom-Daten auf Kontaminationen überprüft und der Grad einer Kontamination in Prozent ausgegeben. Zusätzlich dazu erfolgt die Bestimmung der Anzahl der *SRY-Reads*.

Die Entwicklung des LIMS und die Etablierung der automatisierten *Library*-Präparation befähigen uns, im Hochdurchsatz Proben und Projekte mit hoher Qualität und Standardisierung zu bearbeiten. Dennoch wird es in den kommenden Monaten notwendig sein, eine zweite Automation zu implementieren. Zum einen sollen prä- und post-PCR-Prozesse voneinander getrennt werden und zum anderen ermöglicht eine zweite Automation einen noch höheren Durchsatz. Dieser ist notwendig um die Sequenziermaschinen, deren Leistung weiterhin steigen wird, auszulasten. Im Hinblick auf die fallenden Kosten, wird die Genom-Sequenzierung mehr und mehr in den Fokus der Forschung rücken. Dazu wird es notwendig sein ein automatisiertes qPCR-Protokoll zu entwickeln um reproduzierbarer und in größeren Mengen arbeiten zu können. Die Entwicklung einer Automation zur initialen Verdünnung von DNA für eine Exom- oder Genom-*Library*-Präparation wird das Risiko einer Probenverwechslung- und kontamination auf ein Minimum reduzieren und erlaubt die Einführung einer weiteren Standardisierung.

4.3 Anwendung des LIMS und der Automation

Das in dieser Arbeit beschriebene LIMS zusammen mit der Entwicklung der automatisierten *Library*-Präparation und der Ableitung von Qualitätsstandards für eingehendes DNA-Material, bildete die Grundlage für die Anwendung im Bereich der molekulargenetischen Untersuchung. Es wurden zwei Protokolle entwickelt und getestet, die es erlauben weniger DNA, oder FFPE-DNA in die *Library*-Präparation einzusetzen. Zusätzlich dazu wurden die Auswirkungen einer vergrößerten *Insert*-Größe einer *Library* auf die erhaltenen Qualitätsparameter untersucht. Das Protokoll für die Semi-Automation und der weitgehenden Automation der Exom-*Library*-Präparation, wurde für die genetische Untersuchung bei Patienten mit Morbus Parkinson und Burn-McKeown-Syndrom erfolgreich eingesetzt.

4.3.1 Entwicklung neuer Automationsprotokolle für spezielle Anwendungen

Nach der Etablierung der Automation konnten zusätzlich neue Protokolle ausgetestet werden. So wurde ein Protokoll für die Exom-Sequenzierung getestet, welches eine exomische *Library*-Präparation mit weniger Einsatzmenge bei gleichzeitig größerer *Insert*-Größe ermöglichte. Die Nachfrage nach *Library*-Präparationsprotokollen mit weniger als 3000 ng DNA-Startmaterial, kommt vor allem aus den Bereichen der Tumorforschung oder der Pädiatrie, wo gehäuft nicht ausreichend DNA-Material für ein klassisches Exom zur Verfügung steht. Gleichzeitig sollten Optimierungen des bestehenden Standard-Protokolls hinsichtlich der Fragmentlänge einer *Library* vorgenommen werden, um die Anzahl der überlappenden *Reads* einer Sequenzierung von rund 44 % zu minimieren. Überlappende Sequenier-*Reads* bringen keinen Mehrgewinn bei der Suche nach genetischen Varianten und sind redundant.

Protokolle für eine Reduktion der Einsatzmenge innerhalb des *SureSelect*-Protokolls auf 200 ng stehen von Agilent zur Verfügung, behalten aber die Standard-*Insert*-Größe von 200 bp bei. Die vorliegende Arbeit beschreibt die Austestung eines Automationsprotokolls, dass beide Ansprüche umsetzt. Dabei wurden einzelne Protokollschritte aus drei verschiedenen kommerziellen Protokollen so kombiniert, dass eine *Library*-Präparation mit 500 ng, bei einer gleichzeitigen Erhöhung der *Insert*-Größe auf 350 bp möglich war. Im Ergebnis konnte eine exomische

Library präpariert werden. Die Sequenzierung ergab, dass die anvisierte *Insert*-Größe von 350 bp mit durchschnittlich 390 bp überschritten wurde. Gleichzeitig zeigte sich bei der Analyse der Qualitätsparameter der Testproben, dass sich die *On-Bait*-Rate durch die Größenerhöhung um 20 % verringerte und damit einhergehend die durchschnittliche exomische Abdeckung sank. Die Duplikatrate stieg erwartungsgemäß an. Es ist anzunehmen, dass eine mäßige Erhöhung der *Insert*-Größe von 200 bp auf 250 oder 300 bp die exomische Abdeckung sowie die *On-Bait*-Rate erhöht. Diese Hypothese muss zukünftig in weiterführenden Untersuchungen getestet werden. Eine aktuelle Publikation von Krasnenko *et al.* [192] beschreibt eine optimale Anreicherungs-effizienz bei *Insert*-Größen von 250 bis 330 bp. Dazu kommt, dass circa 80 % der humanen Exons kleiner als 200 bp sind [193]. Krasnenko *et al.* schlussfolgern das die optimale *Insert*-Größe zwischen 250 und 300 bp liegt und eine Annäherung an diese Größe die Sequenzier-effizienz maximiert.

Zusätzlich wurde ein zweites Protokoll für eine *Library*-Präparation von FFPE-DNA entwickelt. FFPE-Material wird für die histopathologische Diagnostik angelegt [131] und ist eine wertvolle Ressource für die molekulargenetische Diagnostik und Forschung [194]. DNA isoliert aus FFPE-Gewebe hat den großen Nachteil, dass sie durch den Fixierungs- und Lagerungsprozess häufig degradiert vorliegt. Zudem kommt es zu Quervernetzungen der DNA, abasischen Seiten durch Depurierung und im Folgenden zu lokalen DNA-Denaturierungen, Strangbrüchen und Cytosin-Desaminierungen [131]. FFPE-Material wurde in der Vergangenheit erfolgreich für Amplikon-basierte Analysen oder *Panel*-Analysen genutzt [195]. Die vorliegende Arbeit beschreibt die technische Machbarkeit einer Exom-*Library*-Präparation aus FFPE-isolierter DNA. Die DNA stammt auf FFPE-Tumoren von Patienten mit Kopf-Hals Tumoren.

In einem initialen Schritt wurde bestimmt, welche Qualität FFPE-DNA aufweisen muss um daraus eine *Library* zu generieren. Es wurde eine Einteilung in Qualitäten auf Grundlage der Fragmentverteilung der jeweiligen DNA vorgenommen. Die Basis dazu war die Messung auf dem Bioanalyzer (DNA 12000) und gliederte sich in gut, mäßig und schlecht. Die DNA-Quantität wurde fluoreszenzbasiert (Qubit) bestimmt. Da auch die Menge des Materials limitierend sein kann, wurde zusätzlich getestet, welche Einsatzmenge für die *Library*-Präparation notwendig ist. Es konnte festgelegt werden, dass 1500 ng einer qualitativ guten

FFPE-DNA ausreichend für eine Präparation sind, während für mäßige und qualitativ schlechte FFPE-DNA 3000 ng Einsatzmenge notwendig waren. Die *Library*-Präparation wurde über die weitgehende Automation nach dem Standard-Anreicherungsprotokoll mit angepasster Einsatzmenge durchgeführt. Ein Vergleich der Sequenzier-Qualitäten von DNA aus peripheralem Blut und FFPE-DNA zeigte keine Abweichungen. Lediglich die Duplikatrate war mit 14 % über den Durchschnittswert von 9 % erhöht, während die *Insert*-Größe um 30 bp verringert war. Die Verkleinerung der *Insert*-Größe beschreiben auch Oh *et al.* in ihrer Studie [194].

Die Arbeit zeigt, dass es technisch möglich ist aus FFPE-Material mit verschiedenen Qualitätsstufen eine exomische *Library* zu erzeugen. Besonders interessant wird die zukünftige bioinformatische Auswertung der Proben sein um die Frage zu beantworten ob es gehäuft zu C>T und G>A Mutationsartefakten durch die Desaminierung der DNA kommt.

Es existieren bisher zwei publizierte Studien zur Exom-Sequenzierung von FFPE-DNA aus Tumormaterial [195, 194]. Beide beschreiben die Ergebnisse der Austestung eines *Library*-Präparationsprotokolls mit unterschiedlichen Anreicherungstechniken. Die Ergebnisse der Studien sind recht ähnlich in Bezug auf die erhaltene Datenqualität. Beide beschreiben eine hohe Vergleichbarkeit von FFPE-DNA und DNA aus gepaartem gefrorenen Material, wobei die C>T und G>A Mutationsartefakte in beiden Studien deutlich sichtbar werden. Bei stark degradierten FFPE-DNA-Proben wird die Rate der Mutationsartefakte mit 1:1.000 bp angegeben [194]. Es wird empfohlen die Sequenziertiefe von FFPE-DNA auf mindestens 80-fach zu fokussieren um falsch-positive Assoziationen bei der Suche nach somatischen Mutationen zu reduzieren [196, 194].

Astolfi *et al.* [195] etablierten zudem eine *Library*-Präparation mit 100 bis 200 ng Startmaterial (gute Qualität/schlechte Qualität) in Verbindung mit dem *Nextera Rapid Capture Exome Enrichment Kit* (Illumina). Auch sie zeigten, dass es keine Abweichungen der Sequenzierqualität im Vergleich zu nicht FFPE-DNA gibt. Mit der Anwendung anderer Präparationsprotokolle ist es somit möglich die Einsatzmenge auf 100 bis 200 ng zu verringern. Gleichzeitig kann das am Institut für Humangenetik entwickelte Protokoll dahingehend optimiert werden, dass die *Insert*-Größe bei entsprechender FFPE-DNA Qualität leicht erhöht wird, um überlappende *Reads* zu reduzieren.

Marosy *et al.* [197] beschreiben ein Protokoll zur Reparatur von FFPE-DNA vor der eigentlichen *Library*-Präparation mit einem Enzym-Mix um Strangbrüche, Desaminierungen und abasische Seiten zu reduzieren. Sie zeigen, dass durch die Reparatur der DNA die Anzahl amplifizierbarer DNA-Fragmente und individueller Moleküle erhöht. Damit einhergehend kommt es zu einer Verbesserung der Datenqualität in Bezug auf die Anzahl von Mutationsartefakten. Die eigentliche Fragmentierung der FFPE-DNA kann damit allerdings nicht behoben werden [197]. In folgenden Studien sollte zusammen mit der Anpassung der *Insert*-Größe auch die DNA-Reparatur getestet werden.

Neben der Exom-Sequenzierung ist es ebenso vorstellbar auch genomische *Libraries* aus FFPE-DNA zu erstellen, um somatische Varianten in nicht-kodierenden Regionen zu identifizieren. So zeigten Robbe *et al.* [131] kürzlich in einer umfassenden Untersuchung die Ergebnisse einer Genomstudie, bei der 52 FFPE-Proben mit 52 gepaarten Gefrierproben und 52 gepaarten Blutproben von Krebspatienten verglichen wurden (52 Trios). Es wird gezeigt, dass sowohl aus FFPE-Material als auch aus Gefrier-Material die selbe Menge DNA isoliert werden kann. Allerdings ist der Anteil Doppelsträngiger DNA aus FFPE-Material signifikant verringert, während die Reinheit der Proben ähnlich war. Wie zu erwarten war zeigte die FFPE-DNA exzessive Degradation im Vergleich zu DNA aus Gefrier-Material [131]. Ähnlich wie bei der vorliegenden Arbeit war die *Insert*-Größe der FFPE-Genom-*Libraries* verringert, bei einer gleichzeitigen Reduktion der durchschnittlichen Sequenziertiefe im Vergleich zum Gefrier-Material. Die FFPE-Genome zeigten eine inhomogene Abdeckung über die Chromosomen und eine abweichende Anzahl somatische Mutationen und Indels in beiden Materialtypen [131]. Die Überlappung von Varianten zwischen beiden Test-Sets lag bei 52 % und wird mit die Heterogenität der Tumore und der Probenentnahme begründet. Varianten mit einer Allelfrequenz unter 0,1 konnten nicht nachgewiesen werden und eine Erhöhung der genomischen Abdeckung wird zur Sensitivitätssteigerung empfohlen. Die Detektion von *copy-number*-Varianten war erschwert und wurde durch eine Anpassung der DNA-Extraktionsmethode (Lösen der Quervernetzungen) verbessert. Jedoch musste eine manuelle Untersuchung der Genom-Daten jeder einzelnen Probe erfolgen [131].

Im Ergebnis der Studien zeigt sich, dass die *Library*-Präparation einer FFPE-DNA für die Exom- und Genom-Sequenzierung in Abhängigkeit von der Qualität

der FFPE-DNA möglich ist. Die Herausforderung liegt bei der bioinformatischen Analyse der erhaltenen Sequenzierdaten und der Filterung von Mutationsartefakten. Generell wird empfohlen die Sequenziertiefe zu erhöhen um die Detektion von Varianten zu vereinfachen [131, 196, 194]. Zudem sollten Standard-Prozeduren zur Präparation von FFPE-Material, der DNA-Isolation und der bioinformatischen Auswertung entwickelt werden [131].

4.3.2 Exom-Sequenzierung bei Patienten mit Morbus Parkinson

In der Studie, zur Identifikation der kausativen genetischen Mutation bei Patienten mit Morbus Parkinson, wurde eine große österreichische Familie untersucht die ein autosomal-dominantes Vererbungsmuster der Erkrankung zeigt. Diese Studie steht exemplarisch für viele weitere Exom-Studien, die seither am Institut für Humangenetik durchgeführt wurden, um die ursächlichen genetischen Varianten bei Patienten mit monogenen Erkrankungen zu identifizieren. Wir sequenzierten die Exome zweier Cousins zweiten Grades und ermittelten die Schnittmenge der genetischen Veränderungen in beiden Patienten. Wir priorisierten die gemeinsamen Varianten durch Ausschluss von Alterationen, welche eine Qualität kleiner als 15 hatten, in weniger als 15 % der Reads vorkamen, die mit einer Allelfrequenz von mehr als 2 % in dbSNP auftraten und die in 72 Kontroll-Exomen identifiziert wurden. So minimierten wir die gemeinsamen genetischen Alterationen der beiden Betroffenen auf 10 heterozygote, nicht-synonyme Varianten. Lediglich eine der zehn Varianten, c.1858G>A (p.Asp620Asn) im Kandidatengen *VPS35*, erfüllte die Kriterien für eine mögliche Pathogenität: sie kosegregiert mit der Erkrankung in der Familie und war nicht detektierbar in 680 KORA-S4 Kontrollen. Zusätzlich liegt die Veränderung im Bereich einer gekoppelten Region innerhalb der Familie. Bei der Untersuchung der genetischen Variante c.1858G>A in *VPS35* in 486 unabhängigen M. Parkinson Patienten, identifizierten wir zusätzlich zwei weitere Familien, bei denen die Mutation auftrat, welche ebenfalls innerhalb der Familien mit der Erkrankung kosegregiert. Die Variante c.1858G>A (p.Asp620Asn) wurde nicht in weiteren 1.014 Kontrollen gefunden. Die Aminosäure Asp620 in *VPS35* ist innerhalb verschiedener Spezies evolutionär hoch konserviert und der Austausch der Aminosäure wird durch die Prediktionsprogramme PolyPhen2, SNAP und SIFT als schädlich eingestuft.

VPS35 ist ein evolutionär konserviertes Protein und eine Untereinheit des Retromer Komplex. Das Retromer ist ein Multiprotein-Komplex (Abb.: 61), bestehend aus einem *VPS26-VPS29-VPS35*-Heterodimer zur Erkennung der Fracht (*cargo-selective complex, CSC*) und einem Homo- oder Heterodimer bestehend aus *SNX1* und/oder *SNX2* zur Membranbindung [198, 199].

Das Retromer ist am retrograden Transport von Membranproteinen vom Endosom zurück zum Trans-Golgi-Netzwerk (TGN) beteiligt [198]. Die bisher am besten beschriebenen Fracht-Proteine des Retromers sind Mannose-6-Phosphat

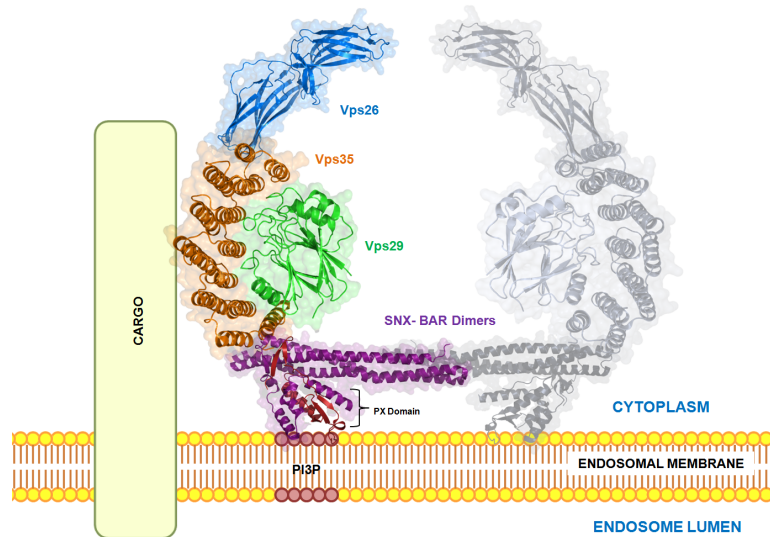


Abbildung 61: Der Retromer Komplex [200]. Das Retromer besteht aus einem VPS26-VPS29-VPS35-Heterodimer und einem Homo- oder Heterodimer bestehend aus SNX1 und/oder SNX2. Dargestellt ist die Bindung des Komplex an ein Frachtmolekül und an die endosomale Membran.

Rezeptoren [201] und Vps10p. Beide Proteine transportieren Hydroxylasen zu den Endosomen und lysosomalen Vakuolen. Sehr viele weitere Membranproteine, wie CIMPR und Sortilin, sind auf den Transport durch das Retromer angewiesen [202, 203, 204]. Ein interessanter Punkt im Zusammenhang mit der vorliegenden Arbeit ist die Beteiligung des Retromers am Transport von SORL1 und Rezeptorproteinen mit einer VPS10P-Domäne, die eine bekannte Rolle in der Entstehung von Alzheimer spielen [205, 206]. Zusätzlich konnte in weiteren Studien gezeigt werden, dass der Retromer Komplex essentiell zur endosomalen Rekrutierung des *WASH*-Komplex (*Wiskott-Aldrich syndrome homologue*) ist [207]. Der *WASH*-Komplex zusammen mit Arp2/3 wiederum ist essentiell für die Bildung von Aktin Strukturen auf dem Endosom um die Protein-Sortierung vom Endosom zur Zelloberfläche zu bewerkstelligen [208]. Membranproteine, die für ihren Transport den *WASH*-Komplex benötigen, sind somit indirekt auch auf die Funktion des Retromers angewiesen [209]. Veränderungen einzelner Komponenten des *WASH*-Komplex wurden bei neurologischen Erkrankungen wie mentaler Retardierung [210] und hereditärer spastischer Paraplegie [211, 212] beschrieben. Die Rekrutierung des *WASH*-Komplex findet über VPS35 im *CSC* des Retromers statt [213, 214, 215]. Tatsächlich konnte in einer kürzlichen Studie gezeigt wer-

den das die von uns identifizierte Variante Asp620Asn in VPS35 die Interaktion zum *WASH*-Komplex über FAM21 destabilisiert, was zu einer reduzierten endosomalen Lokalisation des *WASH*-Komplex führt [209]. Zwei der wichtigsten Transportproteine des Retromer-Komplex sind der Glukosetransporter GLUT 1 und der Kation-unabhängige Mannose 6-Phosphat Rezeptor (CI-MPR). McGough *et al.* [216] zeigten das die von uns identifizierte Mutation Asp620Asn in VPS35 den Transport von GLU1 vom Endosom zur Plasmamembran nicht beeinträchtigt, wohl aber den Transport des CI-MPR vom Endosom zum Trans-Golgi Netzwerk. Sie zeigten zudem, dass die Variante weder den Zusammenbau des Retromers, noch die endosomale Lokalisation von VPS35 beeinflusst. Ähnlich wie Zavodszky *et al.* [209], demonstrieren die Autoren, dass der primäre Effekt der Asp620Asn Mutation, eine Abschwächung der Affinität zu FAM21 ist und als Folge daraus eine reduzierte Bindung an den *WASH*-Komplex und dessen Proteine entstehen [216, 217]. Der Verlust der Bindung zum *WASH*-Komplex führt, durch eine gestörte Bildung der Autophagosomen, zu Defekten in der Autophagozytose der Zelle [209]. Eine beeinträchtigte Autophagozytose ist eine schlüssige Erklärung der Pathogenese von Morbus Parkinson. Der Ausfall des zelleigenen Proteinabbaus führt zu Proteinaggregationen, mitochondrialen Anomalien, erhöhten Konzentrationen an reaktiven Sauerstoffspezies und zu einer gesteigerten Empfindlichkeit gegenüber dem Zelltod.

Die tatsächliche Anzahl der M. Parkinson Fälle, die durch Varianten in *VPS35* entstehen, schätzen wir als eher gering ein. Dennoch eröffnen die Ergebnisse neue Einblicke in die möglichen molekularen Mechanismen der Entstehung des Morbus Parkinson und anderer neurologischer Erkrankungen.

4.3.3 Exom- und Genom-Sequenzierung bei Patienten mit Burn-McKeown-Syndrom

Das Burn-McKeown-Syndrom (BMKS) ist eine seltene Erkrankung, bei der sich die Patienten psychomotorisch normal entwickeln. Betroffene zeigen eine klassische Kombination aus Choanalatresie (Verschluss der hinteren Nasenöffnungen), sensorineuraler Schwerhörigkeit, Herzfehler und typischen kraniofazialen Dysmorphien [167, 168, 169]. Bisher sind weltweit lediglich 14 Betroffene aus elf Familien beschrieben [218]. Das Ziel der Studie war die Aufklärung der genetischen Ursache bei Patienten mit Burn-McKeown-Syndrom in diesen elf Familien.

Wir analysierten mit einer Kombination aus Exom- und Genom-Sequenzierung die DNA in 14 betroffenen Patienten. Die Exom-Sequenzierung bei Patienten ergab in drei der vier untersuchten Familien heterozygote *loss-of-function*-Varianten in *TXNL4A*, die von einem gesunden Elternteil ererbt wurden. Bei weitere Untersuchungen aller Betroffener mittels DNA-Mikroarray wurden in den Patienten sechs Mikrodeletionen identifiziert, die von jeweils einem gesunden Elternteil ererbt wurden. Alle Deletionen überlappen am Genort von *TXNL4A*. Eine MLPA-Analyse ergab zudem eine heterozygote Deletion von Exon 3 im *TXNL4A*-Gen bei einem Patienten. Die darauf folgende Genom-Sequenzierung, zur Suche nach der zweiten genetischen Variante, ergab heterozygote Deletionen im angenommenen Promotorbereich von *TXNL4A*. Es wurde gezeigt, dass sowohl eine *compound*-heterozygote Promoterdeletion (1 Δ), kombiniert mit einer Nullmutation, als auch eine homozygote Promotordeletion (2 Δ) im *TXNL4A*-Genlocus, den größten Teil der Fälle (9 von 11 Familien) mit BMKS verursachen. Die beiden *TXNL4A* negativen Indexpatienten der Familien BMKS003 und BMKS011, wurden initial, aufgrund des überlappenden klinische Phänotyps, als BMKS-Patienten diagnostiziert. Die beiden Mädchen sind klinisch nicht von den *TXNL4A* positiven Patienten zu unterscheiden. Dies unterstützt die Annahme einer genetischen Heterogenität der Erkrankung [94].

Die prognostizierte Frequenz von homozygoten Trägern der Typ 1 Δ -Promotordeletion in der generellen Population von 1:17,300, schließt diesen Genotyp als Ursache für das BMKS aus [94]. Das homozygote 1 Δ -Allel könnte zu einem leichten Phänotyp führen, der möglicherweise der klinischen Untersuchung entgehen kann. Daher sollten größere Kohorten untersucht werden.

Die Ergebnisse zeigen, dass homozygote Typ 2 Δ -Deletionen, wie in der Familie

BMKS008, einen stärkeren Effekt haben als homozygote Typ 1 Δ -Deletionen [94]. Reporter-gen-Assays in HEK293 Zellen und *in vivo* Assays in Hefe zeigen, dass die Promoterdeletionen zu einer verminderten *TXNL4* Expression führen. Untersuchungen an Hefe nach Depletion von *TXNL4A* (*Dib1*) resultierten in der reduzierten Assemblierung des tri-snRNP Komplex, einer Untereinheit des katalytisch aktiven Spleißosoms [94]. Unsere Ergebnisse erweitern die Anzahl, der durch Mutationen in Bestandteilen des Spleißosoms bedingten, erblichen Erkrankungen [94]. Erkrankungen aus dieser Gruppe sind das Nager Syndrom (MIM154400) [219] und die mandibulofaziale Dysostose Typ Guion-Almeida (MIM610536) [219], die ebenso wie das BMKS durch kraniofaziale Fehlbildungen gekennzeichnet sind. Alle drei Syndrome werden durch Mutationen in Komponenten des Spleißosoms verursacht und haben faziale Dysmorphien. Dies führt zu der Annahme, dass die kraniofaziale Entwicklung hoch sensitiv gegenüber Fehlfunktionen des Spleißosoms ist [94]. Wir zeigen in dieser Studie, dass das Burn-McKeown-Syndrom eine autosomal-rezessive Erkrankung ist, die primär durch *compound*-heterozygote Promoterdeletionen in Kombination mit sehr seltenen *loss-of-function*-Mutationen hervorgerufen wird.

Das Burn-McKeown-Syndrom wurde 1992 zum ersten mal beschrieben [169]. 20 Jahre später konnte die genetische Ursache durch die Untersuchung mittels *Next-Generation*-Sequenzierung aufgeklärt werden. Es wird angenommen, dass eine reduzierte *TXNL4A*-Expression das Spleißen bei bestimmten Untergruppen von prä-mRNAs beeinflusst, was zum Gewebe-spezifischen phänotypischen Spektrum der Erkrankung führt [220, 170]. Goos *et al.* konnten kürzlich das Spektrum der kausativen Varianten in *TXNL4A*, bei Patienten mit überlappenden klinischen Symptomen, durch Genom-Sequenzierung erweitern [170]. Sie identifizierten in zwei Familien eine Spleiß-Variante der 3'-Akzeptorstelle strangaufwärts des letzten Exons in *TXNL4A* (c.258-2A>G) zusammen mit der 1 Δ -Deletion im Promotorbereich. Zusätzlich zeigte ein Betroffener einer anderen Familie einen homozygoten Status der 2 Δ -Deletion in *TXNL4A*. In einer vierten Familie identifizierten sie die 1 Δ -Promotordeletion in *TXNL4A* im homozygoten Status. Ferner wurde gezeigt, dass die RNA-Expression durch die 1 Δ -Promotordeletion um 69 % reduziert war [170]. Die Ergebnisse bestätigen und erweitern die Resultate der vorliegenden Studie.

Die Ergebnisse der vorliegenden Arbeit zeigen sowohl die Möglichkeiten der

Exom-Sequenzierung, als auch ihre Limitierung. Nur durch eine Sequenzierung des gesamten Genoms konnten wir die pathogenen Deletionen in der Promotorregion von *TXNL4A* bei Patienten mit BMKS identifizieren. Die Exom-Sequenzierung gibt uns einen Einblick in die wichtigen 1.5 % kodierende genetische Sequenz, enthält uns jedoch regulatorische Bereiche und Abschnitte mit unbekannter Signifikanz vor. Der große Vorteil einer Exom-Sequenzierung ist die kostengünstige und gleichzeitig erfolgversprechende Anwendung zur Identifikation von kausativen genetischen Varianten. Die, im Vergleich zur Genom-Sequenzierung, niedrigen Kosten erlauben dabei Studien mit hohen Patientenzahlen. Die Schwachstellen der Exom-Sequenzierung können durch eine Genom-Sequenzierung ausgeglichen werden. Genom-Daten erlauben eine Analyse von genomischen Varianten in kodierenden und nicht-kodierenden Bereichen, haben eine bessere und gleichmäßigere Abdeckung, haben weniger Verzerrung durch GC-reiche Sequenzen, benötigen weniger Abdeckung und können von jeder Spezies erzeugt werden [42]. Liegt bei einem klaren Studiendesign, einer sehr wahrscheinlichen genetischen Krankheitsursache und einer homogenen Patientenpopulation nach der Exom-Analyse ein negativer Befund vor, lohnt sich die Investition in eine Genom-Sequenzierung.

Die vorliegende Studie zeigt auch Schwachstellen der Genom-Sequenzierung auf, die in der Art der *Library*-Herstellung begründet sind. Der Promotorbereich von *TXNL4A* ist sehr GC-reich. Es zeigte sich, dass die Promotorregionen von *TXNL4A* in einem internen Referenzset von Genomen, die mittels PCR-Schritt innerhalb des *Library*-Präparationsprozesses angereichert wurden, nicht repräsentiert sind. Durch die Anwendung eines PCR-freien *Library*-Präparationsprotokolls konnte die Region abgedeckt und analysiert werden. Es wird daher empfohlen auf den PCR-Anreicherungsschritt innerhalb der *Library*-Herstellung zu verzichten um GC-reiche Genomabschnitte darstellen zu können.

4.4 Fazit und Ausblick

Die vorliegende Arbeit präsentiert die erfolgreiche Etablierung und Anwendung eines *NGS*-LIMS, sowie der Exom- und Genom-Sequenzierung zur Identifizierung pathogener Genvarianten bei Patienten mit monogenen Erkrankungen. Durch den Gebrauch gezielter Priorisierungsstrategien gelang die Identifizierung der kausativen Mutation unter Tausenden genetischen Veränderungen.

Ein wichtiges Ergebnis der vorliegenden Arbeit ist die Demonstration der hohen Sensitivität der Exom-Sequenzierung, als Grundlage für die Durchführung *NGS*-basierter Diagnostik im klinischen Ablauf. Mit der Exom-Sequenzierung können molekulargenetische Fragestellungen bei Patienten schneller, günstiger und umfassender, als mit herkömmlichen molekulargenetischen Methoden bearbeitet werden. So kann die bisher durchgeführte Stufendiagnostik durch die Exom-Analyse abgelöst werden. Vor allem bei stark heterogenen Erkrankung, wie der mentalen Retardierung oder Autismus, bietet die DNA-Sequenzierung über eine *NGS*-Plattform eine hervorragende Möglichkeit zur Identifikation möglicher pathogener genetischer Varianten. Eine gesicherte Diagnose ist erleichternd für die Patienten und die Angehörigen, unterstützt die gezielte Therapie und hilft dem behandelten Mediziner bei der Prognoseabschätzung.

Durch die stetig fallenden Kosten einer Hochdurchsatzsequenzierung wird die Exom-Sequenzierung in den kommenden Jahren von der Genom-Sequenzierung abgelöst. Seit dem Jahr 2016 ist es möglich, ein gesamtes Genom für 1.000 USD zu sequenzieren [15]. Ein Genom beherbergt sehr viele genetische Varianten im Vergleich zum Referenzgenom. Wir haben nur sehr wenig Wissen über die Bereiche außerhalb von Genen und deren Funktion. Die damit einhergehende Schwierigkeit, aus den unzähligen, natürlich vorkommenden genetischen Varianten die ursächliche zu filtern, wird in den nächsten Jahren die Hauptaufgabe sein. Eine erste Ergänzung zu den bekannten Priorisierungsstrategien von genomischen Varianten ist die gleichzeitige Analyse des Transkriptom, um den möglichen Effekt einer genetischen Variante auf die RNA-Expression und/oder RNA-Prozessierung sichtbar zu machen [221]. Vor allem heterogene Erkrankungen, Erkrankungen mit geringer Penetranz und solche, die durch genetische Veränderungen in nicht-kodierenden Bereichen verursacht werden, können in Zukunft untersucht werden. Zusätzlich rücken all jene Patienten in den Fokus, bei denen die Exom-Sequenzierung keine schlüssigen Ergebnisse erbracht hat.

Neben Fortschritten in der Datenauswertung wird die zweite Generation der Sequenzierung, so wie sie in dieser Arbeit angewandt wurde, in den kommenden Jahren durch die dritte Generation Sequenzieretechniken abgelöst. Mit ihnen entfällt der Amplifikationsschritt (*Bridge-* oder *Emulsions-PCR*) vor der Sequenzierung und es werden die individuellen DNA-Fragmente in Echtzeit gelesen. Durch den Wegfall der bisher nötigen Amplifikation der einzelnen Moleküle zur Potenzierung des Signals für eine erfolgreiche Detektion der eingebauten Nukleotide, entstehen keine PCR-bedingten Artefakte, die eine möglicherweise falsche Assoziation oder Verzerrungen der Daten herbeiführen. Zusätzlich ergeben die Sequenzierungen auf Geräten der dritten Generation eine homogene Abdeckung der einzelnen DNA-Fragmente unabhängig von ihrer Nukleotidkomposition. Zudem versprechen die neuen Plattformen eine weitaus höhere Leselänge, als dies bisher möglich ist. So ist mit dem *PACBIO RS II* (Pacific Biosciences) eine Leselänge von bis zu 20 kb möglich, während Oxford Nanopore mit ihrem *MinION* eine Leselänge von mehreren 100 kb verspricht.

Bei der Genom-Sequenzierung auf einer der Illumina Sequenzier-Plattformen, werden Fragmente von 150 bp Länge gelesen. Eine Analyse von größeren strukturellen Variationen oder repetitiven Sequenzen innerhalb des Genoms wird dadurch erschwert. Strukturelle Variationen zählen mit 1,2 % zu den häufigsten genetischen Veränderungen in einem humanen Genom im Vergleich zu Basenpaarsubstitutionen (0,1 %) [222]. Die Aufklärung solcher Veränderungen, die vor allem im Bereich der Krebsforschung enorme Wichtigkeit hat, kann, wie kürzlich gezeigt, durch lange Leselängen erreicht werden [223, 224]. Auch für *de novo* Assemblierungen von Genomen sind große Leselängen vorteilhaft. Der Einsatz von kombinierten *Library*-Präparationen und verschiedenen Sequenzier-Plattformen, in Abhängigkeit von der Fragestellung der Studie, kann die Schwachstellen der einzelnen Methodiken überbrücken.

Neben den Techniken der Exom- und Genom-Sequenzierung zur Identifikation von genetischen Mutationsereignissen ist das Anwendungsspektrum der *NGS*-Technologie noch weitaus größer [225]. Neben der Analyse von DNA ist die Untersuchung von RNA und die Erstellung von differentiellen Expressionprofilen in mutierten und nicht-mutierten Organismen, oder bei Organismen mit spezieller Behandlung (Diäten, Medikamente), eine weitere Möglichkeit zur Anwendung der *NG*-Sequenzierung [226]. Zusätzlich können neue RNA-Varianten und Spleiß-

stellen identifiziert werden. Zur Untersuchung von Proteinbindestellen, wie zum Beispiel Transkriptionsfaktoren, auf der DNA kann eine klassische Chromatin-Immunopräzipitation (ChIP) durchgeführt und anschließend die angereicherte DNA über den Hochdurchsatz sequenziert werden (ChIP-Seq). Eine relativ neue Methodik ist die Analyse des Translatoms nach einem *Ribosome Profiling*, bei dem Translationsstartpunkte und Syntheseraten von Proteinen definiert werden können, nachdem die aktiv an Ribosomen gebundenen mRNA-Moleküle extrahiert und sequenziert werden [227]. Für epigenetische Studien ist es möglich, aus den Daten einer *NG*-Sequenzierung DNA-oder RNA-Methylierungsprofile zu erstellen (MeDIP-Seq, MeRIP-Seq). Und auch Methylierungsmuster von tRNAs [228] können nach einer Bisulfit-Sequenzierung erzeugt und analysiert werden. Neue Einblicke über die Konformation von Chromosomen oder des gesamten Genoms in der Interphase des Zellzyklus gibt die Kombination aus HiC-Technik [229] und *NG*-Sequenzierung. Dabei kann solch eine Studie Aufschluss über die molekularen Grundlagen der DNA-Replikation, Genregulation, DNA-Reparatur und von Rekombinationsvorgängen geben. Die Sequenzierung von RNA oder DNA einer einzelnen Zelle aus einer heterogenen Zellpopulation (scRNA-seq, scDNA-seq) gibt Aufschluss über die Funktion individueller Zellgruppen. Vor allem im Bereich der Krebsforschung oder bei genetischen Mosaiken, sowie bei der Untersuchung zur Neurogenese ist eine Auflösung bis auf Zellebene interessant.

All diese Anwendungen zeigen das enorme Potential der *Next-Generation*-Sequenzieretechnologie und geben einen Vorstellung auf mögliche Forschungsprojekte am Institut für Humangenetik in den kommenden Jahren.

Literatur

- [1] Eric S. Lander. Initial impact of the sequencing of the human genome. *Nature*, 470(7333):187–197, Feb 2011.
- [2] K Neveling and A Hoischen. Exom-sequenzierung zur identifizierung von krankheitsgenen. *medizinische genetik*, 24(1):4–11, 2012.
- [3] Michael J. Bamshad, Sarah B. Ng, Abigail W. Bigham, Holly K. Tabor, Mary J. Emond, Deborah A. Nickerson, and Jay Shendure. Exome sequencing as a tool for mendelian disease gene discovery. *Nat Rev Genet*, 12(11):745–755, Nov 2011.
- [4] D. E. Reich and E. S. Lander. On the allelic spectrum of human disease. *Trends Genet*, 17(9):502–510, Sep 2001.
- [5] Stylianos E. Antonarakis and Jacques S. Beckmann. Mendelian disorders deserve more attention. *Nat Rev Genet*, 7(4):277–282, Apr 2006.
- [6] Ada Hamosh, Alan F Scott, Joanna S Amberger, Carol A Bocchini, and Victor A McKusick. Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*, 33(suppl 1):D514–D517, 2005.
- [7] Jocelyn Kaiser. Affordable'exomes' fill gaps in a catalog of rare diseases, 2010.
- [8] A.W. Kuss. Next generation sequencing in der humangenetik. *Medizinische Genetik*, 2014.
- [9] N. E. MORTON. Sequential tests for the detection of linkage. *Am J Hum Genet*, 7(3):277–318, Sep 1955.
- [10] James E. Hilbert, Tetsuo Ashizawa, John W. Day, Elizabeth A. Luebbe, William B. Martens, Michael P. McDermott, Rabi Tawil, Charles A. Thornton, and Richard T Moxley, 3rd. Diagnostic odyssey of patients with myotonic dystrophy. *J Neurol*, 260(10):2497–2504, Oct 2013.
- [11] Saskia Biskup. Next generation sequencing in genetic diagnostics 1. *LaboratoriumsMedizin*, 34(6), 2010.

-
- [12] Vandana Shashi, Allyn McConkie-Rosell, Bruce Rosell, Kelly Schoch, Kasturi Vellore, Marie McDonald, Yong-Hui Jiang, Pingxing Xie, Anna Need, David B. Goldstein, and David G. Goldstein. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med*, 16(2):176–182, Feb 2014.
- [13] Sebastian Köhler, Sandra C. Doelken, Christopher J. Mungall, Sebastian Bauer, Helen V. Firth, Isabelle Bailleul-Forestier, Graeme C M. Black, Danielle L. Brown, Michael Brudno, Jennifer Campbell, David R. FitzPatrick, Janan T. Eppig, Andrew P. Jackson, Kathleen Freson, Marta Girdea, Ingo Helbig, Jane A. Hurst, Johanna Jähn, Laird G. Jackson, Anne M. Kelly, David H. Ledbetter, Sahar Mansour, Christa L. Martin, Celia Moss, Andrew Mumford, Willem H. Ouwehand, Soo-Mi Park, Erin Rooney Riggs, Richard H. Scott, Sanjay Sisodiya, Steven Van Vooren, Ronald J. Wapner, Andrew O M. Wilkie, Caroline F. Wright, Anneke T. Vulto-van Silfhout, Nicole de Leeuw, Bert B A. de Vries, Nicole L. Washington, Cynthia L. Smith, Monte Westerfield, Paul Schofield, Barbara J. Ruef, Georgios V. Gkoutos, Melissa Haendel, Damian Smedley, Suzanna E. Lewis, and Peter N. Robinson. The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res*, 42(Database issue):D966–D974, Jan 2014.
- [14] F. Sanger, S. Nicklen, and A. R. Coulson. Dna sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74(12):5463–5467, Dec 1977.
- [15] K Neveling and A Hoischen. Einführung in die grundlagen der hochdurchsatzsequenzierung. *medizinische genetik*, 26(2):231–238, 2014.
- [16] Illumina. Illumina inc.. <https://www.illumina.com/informatics/research/sequencing-data-analysis-management/genomics-lims.html>.
- [17] Shankar Balasubramanian, David Klenerman, and David Bentley. Arrayed biomolecules and their use in sequencing, September 7 2004. US Patent 6,787,308.

-
- [18] David R Bentley, Shankar Balasubramanian, Harold P Swerdlow, Geoffrey P Smith, John Milton, Clive G Brown, Kevin P Hall, Dirk J Evers, Colin L Barnes, Helen R Bignell, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, 456(7218):53, 2008.
- [19] Rahul Sinha, Geoff Stanley, Gunsagar Singh Gulati, Camille Ezran, Kyle Joseph Travaglini, Eric Wei, Charles Kwok Fai Chan, Ahmad N Nabhan, Tianying Su, Rachel Marie Morganti, et al. Index switching causes spreading-of-signal among multiplexed samples in illumina hiseq 4000 dna sequencing. *bioRxiv*, page 125724, 2017.
- [20] Jay Shendure, Gregory J. Porreca, Nikos B. Reppas, Xiaoxia Lin, John P. McCutcheon, Abraham M. Rosenbaum, Michael D. Wang, Kun Zhang, Robi D. Mitra, and George M. Church. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741):1728–1732, Sep 2005.
- [21] Illumina. *Illumina Inc.. Genome Analyzer Iix - Illumina*, pub. no. 770-2009-034 current as of 11 january 2010 edition, Janury 2010.
- [22] Peter D. Stenson, Edward V. Ball, Katy Howells, Andrew D. Phillips, Matthew Mort, and David N. Cooper. The human gene mutation database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics*, 4(2):69–72, Dec 2009.
- [23] Murim Choi, Ute I. Scholl, Weizhen Ji, Tiewen Liu, Irina R. Tikhonova, Paul Zumbo, Ahmet Nayir, Ay?in Bakkalo?lu, Seza Ozen, Sami Sanjad, Carol Nelson-Williams, Anita Farhi, Shrikant Mane, and Richard P. Lifton. Genetic diagnosis by whole exome capture and massively parallel dna sequencing. *Proc Natl Acad Sci U S A*, 106(45):19096–19101, Nov 2009.
- [24] R Weißmann and C Gilissen. Ngs datenanalyse und qualitätskontrolle. *medizinische genetik*, 26(2):239–245, 2014.
- [25] Sarah B. Ng, Kati J. Buckingham, Choli Lee, Abigail W. Bigham, Holly K. Tabor, Karin M. Dent, Chad D. Huff, Paul T. Shannon, Ethylin Wang

-
- Jabs, Deborah A. Nickerson, Jay Shendure, and Michael J. Bamshad. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 42(1):30–35, Jan 2010.
- [26] Fredrik Dahl, Johan Stenberg, Simon Fredriksson, Katrina Welch, Michael Zhang, Mats Nilsson, David Bicknell, Walter F. Bodmer, Ronald W. Davis, and Hanlee Ji. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc Natl Acad Sci U S A*, 104(22):9387–9392, May 2007.
- [27] Simon Fredriksson, Johan Banér, Fredrik Dahl, Angela Chu, Hanlee Ji, Katrina Welch, and Ronald W. Davis. Multiplex amplification of all coding sequences within 10 cancer genes by gene-collector. *Nucleic Acids Res*, 35(7):e47, 2007.
- [28] Andreas Gnirke, Alexandre Melnikov, Jared Maguire, Peter Rogov, Emily M LeProust, William Brockman, Timothy Fennell, Georgia Giannoukos, Sheila Fisher, Carsten Russ, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology*, 27(2):182–189, 2009.
- [29] David T. Okou, Karyn Meltz Steinberg, Christina Middle, David J. Cutler, Thomas J. Albert, and Michael E. Zwick. Microarray-based genomic selection for high-throughput resequencing. *Nat Methods*, 4(11):907–909, Nov 2007.
- [30] Gregory J. Porreca, Kun Zhang, Jin Billy Li, Bin Xie, Derek Austin, Sara L. Vassallo, Emily M. LeProust, Bill J. Peck, Christopher J. Emig, Fredrik Dahl, Yuan Gao, George M. Church, and Jay Shendure. Multiplex amplification of large sets of human exons. *Nat Methods*, 4(11):931–936, Nov 2007.
- [31] Thomas J. Albert, Michael N. Molla, Donna M. Muzny, Lynne Nazareth, David Wheeler, Xingzhi Song, Todd A. Richmond, Chris M. Middle, Matthew J. Rodesch, Charles J. Packard, George M. Weinstock, and Richard A. Gibbs. Direct selection of human genomic loci by microarray hybridization. *Nat Methods*, 4(11):903–905, Nov 2007.

-
- [32] Emily H. Turner, Choli Lee, Sarah B. Ng, Deborah A. Nickerson, and Jay Shendure. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods*, 6(5):315–316, May 2009.
- [33] Emily H. Turner, Sarah B. Ng, Deborah A. Nickerson, and Jay Shendure. Methods for genomic partitioning. *Annu Rev Genomics Hum Genet*, 10:263–284, 2009.
- [34] Lira Mamanova, Alison J. Coffey, Carol E. Scott, Iwanka Kozarewa, Emily H. Turner, Akash Kumar, Eleanor Howard, Jay Shendure, and Daniel J. Turner. Target-enrichment strategies for next-generation sequencing. *Nat Methods*, 7(2):111–118, Feb 2010.
- [35] Matthew N. Bainbridge, Min Wang, Daniel L. Burgess, Christie Kovar, Matthew J. Rodesch, Mark D’Ascenzo, Jacob Kitzman, Yuan-Qing Wu, Irene Newsham, Todd A. Richmond, Jeffrey A. Jeddloh, Donna Muzny, Thomas J. Albert, and Richard A. Gibbs. Whole exome capture in solution with 3 gbp of data. *Genome Biol*, 11(6):R62, 2010.
- [36] Rachel Marine, Shawn W. Polson, Jacques Ravel, Graham Hatfull, Daniel Russell, Matthew Sullivan, Fraz Syed, Michael Dumas, and K Eric Wommack. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of dna. *Appl Environ Microbiol*, 77(22):8071–8079, Nov 2011.
- [37] Chandra Sekhar Reddy Chilamakuri, Susanne Lorenz, Mohammed-Amin Madoui, Daniel Vodák, Jinchang Sun, Eivind Hovig, Ola Myklebost, and Leonardo A. Meza-Zepeda. Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*, 15:449, 2014.
- [38] Agilent. Sureselect - how it works; <http://www.genomics.agilent.com/article.jsp?pageid=3083>.
- [39] Kim D. Pruitt, Tatiana Tatusova, and Donna R. Maglott. Ncbi reference sequences (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*, 35(Database issue):D61–D65, Jan 2007.

-
- [40] Paul Flicek, M Ridwan Amode, Daniel Barrell, Kathryn Beal, Simon Brent, Denise Carvalho-Silva, Peter Clapham, Guy Coates, Susan Fairley, Stephen Fitzgerald, Laurent Gil, Leo Gordon, Maurice Hendrix, Thibaut Hourlier, Nathan Johnson, Andreas K. Kähäri, Damian Keefe, Stephen Keenan, Rhoda Kinsella, Monika Komorowska, Gautier Koscielny, Eugene Kulesha, Pontus Larsson, Ian Longden, William McLaren, Matthieu Muffato, Bert Overduin, Miguel Pignatelli, Bethan Pritchard, Harpreet Singh Riat, Graham R S. Ritchie, Magali Ruffier, Michael Schuster, Daniel Sobral, Y Amy Tang, Kieron Taylor, Stephen Trevanion, Jana Vandrovcova, Simon White, Mark Wilson, Steven P. Wilder, Bronwen L. Aken, Ewan Birney, Fiona Cunningham, Ian Dunham, Richard Durbin, Xosé M. Fernández-Suarez, Jennifer Harrow, Javier Herrero, Tim J P. Hubbard, Anne Parker, Glenn Proctor, Giulietta Spudich, Jan Vogel, Andy Yates, Amonida Zadissa, and Stephen M J. Searle. Ensembl 2012. *Nucleic Acids Res*, 40(Database issue):D84–D90, Jan 2012.
- [41] Kim D. Pruitt, Jennifer Harrow, Rachel A. Harte, Craig Wallin, Mark Diekhans, Donna R. Maglott, Steve Searle, Catherine M. Farrell, Jane E. Loveland, Barbara J. Ruff, Elizabeth Hart, Marie-Marthe Suner, Melissa J. Landrum, Bronwen Aken, Sarah Ayling, Robert Baertsch, Julio Fernandez-Banet, Joshua L. Cherry, Val Curwen, Michael Dicuccio, Manolis Kellis, Jennifer Lee, Michael F. Lin, Michael Schuster, Andrew Shkeda, Clara Amid, Garth Brown, Oksana Dukhanina, Adam Frankish, Jennifer Hart, Bonnie L. Maidak, Jonathan Mudge, Michael R. Murphy, Terence Murphy, Jeena Rajan, Bhanu Rajput, Lillian D. Riddick, Catherine Snow, Charles Steward, David Webb, Janet A. Weber, Laurens Wilming, Wenyu Wu, Ewan Birney, David Haussler, Tim Hubbard, James Ostell, Richard Durbin, and David Lipman. The consensus coding sequence (ccds) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res*, 19(7):1316–1323, Jul 2009.
- [42] Genohub. <https://genohub.com/exome-sequencing-library-preparation/>. online homepage.
- [43] Daichi Shigemizu, Yukihide Momozawa, Testuo Abe, Takashi Morizono, Keith A Boroevich, Sadaaki Takata, Kyota Ashikawa, Michiaki Kubo, and

-
- Tatsuhiko Tsunoda. Performance comparison of four commercial human whole-exome capture platforms. *Scientific reports*, 5:12742, 2015.
- [44] Rainer Jonak. Labor-informations- und management-systeme. *GIT Labor-Fachzeitschrift*, 2015(7):13–17, 2015.
- [45] Piergiorgio Palla, Gianfranco Frau, Laura Vargiu, and Patricia Rodriguez-Tomé. Qtreds: a ruby on rails-based platform for omics laboratories. *BMC bioinformatics*, 15(1):S13, 2014.
- [46] Gerst A Gibbon. A brief history of lims. *Laboratory Automation & Information Management*, 32(1):1–5, 1996.
- [47] Pittsburgh conference final program, 1973, papers 39-45 and 1-8. 1973.
- [48] Perkin-Elmer. Laboratory information management is more than a great idea. for perkin-elmer customers, its a magnificent reality. <https://pubs.acs.org/doi/pdf/10.1021/ac00271a756>, 1983.
- [49] Thomas Triplet and Gregory Butler. The enzymetracker: an open-source laboratory information management system for sample tracking. *BMC bioinformatics*, 13(1):15, 2012.
- [50] Kurtiss J Gordon. Spreadsheet or database: Which makes more sense? *Journal of Computing in Higher Education*, 10(2):111–116, 1999.
- [51] Christopher S Tang. The science of decision making: A problem-based approach using excel.(book reviews). *IIE Transactions*, 35(3):326–328, 2003.
- [52] Yalan Chen, Yuxin Lin, Xuye Yuan, and Bairong Shen. Lims and clinical data management. In *Translational Biomedical Informatics*, pages 225–239. Springer, 2016.
- [53] Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14):1754–1760, Jul 2009.
- [54] David B. Goldstein, Andrew Allen, Jonathan Keebler, Elliott H. Margulies, Steven Petrou, Slavé Petrovski, and Shamil Sunyaev. Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet*, 14(7):460–470, Jul 2013.

-
- [55] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup . The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, Aug 2009.
- [56] Mark A DePristo, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, Christopher Hartl, Anthony A Philippakis, Guillermo Del Angel, Manuel A Rivas, Matt Hanna, et al. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature genetics*, 43(5):491, 2011.
- [57] Vincent Plagnol, James Curtis, Michael Epstein, Kin Y. Mok, Emma Stebbings, Sofia Grigoriadou, Nicholas W. Wood, Sophie Hambleton, Siobhan O. Burns, Adrian J. Thrasher, Dinakantha Kumararatne, Rainer Doffinger, and Sergey Nejentsev. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, 28(21):2747–2754, Nov 2012.
- [58] Goo Jun, Matthew Flickinger, Kurt N Hetrick, Jane M Romm, Kimberly F Doheny, Gonçalo R Abecasis, Michael Boehnke, and Hyun Min Kang. Detecting and estimating contamination of human dna samples in sequencing and array-based genotype data. *The American Journal of Human Genetics*, 91(5):839–848, 2012.
- [59] Eugene V Davydov, David L Goode, Marina Sirota, Gregory M Cooper, Arend Sidow, and Serafim Batzoglou. Identifying a high fraction of the human genome to be under selective constraint using gerp++. *PLoS computational biology*, 6(12):e1001025, 2010.
- [60] Katherine S. Pollard, Melissa J. Hubisz, Kate R. Rosenbloom, and Adam Siepel. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*, 20(1):110–121, Jan 2010.
- [61] R Grantham. Amino acid difference formula to help explain protein evolution. *Science*, 185(4154):862–864, 1974.

-
- [62] Prateek Kumar, Steven Henikoff, and Pauline C Ng. Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. *Nature protocols*, 4(7):1073, 2009.
- [63] Ivan A. Adzhubei, Steffen Schmidt, Leonid Peshkin, Vasily E. Ramensky, Anna Gerasimova, Peer Bork, Alexey S. Kondrashov, and Shamil R. Sunyaev. A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4):248–249, Apr 2010.
- [64] M. Krawczak and D. N. Cooper. The human gene mutation database. *Trends Genet*, 13(3):121–122, Mar 1997.
- [65] S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin. dbsnp: the ncbi database of genetic variation. *Nucleic Acids Res*, 29(1):308–311, Jan 2001.
- [66] NHLBI GO Exome Sequencing Project (ESH). Exome variant server, <http://evs.gs.washington.edu/evs/>.
- [67] Monkol Lek, Konrad J Karczewski, Eric V Minikel, Kaitlin E Samocha, Eric Banks, Timothy Fennell, Anne H Donnell-Luria, James S Ware, Andrew J Hill, Beryl B Cummings, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616):285–291, 2016.
- [68] Tom Strachan and Andrew P. Read. *Molekulare Humangentik*, volume 3. Spektrum Akademischer Verlag, 2005.
- [69] Alexander Zimprich, Anna Benet-Pagès, Walter Struhal, Elisabeth Graf, Sebastian H Eck, Marc N Offman, Dietrich Haubenberger, Sabine Spielberger, Eva C Schulte, Peter Lichtner, Shaila C Rossle, Norman Klopp, Elisabeth Wolf, Klaus Seppi, Walter Pirker, Stefan Presslauer, Brit Mollenhauer, Regina Katzenschlager, Thomas Foki, Christoph Hotzy, Eva Reinthaler, Ashot Harutyunyan, Robert Kralovics, Annette Peters, Fritz Zimprich, Thomas Brücke, Werner Poewe, Eduard Auff, Claudia Trenkwalder, Burkhard Rost, Gerhard Ransmayr, Juliane Winkelmann, Thomas Meitinger, and Tim M Strom. A mutation in vps35, encoding a subunit of the retromer complex, causes late-onset parkinson disease. *Am J Hum Genet*, 89(1):168–175, Jul 2011.

-
- [70] Jeanette Erdmann, Klaus Stark, Ulrike B. Esslinger, Philipp Moritz Rumpf, Doris Koesling, Cor de Wit, Frank J. Kaiser, Diana Braunholz, Anja Medack, Marcus Fischer, Martina E. Zimmermann, Stephanie Tennstedt, Elisabeth Graf, Sebastian Eck, Zouhair Aherrahrou, Janja Nahrstaedt, Christina Willenborg, Petra Bruse, Ingrid Brænne, Markus M. Nöthen, Per Hofmann, Peter S. Braund, Evanthia Mergia, Wibke Reinhard, Christof Burgdorf, Stefan Schreiber, Anthony J. Balmforth, Alistair S. Hall, Lars Bertram, Elisabeth Steinhagen-Thiessen, Shu-Chen Li, Winfried März, Muredach Reilly, Sekar Kathiresan, Ruth McPherson, Ulrich Walter, C. A. R. D. IoG. R. A. M. , Jurg Ott, Nilesh J. Samani, Tim M. Strom, Thomas Meitinger, Christian Hengstenberg, and Heribert Schunkert. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature*, 504(7480):432–436, Dec 2013.
- [71] Lisenka E L M. Vissers, Joep de Ligt, Christian Gilissen, Irene Janssen, Marloes Steehouwer, Petra de Vries, Bart van Lier, Peer Arts, Nienke Wieskamp, Marisol del Rosario, Bregje W M. van Bon, Alexander Hoischen, Bert B A. de Vries, Han G. Brunner, and Joris A. Veltman. A de novo paradigm for mental retardation. *Nat Genet*, 42(12):1109–1112, Dec 2010.
- [72] Benjamin M. Neale, Yan Kou, Li Liu, Avi Ma’ayan, Kaitlin E. Samocha, Aniko Sabo, Chiao-Feng Lin, Christine Stevens, Li-San Wang, Vladimir Makarov, Paz Polak, Seungtae Yoon, Jared Maguire, Emily L. Crawford, Nicholas G. Campbell, Evan T. Geller, Otto Valladares, Chad Schafer, Han Liu, Tuo Zhao, Guiqing Cai, Jayon Lihm, Ruth Dannenfelser, Omar Jabbado, Zuleyma Peralta, Uma Nagaswamy, Donna Muzny, Jeffrey G. Reid, Irene Newsham, Yuanqing Wu, Lora Lewis, Yi Han, Benjamin F. Voight, Elaine Lim, Elizabeth Rossin, Andrew Kirby, Jason Flannick, Menachem Fromer, Khalid Shakir, Tim Fennell, Kiran Garimella, Eric Banks, Ryan Poplin, Stacey Gabriel, Mark DePristo, Jack R. Wimbish, Braden E. Boone, Shawn E. Levy, Catalina Betancur, Shamil Sunyaev, Eric Boerwinkle, Joseph D. Buxbaum, Edwin H Cook, Jr, Bernie Devlin, Richard A. Gibbs, Kathryn Roeder, Gerard D. Schellenberg, James S. Sutcliffe, and Mark J. Daly. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, 485(7397):242–245, May 2012.

-
- [73] Brian J. O’Roak, Pelagia Deriziotis, Choli Lee, Laura Vives, Jerrod J. Schwartz, Santhosh Girirajan, Emre Karakoc, Alexandra P. Mackenzie, Sarah B. Ng, Carl Baker, Mark J. Rieder, Deborah A. Nickerson, Raphael Bernier, Simon E. Fisher, Jay Shendure, and Evan E. Eichler. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet*, 43(6):585–589, Jun 2011.
- [74] Simon L. Girard, Julie Gauthier, Anne Noreau, Lan Xiong, Sirui Zhou, Loubna Jouan, Alexandre Dionne-Laporte, Dan Spiegelman, Edouard Henrion, Ousmane Diallo, Pascale Thibodeau, Isabelle Bachand, Jessie Y J. Bao, Amy Hin Yan Tong, Chi-Ho Lin, Bruno Millet, Nematollah Jaafari, Ridha Joobar, Patrick A. Dion, Si Lok, Marie-Odile Krebs, and Guy A. Rouleau. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet*, 43(9):860–863, Sep 2011.
- [75] Anita Rauch*, Dagmar Wieczorek*, Elisabeth Graf*, Thomas Wieland*, Sabine Ende, Thomas Schwarzmayr, Beate Albrecht, Deborah Bartholdi, Jasmin Beygo, Nataliya Di Donato, Andreas Dufke, Kirsten Cremer, Maja Hempel, Denise Horn, Juliane Hoyer, Pascal Joset, Albrecht Röpke, Ute Moog, Angelika Riess, Christian T. Thiel, Andreas Tzschach, Antje Wiesener, Eva Wohlleber, Christiane Zweier, Arif B. Ekici, Alexander M. Zink, Andreas Rump, Christa Meisinger, Harald Grallert, Heinrich Sticht, Annette Schenck, Hartmut Engels, Gudrun Rappold, Evelin Schröck, Peter Wieacker, Olaf Riess, Thomas Meitinger, André Reis, and Tim M. Strom. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*, 380(9854):1674–1682, Nov 2012.
- [76] Felix Beuschlein, Sheerazed Boulkroun, Andrea Osswald, Thomas Wieland, Hang N. Nielsen, Urs D. Lichtenauer, David Penton, Vivien R. Schack, Laurence Amar, Evelyn Fischer, Anett Walther, Philipp Tauber, Thomas Schwarzmayr, Susanne Diener, Elisabeth Graf, Bruno Allolio, Benoit Samson-Couterie, Arndt Benecke, Marcus Quinkler, Francesco Fallo, Pierre-Francois Plouin, Franco Mantero, Thomas Meitinger, Paolo Mularo, Xavier Jeunemaitre, Richard Warth, Bente Vilsen, Maria-Christina Zennaro, Tim M. Strom, and Martin Reincke. Somatic mutations in *atp1a1*

-
- and *atp2b3* lead to aldosterone-producing adenomas and secondary hypertension. *Nat Genet*, 45(4):440–4, 444e1–2, Apr 2013.
- [77] Martin Reincke, Silviu Sbiera, Akira Hayakawa, Marily Theodoropoulou, Andrea Osswald, Felix Beuschlein, Thomas Meitinger, Emi Mizuno-Yamasaki, Kohei Kawaguchi, Yasushi Saeki, Keiji Tanaka, Thomas Wieland, Elisabeth Graf, Wolfgang Saeger, Cristina L. Ronchi, Bruno Allolio, Michael Buchfelder, Tim M. Strom, Martin Fassnacht, and Masayuki Komada. Mutations in the deubiquitinase gene *usp8* cause cushing’s disease. *Nat Genet*, 47(1):31–38, Jan 2015.
- [78] Sarah B. Ng, Emily H. Turner, Peggy D. Robertson, Steven D. Flygare, Abigail W. Bigham, Choli Lee, Tristan Shaffer, Michelle Wong, Arindam Bhattacharjee, Evan E. Eichler, Michael Bamshad, Deborah A. Nickerson, and Jay Shendure. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261):272–276, Sep 2009.
- [79] Alexander Hoischen, Bregje W M. van Bon, Benjamín Rodríguez-Santiago, Christian Gilissen, Lisenka E L M. Vissers, Petra de Vries, Irene Janssen, Bart van Lier, Rob Hastings, Sarah F. Smithson, Ruth Newbury-Ecob, Susanne Kjaergaard, Judith Goodship, Ruth McGowan, Deborah Bartholdi, Anita Rauch, Maarit Peippo, Jan M. Cobben, Dagmar Wieczorek, Gabriele Gillessen-Kaesbach, Joris A. Veltman, Han G. Brunner, and Bert B B A. de Vries. De novo nonsense mutations in *asxl1* cause bohring-opitz syndrome. *Nat Genet*, 43(8):729–731, Aug 2011.
- [80] Dagmar Wieczorek, William G. Newman, Thomas Wieland, Tea Berulava, Maria Kaffe, Daniela Falkenstein, Christian Beetz, Elisabeth Graf, Thomas Schwarzmayr, Sofia Douzgou, Jill Clayton-Smith, Sarah B. Daly, Simon G. Williams, Sanjeev S. Bhaskar, Jill E. Urquhart, Beverley Anderson, James O’Sullivan, Odile Boute, Jasmin Gundlach, Johanna Christina Czeschik, Anthonie J. van Essen, Filiz Hazan, Sarah Park, Anne Hing, Alma Kuechler, Dietmar R. Lohmann, Kerstin U. Ludwig, Elisabeth Mangold, Laura Steenpaß, Michael Zeschnigk, Johannes R. Lemke, Charles Marques Lourenco, Ute Hehr, Eva-Christina Prott, Melanie Waldenberger, Anne C. Böhmer, Bernhard Horsthemke, Raymond T. O’Keefe, Thomas Meitinger,

-
- John Burn, Hermann-Josef Lüdecke, and Tim M. Strom. Compound heterozygosity of low-frequency promoter deletions and rare loss-of-function mutations in *txnl4a* causes burn-mckeown syndrome. *Am J Hum Genet*, 95(6):698–707, Dec 2014.
- [81] Lia Crotti, Christopher N. Johnson, Elisabeth Graf, Gaetano M. De Ferrari, Bettina F. Cuneo, Marc Ovadia, John Papagiannis, Michael D. Feldkamp, Subodh G. Rathi, Jennifer D. Kunic, Matteo Pedrazzini, Thomas Wieland, Peter Lichtner, Britt-Maria Beckmann, Travis Clark, Christian Shaffer, D Woodrow Benson, Stefan Kääh, Thomas Meitinger, Tim M. Strom, Walter J. Chazin, Peter J. Schwartz, and Alfred L George, Jr. Calmodulin mutations associated with recurrent cardiac arrest in infants. *Circulation*, 127(9):1009–1017, Mar 2013.
- [82] Jan Murken. *Humangenetik - Inzuchts- und Verwandtschaftskoeffizient bei verschiedenen Verwandtschaftsverhältnissen*. Georg Thieme Verlag, 2006.
- [83] Elisabeth Stogmann, Eva Reinthaler, Salwa Eltawil, Mohammed A. El Etribi, Mahmoud Hemed, Nevine El Nahhas, Ahmed M. Gaber, Amal Fouad, Sherif Edris, Anna Benet-Pages, Sebastian H. Eck, Ekaterina Patarai, Davide Mei, Alexis Brice, Suzanne Lesage, Renzo Guerrini, Friedrich Zimprich, Tim M. Strom, and Alexander Zimprich. Autosomal recessive cortical myoclonic tremor and epilepsy: association with a mutation in the potassium channel associated gene *cntn2*. *Brain*, 136(Pt 4):1155–1160, Apr 2013.
- [84] Sarah B. Pierce, Tom Walsh, Karen M. Chisholm, Ming K. Lee, Anne M. Thornton, Agata Fiumara, John M. Opitz, Ephrat Levy-Lahad, Rachel E. Klevit, and Mary-Claire King. Mutations in the *dbp*-deficiency protein *hsd17b4* cause ovarian dysgenesis, hearing loss, and ataxia of perrault syndrome. *Am J Hum Genet*, 87(2):282–288, Aug 2010.
- [85] Ngak-Leng Sim, Prateek Kumar, Jing Hu, Steven Henikoff, Georg Schneider, and Pauline C. Ng. Sift web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*, 40(Web Server issue):W452–W457, Jul 2012.

-
- [86] Jana Marie Schwarz, Christian Rödelsperger, Markus Schuelke, and Dominik Seelow. Mutationtaster evaluates disease-causing potential of sequence alterations. *Nat Methods*, 7(8):575–576, Aug 2010.
- [87] Daniel M. Jordan, Vasily E. Ramensky, and Shamil R. Sunyaev. Human allelic variation: perspective from protein function, structure, and evolution. *Curr Opin Struct Biol*, 20(3):342–350, Jun 2010.
- [88] Yaniv Erlich, Simon Edvardson, Emily Hodges, Shamir Zenvirt, Pramod Thekkat, Avraham Shaag, Talya Dor, Gregory J. Hannon, and Orly Elpeleg. Exome sequencing and disease-network analysis of a single family implicate a mutation in kif1a in hereditary spastic paraparesis. *Genome Res*, 21(5):658–664, May 2011.
- [89] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.
- [90] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. Primer-directed enzymatic amplification of dna with a thermostable dna polymerase. *Science*, 239(4839):487–491, Jan 1988.
- [91] Rolf H A M. Vossen, Emmelien Aten, Anja Roos, and Johan T. den Dunnen. High-resolution melting analysis (hrma): more than just sequence variant screening. *Hum Mutat*, 30(6):860–866, Jun 2009.
- [92] Rui Chen, Hogune Im, and Michael Snyder. Whole-exome enrichment with the illumina truseq exome enrichment platform. *Cold Spring Harbor Protocols*, 2015(7):pdb-prot084863, 2015.
- [93] Agilent. Sureselect exome.
- [94] Dagmar Wiczorek, William G. Newman, Thomas Wieland, Tea Berulava, Maria Kaffe, Daniela Falkenstein, Christian Beetz, Elisabeth Graf, Thomas Schwarzmayr, Sofia Douzgou, Jill Clayton-Smith, Sarah B. Daly, Simon G.

-
- Williams, Sanjeev S. Bhaskar, Jill E. Urquhart, Beverley Anderson, James O’Sullivan, Odile Boute, Jasmin Gundlach, Johanna Christina Czeschik, Anthonie J. van Essen, Filiz Hazan, Sarah Park, Anne Hing, Alma Kuechler, Dietmar R. Lohmann, Kerstin U. Ludwig, Elisabeth Mangold, Laura Steenpaß, Michael Zeschnigk, Johannes R. Lemke, Charles Marques Lourenco, Ute Hehr, Eva-Christina Prott, Melanie Waldenberger, Anne C. Böhmer, Bernhard Horsthemke, Raymond T. O’Keefe, Thomas Meitinger, John Burn, Hermann-Josef Lüdecke, and Tim M. Strom. Compound heterozygosity of low-frequency promoter deletions and rare loss-of-function mutations in *txnl4a* causes burn-mckeown syndrome. *Am J Hum Genet*, 95(6):698–707, Dec 2014.
- [95] Iraad F Bronner, Michael A Quail, Daniel J Turner, and Harold Swerdlow. Improved protocols for illumina sequencing. *Current Protocols in Human Genetics*, pages 18–2, 2014.
- [96] Michael Andrew Quail. Dna: Mechanical breakage. *e LS*, 2001.
- [97] Covaris. <https://covaris.com/products/afa-ultrasonication/>.
- [98] Illumina Inc. <https://emea.illumina.com/science/technology/next-generation-sequencing/illumina-sequencing-history/decade-in-sequencing.html>.
- [99] Martin Kircher, Udo Stenzel, and Janet Kelso. Improved base calling for the illumina genome analyzer using machine learning strategies. *Genome biology*, 10(8):R83, 2009.
- [100] Fan Hsu, W James Kent, Hiram Clawson, Robert M Kuhn, Mark Diekhans, and David Haussler. The ucsc known genes. *Bioinformatics*, 22(9):1036–1046, 2006.
- [101] Kim D Pruitt and Donna R Maglott. Refseq and locuslink: Ncbi gene-centered resources. *Nucleic acids research*, 29(1):137–140, 2001.
- [102] Stephen T Sherry, M-H Ward, M Kholodov, J Baker, Lon Phan, Elizabeth M Smigielski, and Karl Sirotkin. dbsnp: the ncbi database of genetic variation. *Nucleic acids research*, 29(1):308–311, 2001.

-
- [103] Melissa J Landrum, Jennifer M Lee, Mark Benson, Garth Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, Jennifer Hart, Douglas Hoffman, Jeffrey Hoover, et al. Clinvar: public archive of interpretations of clinically relevant variants. *Nucleic acids research*, 44(D1):D862–D868, 2015.
- [104] James D Watson. The human genome project: past, present, and future. *Science*, 248(4951):44–49, 1990.
- [105] KJ Karczewski and L Francioli. The genome aggregation database (gnomad). *MacArthur Lab*, 2017.
- [106] James T Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. Integrative genomics viewer. *Nature biotechnology*, 29(1):24, 2011.
- [107] Genomics. Genomics clarity lims. <https://www.genomics.com/clarity-lims/>.
- [108] Timothy G Bath, Selcuk Bozdog, Vackar Afzal, and Daniel Crowther. Limsportal and bonsailims: development of a lab information management system for translational medicine. *Source code for biology and medicine*, 6(1):9, 2011.
- [109] Susan M Grimes and Hanlee P Ji. Mendelims: a web-based laboratory information management system for clinical genome sequencing. *BMC bioinformatics*, 15(1):290, 2014.
- [110] Danielle Dionne; Ryan Hegarty; Justin Abreu; Matthew Coole; Laurie Holmes; Matthew Lee; Sheli McDonough; Alvin Tam; Tim DeSmet; Bertrand Haas; Yossi Farjoun; Danielle Perrin; Niall Lennon; Sheila Fisher; Stacey Gabriel. Broad institute. enabling population-scale pcr-free whole genome sequencing. In *AGBT-Conference Poster Presentation*, 2016.
- [111] Maura Costello, Mark Fleharty, Justin Abreu, Yossi Farjoun, Steven Ferreira, Laurie Holmes, Brian Granger, Lisa Green, Tom Howd, Tamara Mason, et al. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC genomics*, 19(1):332, 2018.

-
- [112] Sheila Fisher, Andrew Barry, Justin Abreu, Brian Minie, Jillian Nolan, Toni M Delorey, Geneva Young, Timothy J Fennell, Alexander Allen, Lauren Ambrogio, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome biology*, 12(1):1, 2011.
- [113] Konstantinos Nikopoulos, Christian Gilissen, Alexander Hoischen, C Erik van Nouhuys, F Nienke Boonstra, Ellen AW Blokland, Peer Arts, Nienke Wieskamp, Tim M Strom, Carmen Ayuso, et al. Next-generation sequencing of a 40 mb linkage interval reveals *tspan12* mutations in patients with familial exudative vitreoretinopathy. *The American Journal of Human Genetics*, 86(2):240–247, 2010.
- [114] Stephen P Daiger, Lori S Sullivan, Sara J Bowne, David G Birch, John R Heckenlively, Eric A Pierce, and George M Weinstock. Targeted high-throughput dna sequencing for gene discovery in retinitis pigmentosa. In *Retinal Degenerative Diseases*, pages 325–331. Springer, 2010.
- [115] Daniel Summerer, Nadine Schracke, Haiguo Wu, Yang Cheng, Stephan Bau, Cord F Stähler, Peer F Stähler, and Markus Beier. Targeted high throughput sequencing of a cancer-related exome subset by specific sequence capture with a fully automated microarray platform. *Genomics*, 95(4):241–246, 2010.
- [116] CA 95051 USA Agilent Technologies, Inc. 5301 Stevens Creek Rd Santa Clara. *SureSelect Target Enrichment System for Illumina Paired-End Sequencing Library Protocol Version 2.0.1, May 2010*, 2010.
- [117] Margaret M DeAngelis, David G Wang, and Trevor L Hawkins. Solid-phase reversible immobilization for the isolation of pcr products. *Nucleic acids research*, 23(22):4742, 1995.
- [118] Agencourt Bioscience. *AGENCOURT AMPURE XP PCR PURIFICATION*. Beckman Coulter Inc.
- [119] Charles Robertson. Spectrophotometry’s next-generation technology. *American biotechnology laboratory*, 21(13):23–24, 2003.

-
- [120] M O’neill, J McPartlin, K Arthure, S Riedel, and ND McMillan. Comparison of the tlda with the nanodrop and the reference qubit system. In *Journal of Physics: Conference series*, volume 307, page 012047. IOP Publishing, 2011.
- [121] Anu Sironen, Pekka Uimari, Johanna Vilkki, et al. Comparison of different dna extraction methods from hair root follicles to genotype finnish landrace boars with the illumina porcinesnp60 beadchip. 2011.
- [122] Michele Simbolo, Marisa Gottardi, Vincenzo Corbo, Matteo Fassan, Andrea Mafficini, Giorgio Malpeli, Rita T Lawlor, and Aldo Scarpa. Dna qualification workflow for next generation sequencing of histopathological samples. *PLoS One*, 8(6):e62692, 2013.
- [123] A. I. Dragan, J. R. Casas-Finet, E. S. Bishop, R. J. Strouse, M. A. Schermerman, and C. D. Geddes. Characterization of picogreen interaction with dsdna and the origin of its fluorescence enhancement upon binding. *Biophys J*, 99(9):3010–3019, Nov 2010.
- [124] V. L. Singer, L. J. Jones, S. T. Yue, and R. P. Haugland. Characterization of picogreen reagent and development of a fluorescence-based solution assay for double-stranded dna quantitation. *Anal Biochem*, 249(2):228–238, Jul 1997.
- [125] <http://www.u.arizona.edu/~gwatts/azcc/InterpretingSpec.pdf>. Interpreting nanodrop (spectrophotometric) results.
- [126] SA Miller, DD Dykes, and HFRN Polesky. A simple salting out procedure for extracting dna from human nucleated cells. *Nucleic acids research*, 16(3):1215, 1988.
- [127] <https://www.qiagen.com/us/resources/technologies/ngs/sample-technologies/>.
- [128] SATOSHI Sasaki and HITOYATA Shimokawa. The amelogenin gene. *International Journal of Developmental Biology*, 39(1):127–133, 2003.

-
- [129] Hongdo Do and Alexander Dobrovic. Sequence artifacts in dna from formalin-fixed tissues: causes and strategies for minimization. *Clinical chemistry*, 61(1):64–71, 2015.
- [130] Sebastian Stintzing, Volker Heinemann, Andreas Jung, Nicolas Moosmann, Wolfgang Hiddemann, and Thomas Kirchner. The treatment of colorectal carcinoma with monoclonal antibodies: the importance of kras mutation analysis and egfr status. *Deutsches Ärzteblatt International*, 106(12):202, 2009.
- [131] Pauline Robbe, Niko Popitsch, Samantha JL Knight, Pavlos Antoniou, Jennifer Becq, Miao He, Alexander Kanapin, Anastasia Samsonova, Dimitrios V Vavoulis, Mark T Ross, et al. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 genomes project. *Genetics in Medicine*, 2018.
- [132] Lei Wen, Fu-Lei Tang, Yan Hong, Shi-Wen Luo, Chun-Lei Wang, Wanxia He, Chengyong Shen, Ji-Ung Jung, Fei Xiong, Dae-hoon Lee, et al. Vps35 haploinsufficiency increases alzheimer disease neuropathology. *J Cell Biol*, pages jcb–201105109, 2011.
- [133] A. E. Lang and A. M. Lozano. Parkinson’s disease. first of two parts. *N Engl J Med*, 339(15):1044–1053, Oct 1998.
- [134] E. R. Dorsey, R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway, K. Kieburtz, F. J. Marshall, B. M. Ravina, G. Schifitto, A. Siderowf, and C. M. Tanner. Projected number of people with parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5):384–386, Jan 2007.
- [135] M. H. Polymeropoulos, C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe, and R. L. Nussbaum. Mutation in the alpha-synuclein gene identified in families with parkinson’s disease. *Science*, 276(5321):2045–2047, Jun 1997.

-
- [136] T. Kitada, S. Asakawa, N. Hattori, H. Matsumine, Y. Yamamura, S. Minoshima, M. Yokochi, Y. Mizuno, and N. Shimizu. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*, 392(6676):605–608, Apr 1998.
- [137] T. Gasser, B. Müller-Myhsok, Z. K. Wszolek, R. Oehlmann, D. B. Calne, V. Bonifati, B. Bereznoi, E. Fabrizio, P. Vieregge, and R. D. Horstmann. A susceptibility locus for parkinson’s disease maps to chromosome 2p13. *Nat Genet*, 18(3):262–265, Mar 1998.
- [138] E. Leroy, R. Boyer, G. Auburger, B. Leube, G. Ulm, E. Mezey, G. Harta, M. J. Brownstein, S. Jonnalagada, T. Chernova, A. Dehejia, C. Lavedan, T. Gasser, P. J. Steinbach, K. D. Wilkinson, and M. H. Polymeropoulos. The ubiquitin pathway in parkinson’s disease. *Nature*, 395(6701):451–452, Oct 1998.
- [139] Enza Maria Valente, Sergio Salvi, Tamara Ialongo, Roberta Marongiu, Antonio Emanuele Elia, Viviana Caputo, Luigi Romito, Alberto Albanese, Bruno Dallapiccola, and Anna Rita Bentivoglio. Pink1 mutations are associated with sporadic early-onset parkinsonism. *Ann Neurol*, 56(3):336–341, Sep 2004.
- [140] Vincenzo Bonifati, Patrizia Rizzu, Marijke J. van Baren, Onno Schaap, Guido J. Breedveld, Elmar Krieger, Marieke C J. Dekker, Ferdinando Squitieri, Pablo Ibanez, Marijke Joosse, Jeroen W. van Dongen, Nicola Vanacore, John C. van Swieten, Alexis Brice, Giuseppe Meco, Cornelia M. van Duijn, Ben A. Oostra, and Peter Heutink. Mutations in the dj-1 gene associated with autosomal recessive early-onset parkinsonism. *Science*, 299(5604):256–259, Jan 2003.
- [141] Coro Paisán-Ruíz, Shushant Jain, E Whitney Evans, William P. Gilks, Javier Simón, Marcel van der Brug, Adolfo López de Munain, Silvia Aparicio, Angel Martínez Gil, Naheed Khan, Janel Johnson, Javier Ruiz Martinez, David Nicholl, Itxaso Marti Carrera, Amets Saénz Pena, Rohan de Silva, Andrew Lees, José Félix Martí-Massó, Jordi Pérez-Tur, Nick W. Wood, and Andrew B. Singleton. Cloning of the gene containing mutations that cause park8-linked parkinson’s disease. *Neuron*, 44(4):595–600, Nov 2004.

-
- [142] Alexander Zimprich, Saskia Biskup, Petra Leitner, Peter Lichtner, Matthew Farrer, Sarah Lincoln, Jennifer Kachergus, Mary Hulihan, Ryan J. Uitti, Donald B. Calne, A Jon Stoessl, Ronald F. Pfeiffer, Nadja Patenge, Iria Carballo Carbajal, Peter Vieregge, Friedrich Asmus, Bertram Müller-Myhsok, Dennis W. Dickson, Thomas Meitinger, Tim M. Strom, Zbigniew K. Wszolek, and Thomas Gasser. Mutations in *lrrk2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*, 44(4):601–607, Nov 2004.
- [143] Alfredo Ramirez, André Heimbach, Jan Gründemann, Barbara Stiller, Dan Hampshire, L Pablo Cid, Ingrid Goebel, Ammar F. Mubaidin, Abdul-Latif Wriekat, Jochen Roeper, Amir Al-Din, Axel M. Hillmer, Meliha Karsak, Birgit Liss, C Geoffrey Woods, Maria I. Behrens, and Christian Kubisch. Hereditary parkinsonism with dementia is caused by mutations in *atp13a2*, encoding a lysosomal type 5 p-type atpase. *Nat Genet*, 38(10):1184–1191, Oct 2006.
- [144] Andrew A. Hicks, Hjörvar Pétursson, Thorlákur Jónsson, Hreinn Stefánsson, Hrefna S. Jóhannsdóttir, Jesus Sainz, Michael L. Frigge, Augustine Kong, Jeffrey R. Gulcher, Kári Stefánsson, and Sigurlaug Sveinbjörnsdóttir. A susceptibility gene for late-onset idiopathic parkinson’s disease. *Ann Neurol*, 52(5):549–555, Nov 2002.
- [145] Corinne Lautier, Stefano Goldwurm, Alexandra Dürr, Barbara Giovannone, William G. Tsiras, Gianni Pezzoli, Alexis Brice, and Robert J. Smith. Mutations in the *gigyl2* (*tnrc15*) gene at the *park11* locus in familial parkinson disease. *Am J Hum Genet*, 82(4):822–833, Apr 2008.
- [146] Nathan Pankratz, William C. Nichols, Sean K. Uniacke, Cheryl Halter, Jill Murrell, Alice Rudolph, Clifford W. Shults, P Michael Conneally, Tatiana Foroud, and Parkinson Study Group . Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex parkinson disease families. *Hum Mol Genet*, 12(20):2599–2608, Oct 2003.
- [147] Karsten M. Strauss, L Miguel Martins, Helene Plun-Favreau, Frank P. Marx, Sabine Kautzmann, Daniela Berg, Thomas Gasser, Zbigniew Wszolek, Thomas Müller, Antje Bornemann, Hartwig Wolburg, Julian Down-

-
- ward, Olaf Riess, Jörg B. Schulz, and Rejko Krüger. Loss of function mutations in the gene encoding *omi/htra2* in parkinson's disease. *Hum Mol Genet*, 14(15):2099–2111, Aug 2005.
- [148] Coro Paisan-Ruiz, Kailash P. Bhatia, Abi Li, Dena Hernandez, Mary Davis, Nick W. Wood, John Hardy, Henry Houlden, Andrew Singleton, and Susanne A. Schneider. Characterization of *pla2g6* as a locus for dystonia-parkinsonism. *Ann Neurol*, 65(1):19–23, Jan 2009.
- [149] Seyedmehdi Shojaee, Farzad Sina, Setareh Sadat Banihosseini, Mohammad Hossein Kazemi, Reza Kalhor, Gholam-Ali Shahidi, Hossein Fakhrai-Rad, Mostafa Ronaghi, and Elahe Elahi. Genome-wide linkage analysis of a parkinsonian-pyramidal syndrome pedigree by 500 k snp arrays. *Am J Hum Genet*, 82(6):1375–1384, Jun 2008.
- [150] Wataru Satake, Yuko Nakabayashi, Ikuko Mizuta, Yushi Hirota, Chiyomi Ito, Michiaki Kubo, Takahisa Kawaguchi, Tatsuhiko Tsunoda, Masahiko Watanabe, Atsushi Takeda, Hiroyuki Tomiyama, Kenji Nakashima, Kazuko Hasegawa, Fumiya Obata, Takeo Yoshikawa, Hideshi Kawakami, Saburo Sakoda, Mitsutoshi Yamamoto, Nobutaka Hattori, Miho Murata, Yusuke Nakamura, and Tatsushi Toda. Genome-wide association study identifies common variants at four loci as genetic risk factors for parkinson's disease. *Nat Genet*, 41(12):1303–1307, Dec 2009.
- [151] Carles Vilariño-Güell, Christian Wider, Owen A. Ross, Justus C. Daxsel, Jennifer M. Kachergus, Sarah J. Lincoln, Alexandra I. Soto-Ortolaza, Stephanie A. Cobb, Gregory J. Wilhoite, Justin A. Bacon, Bahareh Behrouz, Heather L. Melrose, Emna Hentati, Andreas Puschmann, Daniel M. Evans, Elizabeth Conibear, Wyeth W. Wasserman, Jan O. Aasly, Pierre R. Burkhard, Ruth Djaldetti, Joseph Ghika, Faycal Hentati, Anna Krygowska-Wajs, Tim Lynch, Eldad Melamed, Alex Rajput, Ali H. Rajput, Alessandra Solida, Ruey-Meei Wu, Ryan J. Uitti, Zbigniew K. Wszolek, François Vingerhoets, and Matthew J. Farrer. *Vps35* mutations in parkinson disease. *Am J Hum Genet*, 89(1):162–167, Jul 2011.
- [152] Marie-Christine Chartier-Harlin, Justus C. Daxsel, Carles Vilariño-Güell, Sarah J. Lincoln, Frédéric Leprêtre, Mary M. Hulihan, Jennifer Kachergus,

Austen J. Milnerwood, Lucia Tapia, Mee-Sook Song, Emilie Le Rhun, Eugénie Mutez, Lydie Larvor, Aurélie Dufflot, Christel Vanbesien-Mailliot, Alexandre Kreisler, Owen A. Ross, Kenya Nishioka, Alexandra I. Soto-Ortolaza, Stephanie A. Cobb, Heather L. Melrose, Bahareh Behrouz, Brett H. Keeling, Justin A. Bacon, Emna Hentati, Lindsey Williams, Akiko Yanagiya, Nahum Sonenberg, Paul J. Lockhart, Abba C. Zubair, Ryan J. Uitti, Jan O. Aasly, Anna Krygowska-Wajs, Grzegorz Opala, Zbigniew K. Wszolek, Roberta Frigerio, Demetrius M. Maraganore, David Gosal, Tim Lynch, Michael Hutchinson, Anna Rita Bentivoglio, Enza Maria Valente, William C. Nichols, Nathan Pankratz, Tatiana Foroud, Rachel A. Gibson, Faycal Hentati, Dennis W. Dickson, Alain Destée, and Matthew J. Farrer. Translation initiator eif4g1 mutations in familial parkinson disease. *Am J Hum Genet*, 89(3):398–406, Sep 2011.

- [153] Simon Edvardson, Yuval Cinnamon, Asaf Ta-Shma, Avraham Shaag, Yang-In Yim, Shamir Zenvirt, Chaim Jalas, Suzanne Lesage, Alexis Brice, Albert Taraboulos, Klaus H. Kaestner, Lois E. Greene, and Orly Elpeleg. A deleterious mutation in dnajc6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. *PLoS One*, 7(5):e36458, 2012.
- [154] Catharine E. Krebs, Siamak Karkheiran, James C. Powell, Mian Cao, Vladimir Makarov, Hossein Darvish, Gilbert Di Paolo, Ruth H. Walker, Gholam Ali Shahidi, Joseph D. Buxbaum, Pietro De Camilli, Zhenyu Yue, and Coro Paisán-Ruiz. The sac1 domain of synj1 identified mutated in a family with early-onset progressive parkinsonism with generalized seizures. *Hum Mutat*, 34(9):1200–1207, Sep 2013.
- [155] Marialuisa Quadri, Mingyan Fang, Marina Picillo, Simone Olgiati, Guido J. Breedveld, Josja Graafland, Bin Wu, Fengping Xu, Roberto Erro, Marianna Amboni, Sabina Pappatà, Mario Quarantelli, Grazia Annesi, Aldo Quattrone, Hsin F. Chien, Egberto R. Barbosa, International Parkinsonism Genetics Network, Ben A. Oostra, Paolo Barone, Jun Wang, and Vincenzo Bonifati. Mutation in the synj1 gene associated with autosomal recessive, early-onset parkinsonism. *Hum Mutat*, 34(9):1208–1215, Sep 2013.

-
- [156] Carles Vilariño-Güell, Alex Rajput, Austen J Milnerwood, Brinda Shah, Chelsea Szu-Tu, Joanne Trinh, Irene Yu, Mary Encarnacion, Lise N Munsie, Lucia Tapia, et al. Dnajc13 mutations in parkinson disease. *Human molecular genetics*, 23(7):1794–1801, 2013.
- [157] Han-Xiang Deng, Yong Shi, Yi Yang, Kreshnik B Ahmeti, Nimrod Miller, Cao Huang, Lijun Cheng, Hong Zhai, Sheng Deng, Karen Nuytemans, et al. Identification of tmem230 mutations in familial parkinson’s disease. *Nature genetics*, 48(7):733, 2016.
- [158] Manabu Funayama, Kenji Ohe, Taku Amo, Norihiko Furuya, Junji Yamaguchi, Shinji Saiki, Yuanzhe Li, Kotaro Ogaki, Maya Ando, Hiroyo Yoshino, et al. Chchd2 mutations in autosomal dominant late-onset parkinson’s disease: a genome-wide linkage and sequencing study. *The Lancet Neurology*, 14(3):274–282, 2015.
- [159] Suzanne Lesage, Valérie Drouet, Elisa Majounie, Vincent Deramecourt, Maxime Jacoupy, Aude Nicolas, Florence Cormier-Dequaire, Sidi Mohamed Hassoun, Claire Pujol, Sorana Ciura, et al. Loss of vps13c function in autosomal-recessive parkinsonism causes mitochondrial dysfunction and increases pink1/parkin-dependent mitophagy. *The American Journal of Human Genetics*, 98(3):500–513, 2016.
- [160] K Kalinderi, L Fidani, Z Katsarou, and S Bostantjopoulou. Pharmacological treatment and the prospect of pharmacogenetics in parkinsons disease. *International journal of clinical practice*, 65(12):1289–1294, 2011.
- [161] Aitor Hierro, Adriana L. Rojas, Raul Rojas, Namita Murthy, Grégory Efantin, Andrey V. Kajava, Alasdair C. Steven, Juan S. Bonifacino, and James H. Hurley. Functional architecture of the retromer cargo-recognition complex. *Nature*, 449(7165):1063–1067, Oct 2007.
- [162] Georgii G. Krivov, Maxim V. Shapovalov, and Roland L Dunbrack, Jr. Improved prediction of protein side-chain conformations with scwrl4. *Proteins*, 77(4):778–795, Dec 2009.
- [163] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecu-

-
- lar simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.
- [164] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang, and Peter Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem*, 24(16):1999–2012, Dec 2003.
- [165] Michael W Mahoney and William L Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics*, 112(20):8910–8922, 2000.
- [166] Sunbin Liu, Reinhard Rauhut, Hans-Peter Vornlocher, and Reinhard Lührmann. The network of protein-protein interactions within the human u4/u6.u5 tri-snrnp. *RNA*, 12(7):1418–1430, Jul 2006.
- [167] H. V. Toriello and J. V. Higgins. A boy with choanal atresia and cardiac defect: Burn-mckeown syndrome? *Clin Dysmorphol*, 8(2):143–145, Apr 1999.
- [168] Dagmar Wiczorek, Ozge Altug Teber, Dietmar Lohmann, and Gabriele Gillessen-Kaesbach. Two brothers with burn-mckeown syndrome. *Clin Dysmorphol*, 12(3):171–174, Jul 2003.
- [169] J. Burn, C. McKeown, J. Wagget, R. Bray, and J. Goodship. New dysmorphic syndrome with choanal atresia in siblings. *Clin Dysmorphol*, 1(3):137–144, Jul 1992.
- [170] Jacqueline AC Goos, Sigrid MA Swagemakers, Stephen RF Twigg, Marieke F van Dooren, A Jeannette M Hoogeboom, Christian Beetz, Sven Günther, Frank J Magielsen, Charlotte W Ockeloen, Maria A Ramos-Arroyo, et al. Identification of causative variants in *txnl4a* in burn-mckeown syndrome and isolated choanal atresia. *European Journal of Human Genetics*, 25(10):1126, 2017.
- [171] Melissa S. Jurica and Melissa J. Moore. Pre-mrna splicing: awash in a sea of proteins. *Mol Cell*, 12(1):5–14, Jul 2003.

-
- [172] Bastian Linder, Anja Hirmer, Andreas Gal, Klaus Rütter, Hanno Jörn Bolz, Christoph Winkler, Bernhard Laggerbauer, and Utz Fischer. Identification of a prpf4 loss-of-function variant that abrogates u4/u6.u5 tri-snrnp integration and is associated with retinitis pigmentosa. *PLoS One*, 9(11):e111754, 2014.
- [173] Nuno André Faustino and Thomas A. Cooper. Pre-mrna splicing and human disease. *Genes Dev*, 17(4):419–437, Feb 2003.
- [174] Thomas A. Cooper, Lili Wan, and Gideon Dreyfuss. Rna and disease. *Cell*, 136(4):777–793, Feb 2009.
- [175] Patrick Edery, Charles Marcaillou, Mourad Sahbatou, Audrey Labalme, Joëlle Chastang, Renaud Touraine, Emmanuel Tubacher, Faiza Senni, Michael B. Bober, Sheela Nampoothiri, Pierre-Simon Jouk, Elisabeth Steichen, Siren Berland, Annick Toutain, Carol A. Wise, Damien Sanlaville, Francis Rousseau, Françoise Clerget-Darpoux, and Anne-Louise Leutenegger. Association of tals developmental disorder with defect in minor splicing component u4atac snrna. *Science*, 332(6026):240–243, Apr 2011.
- [176] Huiling He, Sandya Liyanarachchi, Keiko Akagi, Rebecca Nagy, Jingfeng Li, Rosemary C. Dietrich, Wei Li, Nikhil Sebastian, Bernard Wen, Baozhong Xin, Jarnail Singh, Pearly Yan, Hansjuerg Alder, Eric Haan, Dagmar Wiczorek, Beate Albrecht, Erik Puffenberger, Heng Wang, Judith A. Westman, Richard A. Padgett, David E. Symer, and Albert de la Chapelle. Mutations in u4atac snrna, a component of the minor spliceosome, in the developmental disorder mopd i. *Science*, 332(6026):238–240, Apr 2011.
- [177] Yichang Jia, John C. Mu, and Susan L. Ackerman. Mutation of a u2 snrna gene causes global disruption of alternative splicing and neurodegeneration. *Cell*, 148(1-2):296–308, Jan 2012.
- [178] Sandra M. Pasternack, Melanie Refke, Elham Paknia, Hans Christian Hennies, Thomas Franz, Niklas Schäfer, Alan Fryer, Maurice van Steensel, Elizabeth Sweeney, Miquel Just, Clemens Grimm, Roland Kruse, Carlos Ferrándiz, Markus M. Nöthen, Utz Fischer, and Regina C. Betz. Mutations in snrpe, which encodes a core protein of the spliceosome, cause autosomal-dominant hypotrichosis simplex. *Am J Hum Genet*, 92(1):81–87, Jan 2013.

-
- [179] Anne V. Hing, Christy Leblond, Raymond W. Sze, Jacqueline R. Starr, Stephanie Monks, and Melissa A. Parisi. A novel oculo-oto-facial dysplasia in a native alaskan community with autosomal recessive inheritance. *Am J Med Genet A*, 140(8):804–812, Apr 2006.
- [180] John M. Opitz and John Burn. Re: Correspondence from wieczorek & gillessen-kaesbach and hing & parisi. *Am J Med Genet A*, 140(21):2385, Nov 2006.
- [181] K. Reuter, S. Nottrott, P. Fabrizio, R. Lührmann, and R. Ficner. Identification, characterization and crystal structure analysis of the human spliceosomal u5 snrnp-specific 15 kd protein. *J Mol Biol*, 294(2):515–525, Nov 1999.
- [182] Illumina. <https://www.illumina.com/informatics/research/sequencing-data-analysis-management/genologics-lims.html>.
- [183] Angela Bauch, Izabela Adamczyk, Piotr Buczek, Franz-Josef Elmer, Kaloyan Enimanev, Pawel Glyzewski, Manuel Kohler, Tomasz Pylak, Andreas Quandt, Chandrasekhar Ramakrishnan, et al. openbis: a flexible framework for managing and analyzing complex data in biology research. *BMC bioinformatics*, 12(1):468, 2011.
- [184] Thea Van Rossum, Ben Tripp, and Denise Daley. Slims a user friendly sample operations and inventory management system for genotyping labs. *Bioinformatics*, 26(14):1808–1810, 2010.
- [185] Andreas Dander, Stephan Pabinger, Michael Sperk, Maria Fischer, Gernot Stocker, and Zlatko Trajanoski. Seqbench: Integrated solution for the management and analysis of exome sequencing data. *BMC research notes*, 7(1):43, 2014.
- [186] David A Nix, Tonya L Di Sera, Brian K Dalley, Brett A Milash, Robert M Cundick, Kevin S Quinn, and Samir J Courdy. Next generation tools for genomic data generation, distribution, and visualization. *BMC bioinformatics*, 11(1):455, 2010.

-
- [187] Jitao Yang, Shuai Zhu, Guangpeng Xu, and Chao Han. Laboratory information management system for ngs genomics labs. In *HEALTHINF*, pages 326–333, 2017.
- [188] Francesco Venco, Yuriy Vaskin, Arnaud Ceol, and Heiko Muller. Smith: a lims for handling next-generation sequencing workflows. *BMC bioinformatics*, 15(14):S3, 2014.
- [189] Jonathan A Griffiths, Aaron TL Lun, Arianne C Richard, Karsten Bach, and John C Marioni. Detection and removal of barcode swapping in single-cell rna-seq data. *bioRxiv*, page 177048, 2017.
- [190] Daniel Vodák, Susanne Lorenz, Sigve Nakken, Lars Birger Aasheim, Harald Holte, Baoyan Bai, Ola Myklebost, Leonardo A Meza-Zepeda, and Eivind Hovig. Sample-index misassignment impacts tumour exome sequencing. *Scientific reports*, 8(1):5307, 2018.
- [191] Illumina. Effects of index misassignment on multiplexing and downstream analysis., 2018.
- [192] Stetsenko I Klimchuk O Plotnikov N Surkova E Ilinsky V Krasnenko A, Tsukanov K. Effect of dna insert length on whole-exome sequencing enrichment efficiency: an observational study. *Dovepress*, Volume 2018:8(12 June 2018):Pages 13–15, June 2018.
- [193] Meena Kishore Sakharkar, Vincent TK Chow, and Pandjassarame Kanguene. Distributions of exons and introns in the human genome. *In silico biology*, 4(4):387–393, 2004.
- [194] Ensel Oh, Yoon-La Choi, Mi Jeong Kwon, Ryong Nam Kim, Yu Jin Kim, Ji-Young Song, Kyung Soo Jung, and Young Kee Shin. Comparison of accuracy of whole-exome sequencing with formalin-fixed paraffin-embedded and fresh frozen tissue samples. *PloS one*, 10(12):e0144162, 2015.
- [195] Annalisa Astolfi, Milena Urbini, Valentina Indio, Margherita Nannini, Chiara Giusy Genovese, Donatella Santini, Maristella Saponara, Anna Mandrioli, Giorgio Ercolani, Giovanni Brandi, et al. Whole exome sequencing (wes) on formalin-fixed, paraffin-embedded (ffpe) tumor tissue in gastrointestinal stromal tumors (gist). *BMC genomics*, 16(1):892, 2015.

-
- [196] Martin Kerick, Melanie Isau, Bernd Timmermann, Holger Sültmann, Ralf Herwig, Sylvia Krobitch, Georg Schaefer, Irmgard Verdorfer, Georg Bartsch, Helmut Klocker, et al. Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (ffpe) tumor tissues, input amount and tumor heterogeneity. *BMC medical genomics*, 4(1):68, 2011.
- [197] Beth A Marosy, Brian D Craig, Kurt N Hetrick, P Dane Witmer, Hua Ling, Sean M Griffith, Benjamin Myers, Elaine A Ostrander, Janet L Stanford, Lawrence C Brody, et al. Generating exome enriched sequencing libraries from formalin-fixed, paraffin-embedded tissue dna for next-generation sequencing. *Current protocols in human genetics*, pages 18–10, 2017.
- [198] Juan S. Bonifacino and Raul Rojas. Retrograde transport from endosomes to the trans-golgi network. *Nat Rev Mol Cell Biol*, 7(8):568–579, Aug 2006.
- [199] Raul Rojas, Satoshi Kametaka, Carol R. Haft, and Juan S. Bonifacino. Interchangeable but essential functions of snx1 and snx2 in the association of retromer with endosomes and the trafficking of mannose 6-phosphate receptors. *Mol Cell Biol*, 27(3):1112–1124, Feb 2007.
- [200] Kianzoidlee. A schematic diagram showing the binding of the retromer complex with a cargo molecule from the endosome. the retromer complex comprises the vps trimer (vps35-29-26) and snx-bar dimers, 2011.
- [201] Ester Damen, Elmar Krieger, Jens E. Nielsen, Jelle Eygensteyn, and Jeroen E M. van Leeuwen. The human vps29 retromer component is a metallo-phosphoesterase for a cation-independent mannose 6-phosphate receptor substrate peptide. *Biochem J*, 398(3):399–409, Sep 2006.
- [202] Emélie Braschi, Vanessa Goyon, Rodolfo Zunino, Abhishek Mohanty, Liquan Xu, and Heidi M. McBride. Vps35 mediates vesicle transport between the mitochondria and peroxisomes. *Curr Biol*, 20(14):1310–1315, Jul 2010.
- [203] Viktor I. Korolchuk, Martin M. Schütz, Carolina Gómez-Llorente, João Rocha, Nico R. Lansu, Stephanie M. Collins, Yogesh P. Wairkar, Iain M. Robinson, and Cahir J. O’Kane. Drosophila vps35 function is necessary

-
- for normal endocytic trafficking and actin cytoskeleton organisation. *J Cell Sci*, 120(Pt 24):4367–4376, Dec 2007.
- [204] Matthew N J. Seaman. The retromer complex - endosomal protein recycling and beyond. *J Cell Sci*, 125(Pt 20):4693–4702, Oct 2012.
- [205] Thomas E. Willnow, Claus M. Petersen, and Anders Nykjaer. Vps10p-domain receptors - regulators of neuronal viability and function. *Nat Rev Neurosci*, 9(12):899–909, Dec 2008.
- [206] Ekaterina Rogaeva, Yan Meng, Joseph H. Lee, Yongjun Gu, Toshitaka Kawarai, Fanggeng Zou, Taiichi Katayama, Clinton T. Baldwin, Rong Cheng, Hiroshi Hasegawa, Fusheng Chen, Nobuto Shibata, Kathryn L. Lunetta, Raphaelle Pardossi-Piquard, Christopher Bohm, Yosuke Wakutani, L Adrienne Cupples, Karen T. Cuenco, Robert C. Green, Lorenzo Pinessi, Innocenzo Rainero, Sandro Sorbi, Amalia Bruni, Ranjan Duara, Robert P. Friedland, Rivka Inzelberg, Wolfgang Hampe, Hideaki Bujo, You-Qiang Song, Olav M. Andersen, Thomas E. Willnow, Neill Graff-Radford, Ronald C. Petersen, Dennis Dickson, Sandy D. Der, Paul E. Fraser, Gerold Schmitt-Ulms, Steven Younkin, Richard Mayeux, Lindsay A. Farrer, and Peter St George-Hyslop. The neuronal sortilin-related receptor sorl1 is genetically associated with alzheimer disease. *Nat Genet*, 39(2):168–177, Feb 2007.
- [207] Michael E. Harbour, Sophia Y A. Breusegem, Robin Antrobus, Caroline Freeman, Evan Reid, and Matthew N J. Seaman. The cargo-selective retromer complex is a recruiting hub for protein complexes that regulate endosomal tubule dynamics. *J Cell Sci*, 123(Pt 21):3703–3717, Nov 2010.
- [208] Emmanuel Derivery, Carla Sousa, Jérémie J. Gautier, Bérangère Lombard, Damaris Loew, and Alexis Gautreau. The arp2/3 activator wash controls the fission of endosomes through a large multiprotein complex. *Dev Cell*, 17(5):712–723, Nov 2009.
- [209] Eszter Zavodszky, Matthew N J. Seaman, Kevin Moreau, Maria Jimenez-Sanchez, Sophia Y. Breusegem, Michael E. Harbour, and David C. Rubinstein. Mutation in vps35 associated with parkinson’s disease impairs wash complex association and inhibits autophagy. *Nat Commun*, 5:3828, 2014.

-
- [210] Fabienne Ropers, Emmanuel Derivery, Hao Hu, Masoud Garshasbi, Mohsen Karbasiyan, Martin Herold, Gudrun Nürnberg, Reinhard Ullmann, Alexis Gautreau, Karl Sperling, Raymonda Varon, and Anna Rajab. Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: the wash complex member swip. *Hum Mol Genet*, 20(13):2585–2590, Jul 2011.
- [211] Caroline Freeman, Matthew N J. Seaman, and Evan Reid. The hereditary spastic paraplegia protein strumpellin: characterisation in neurons and of the effect of disease mutations on wash complex assembly and function. *Biochim Biophys Acta*, 1832(1):160–173, Jan 2013.
- [212] Paul N. Valdmanis, Inge A. Meijer, Annie Reynolds, Adrienne Lei, Patrick MacLeod, David Schlesinger, Mayana Zatz, Evan Reid, Patrick A. Dion, Pierre Drapeau, and Guy A. Rouleau. Mutations in the kiaa0196 gene at the spg8 locus cause hereditary spastic paraplegia. *Am J Hum Genet*, 80(1):152–161, Jan 2007.
- [213] Michael E. Harbour, Sophia Y. Breusegem, and Matthew N J. Seaman. Recruitment of the endosomal wash complex is mediated by the extended 'tail' of fam21 binding to the retromer protein vps35. *Biochem J*, 442(1):209–220, Feb 2012.
- [214] Emmanuèle Helfer, Michael E. Harbour, Véronique Henriot, Goran Lakisic, Carla Sousa-Blin, Larisa Volceanov, Matthew N J. Seaman, and Alexis Gautreau. Endosomal recruitment of the wash complex: active sequences and mutations impairing interaction with the retromer. *Biol Cell*, 105(5):191–207, May 2013.
- [215] Da Jia, Timothy S. Gomez, Daniel D. Billadeau, and Michael K. Rosen. Multiple repeat elements within the fam21 tail link the wash actin regulatory complex to the retromer. *Mol Biol Cell*, 23(12):2352–2361, Jun 2012.
- [216] Ian J McGough, Florian Steinberg, Da Jia, Peter A Barbuti, Kirsty J McMillan, Kate J Heesom, Alan L Whone, Maeve A Caldwell, Daniel D Billadeau, Michael K Rosen, et al. Retromer binding to fam21 and the wash complex is perturbed by the parkinson disease-linked vps35 (d620n) mutation. *Current Biology*, 24(14):1670–1676, 2014.

-
- [217] Megha Mohan and George D Mellick. Role of the vps35 d620n mutation in parkinson's disease. *Parkinsonism & Related Disorders*, 2016.
- [218] Hermann-Josef Lüdecke and Dagmar Wieczorek. Burn-mckeown syndrome. 2016.
- [219] Francois P. Bernier, Oana Caluseriu, Sarah Ng, Jeremy Schwartzentruber, Kati J. Buckingham, A Micheil Innes, Ethylin Wang Jabs, Jeffrey W. Innis, Jane L. Schuette, Jerome L. Gorski, Peter H. Byers, Gregor Andelfinger, Victoria Siu, Julie Lauzon, Bridget A. Fernandez, Margaret McMillin, Richard H. Scott, Hilary Racher, F. O. R. G. E Canada Consortium , Jacek Majewski, Deborah A. Nickerson, Jay Shendure, Michael J. Bamshad, and Jillian S. Parboosingh. Haploinsufficiency of sf3b4, a component of the pre-mrna spliceosomal complex, causes nager syndrome. *Am J Hum Genet*, 90(5):925–933, May 2012.
- [220] D Lehalle, D Wieczorek, RM Zechi-Ceide, M Rita Passos-Bueno, S Lyonnet, J Amiel, and CT Gordon. A review of craniofacial disorders caused by spliceosomal defects. *Clinical genetics*, 88(5):405–415, 2015.
- [221] Laura S Kremer, Daniel M Bader, Christian Mertes, Robert Kopajtich, Garwin Pichler, Arcangela Iuso, Tobias B Haack, Elisabeth Graf, Thomas Schwarzmayr, Caterina Terrile, et al. Genetic diagnosis of mendelian disorders via rna sequencing. *Nature communications*, 8:ncomms15824, 2017.
- [222] Andy W Pang, Jeffrey R MacDonald, Dalila Pinto, John Wei, Muhammad A Rafiq, Donald F Conrad, Hansoo Park, Matthew E Hurles, Charles Lee, J Craig Venter, et al. Towards a comprehensive structural variation map of an individual human genome. *Genome biology*, 11(5):R52, 2010.
- [223] Liang Gong, Chee-Hong Wong, Wei-Chung Cheng, Harianto Tjong, Francesca Menghi, Chew Yee Ngan, Edison T Liu, and Chia-Lin Wei. Nanopore sequencing reveals high-resolution structural variation in the cancer genome. *bioRxiv*, page 209718, 2017.
- [224] Jason D Merker, Aaron M Wenger, Tam Sneddon, Megan Grove, Zachary Zappala, Laure Fresard, Daryl Waggott, Sowmi Utiramerur, Yanli Hou, Ke-

vin S Smith, et al. Long-read genome sequencing identifies causal structural variation in a mendelian disease. *Genetics in Medicine*, 2017.

- [225] AW Kuss. Das methodische potenzial der neuen sequenziertechnologien jenseits der mutationssuche. *medizinische genetik*, 26(2):264–272, 2014.
- [226] Stephen B. Montgomery, Micha Sammeth, Maria Gutierrez-Arcelus, Radoslaw P. Lach, Catherine Ingle, James Nisbett, Roderic Guigo, and Emmanouil T. Dermitzakis. Transcriptome genetics using second generation sequencing in a caucasian population. *Nature*, 464(7289):773–777, Apr 2010.
- [227] Nicholas T. Ingolia, Gloria A. Brar, Silvia Rouskin, Anna M. McGeachy, and Jonathan S. Weissman. Genome-wide annotation and quantitation of translation by ribosome profiling. *Curr Protoc Mol Biol*, Chapter 4:Unit 4.18, Jul 2013.
- [228] Francesca Tuorto, Reinhard Liebers, Tanja Musch, Matthias Schaefer, Sarah Hofmann, Stefanie Kellner, Michaela Frye, Mark Helm, Georg Stoecklin, and Frank Lyko. Rna cytosine methylation by dnmt2 and nsun2 promotes trna stability and protein synthesis. *Nat Struct Mol Biol*, 19(9):900–905, Sep 2012.
- [229] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, Feb 2002.

Abbildungsverzeichnis

1	Schematische Darstellung der Library-Präparation.	19
2	Cluster Generation.	19
3	Sequencing-by-Synthesis.	20
4	Übersicht zur Exom-Anreicherungstechnik der Firma Agilent.	22
5	Priorisierungsmöglichkeiten von Exom-Daten.	29
6	Die Architektur des LIMS.	47
7	Die Menüleiste der Probenebene mit ihren einzelnen Funktionen.	48
8	Der Menüpunkt New Sample	49
9	Die Menüpunkte New Disease, New Cooperation und New Project.	50
10	Der Menüpunkt Overview	50
11	Übersicht über die Funktionen der Prozessebene.	51
12	Der Menüpunkt Create Library.	52
13	Übersicht des Menüpunkts Pooling.	53
14	Die Funktion New Run	54
15	Qualitätskontrolle innerhalb der Analyseebene.	56
16	Das Sample Manifest.	58
17	Ausschnitt aus dem LIMS-generierten Sample Tracking Sheet.	60
18	Übersicht zur Anwendung von 2D-Barcodes innerhalb der Probenprozessierung.	62
19	Rekalkulation von genomischen Library-Pools.	64
20	Zeitlicher Verlauf der prozentualen Kontamination von Exom-Libraries.	66
21	Kontamination exomischer Libraries auf dem HiSeq2500 im Vergleich zum HiSeq4000	67
22	Schematische Darstellung des Adapter Hoppings.	67
23	Kontamination genomischer Libraries.	68
24	Schematische Darstellung des Adapter-Blocking.	69
25	Übersicht der Arbeitsschritte einer Library-Präparation im manuellen, semi-automatisierten und automatisierten Modus.	72
26	Vergleichende graphische Darstellung der durchschnittlichen Abdeckung eines Exoms.	73
27	Übersicht zur On-Bait-Rate.	73
28	Die Duplikatrate im Überblick.	74

29	Die Schritte der AMPure XP Bead Aufreinigung.	75
30	Plattenanordnung zur AMPure XP Bead-Aufreinigung.	76
31	Aufbau der Bravo-Oberfläche zur AMPure XP Bead-Aufreinigung.	77
32	Abfrage der Parameter zur automatisierten AMPure XP Bead Auf- reinigung.	78
33	Vergleichende Darstellung der finalen Konzentrationen nach auto- matisierter versus manueller AMPure XP Bead-Aufreinigung. . . .	82
34	Anordnung der Platten vor dem Start des Anreicherungs-Protokolls.	83
35	Ausschnitt aus dem Automationsprotokoll zur Exom-Library-Präpara- tion.	90
36	Ausschnitt des <i>Lab Tracking</i> -Formulars für die Exom- <i>Library</i> -Präpara- tion.	91
37	Arbeitsablauf zur Bestimmung der DNA-Qualifikation für die <i>NG</i> - Sequenzierung.	94
38	Ergebnis der <i>Caliper LabChip GX</i> Messung von genomischer DNA in unterschiedlichen Qualitäten im Vergleich zur Agarosegel-Elektro- phorese.	98
39	Die Prozessschritte und Resultate der <i>Library</i> -Präparation und Se- quenzierung von eingesandten DNAs.	100
40	Ergebnisse einer Amelogenin-PCR.	106
41	Schematische Darstellung eines <i>Library</i> -Moleküls.	108
42	Schematische Darstellung des Effekts verschiedener <i>Insert</i> -Größen auf die Sequenzierung.	109
43	Größenverteilung einer exomischen <i>Library</i>	110
44	Resultate der Exom-Sequenzierung von <i>Libraries</i> mit 500 ng DNA- Einsatzmenge und höherer <i>Insert</i> -Größe.	113
45	Zeitachse der Evolution der <i>NGS</i> -Plattform am Institut für Hu- mangenetik.	115
46	Übersicht zum Durchsatz der jährlichen <i>Library</i> -Präparation. . . .	116
47	Anzahl der präparierten <i>Libraries</i> von 2009 bis November 2017. .	116
48	Generierte Datenmenge pro Jahr von 2009 bis 2017.	116
49	Übersicht zur Verteilung der untersuchten Erkrankungen.	117
50	Anzahl <i>NGS</i> -basierter Publikationen aus dem Institut für Human- genetik pro Jahr.	117

51	Elektropherogramme der unterschiedlichen FFPE-DNA-Qualitäten.	121
52	Elektropherogramme der Proben nach der Fragmentierung.	123
53	Elektropherogramme der <i>Libraries</i> vor der Hybridisierung.	124
54	Stammbäume der Familien A, B und C	131
55	Analyse der Moleküldynamik im wiltyp- und mutanten VPS35-C-terminalen Bereich	135
56	<i>Root Mean Square Fluctuation</i> (RMSF) Profil	136
57	Übersicht zum Zusammenbau des Spleißosoms und dem anschließenden Spleißprozess.	139
58	Genstruktur von <i>TXNL4A</i> mit den durch MLPA sowie Genom- und Exom-Sequenzierung identifizierten Mutationen.	141
59	Identifizierte Mikrodeletionen nach der Untersuchung der DNA mittels <i>Mikroarray</i>	142
60	Stammbäume der Indexpatienten mit BMKS und biallelen Mutationen in <i>TXNL4A</i>	144
61	Der Retromer Komplex.	167

Tabellenverzeichnis

1	Vergleich der gegenwärtig aktuellen Anreicherungskits.	23
2	Definition der Parameter zur Beurteilung der Sequenzierqualität. .	26
3	Übersicht zur Art und Anzahl von genetischen Varianten in einem Exom.	27
4	Vergleich der gemessenen Quantitäten zwischen UV-Vis und fluoreszenzbasierten Messmethodiken.	95
5	Übersicht zur Sequenzier-Statistik von Exomen mit unterschiedlichen Reinheiten.	96
6	Übersicht zur Sequenzier-Statistik von unaufgereinigter DNA versus aufgereinigter DNA.	102
7	Optimierung der Laborausstattung zur Erhöhung des Durchsatzes und zur Verminderung der Anwenderzeit.	104
8	Übersicht zum Qualitätsmanagement.	105
9	Kombinatorische Anwendung dreier kommerzieller <i>Library</i> -Präparationskits zur Herstellung von exomischen <i>Libraries</i> mit weniger Einsatzmenge und größerer <i>Insert</i> -Größe.	111
10	Übersicht der Parameter von drei ausgewählten FFPE-DNAs für eine exomische <i>Library</i> -Präparation.	120
11	Übersicht zur Sequenzier-Statistik von FFPE-DNA versus nicht-FFPE-DNA.	126
12	Morbus Parkinson assoziierte Gene und genetische Loci.	129
13	Statistik der Sequenzierung	132
14	Identifizierte Varianten nach der ersten Priorisierung des Datensets	133
15	Seltene, nicht-synonyme <i>VPS35</i> -Varianten in Patienten und Kontrollen.	134
16	Statistik der Exom-Sequenzierung bei Patienten mit BMKS. . . .	141
17	Statistik der Genom-Sequenzierung bei Patienten mit BMKS. . .	143
18	Übersicht der in <i>TXNL4A</i> identifizierten Mutationen und Deletionen.	147
19	Intronische Primer für die Amplifikation der kodierenden Sequenz von <i>VPS35</i>	273

Abkürzungsverzeichnis

A	Adenin
Abb	Abbildung
AGBT	Advances in Genome Biology and Technology
Arg	Arginin
Asn	Asparagin
Asp	Asparaginsäure
BMKS	Burn-McKeown-Syndrom
bp	Basenpaar
C	Cytosin
CNV	Copy number variation
csv	comma-separated values
dATP	2'-Desoxyadenosin-5'-triphosphat
dbSNP	Single Nucleotide Polymorphism database
ddNTP	2'-Didesoxynukleosid-5'-triphosphat
DNA	Desoxyribonukleotidsäure
dNTP	2'-Desoxynukleotid-5'-triphosphat
EDTA	Ethylendiamintetraacetat
FFPE	Formalin-fixiertes Paraffin-eingebettetes Gewebe
G	Guanin
Gb	Gigabasen
Gly	Glycin
GWAS	Genomweite-Assoziationsstudie
ID	Identifikator
Ile	Isoleucin
Indel	Insertion und Deletion
KA	Kopplungsanalyse
kb	Kilobasen
KGA	Kandidatengenanalyse
KORA	Kooperative Gesundheitsforschung in der Region Augsburg
kpb	Kilo-Basenpaare
l	Liter
Leu	Leucin

LIMS	Labor-Informations- und Management-System
M	Molar
Mb	Megabasen
Met	Methionin
min	Minute
MLPA	multiplex ligation-dependent probe amplification
mRNA	messenger RNA
NaCl	Natriumchlorid
ng	Nanogramm
NGFN	Nationalen Genomforschungsnetz
NGS	Next-Generation-Sequenzierung
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerasekettenreaktion
PEG	Polyethylenglycol
qPCR	quantitative PCR
RMSF	Root-mean-square-fluctuation
RNA	Ribonukleinsäure
sec	Sekunde
Ser	Serin
SNP	single nucleotide polymorphism
snRNA	small nuclear ribonucleic acid
SNV	single nucleotide variant
SV	structural variant
T	Thymin
Tab	Tabelle
Tb	Terrabasen
Thr	Threonin
Trp	Tryptophan
TXNL4A	Thioredoxin-like protein 4A
U	Unit (Enzymeinheit)
U1-U6	spleissosomale Untereinheit 1-6
UTR	untranslatierter Bereich
VPS35	Vacuolar protein sorting-associated protein 35
WES	whole exome sequencing
WGS	whole genome sequencing

5 Anhang

5.1 Protokoll zur automatisierten Exom-*Library*-Präparation auf der *Bravo Liquid Handling Station*

BRAVO Automated Protocol for Exome Targeted Capture Library Preparation – From DNA to Exome Library*



Elisabeth Graf

Version 5.0

10.05.2012

12-48 sample Version

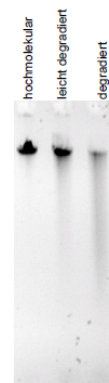
*adapted from SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library V1.1, Jan 2011

Step 1. Verifying of DNA quality and quantity

1. Use the **NanoDrop system** to determine the concentration and quality of your gDNA sample.
2. Before starting the measurement clean the NanoDrop pedestals by adding 2µl H₂O. Leave the solution on for 2 minutes.
3. Start the NanoDrop program using the password **XXX**.
4. Between each sample, wipe down the pedestals with H₂O followed by a dry tissue.
5. Follow the instructions to initialize the instrument.
6. Perform a blank measurement with 1.5µl low TE buffer.
7. Check the accuracy of the instrument by measurement of **two internal controls** (110ng/µl and 55ng/µl).
8. Proceed with the measurement of your DNA samples.
9. Therefore add **1.5µl of each sample** and add the sample ID.
10. Evaluate the output graph and quality data.
11. The gDNA must be of high quality with an A260/A280 ratio between 1.8 and 2.0. A minimal concentration of 60ng/µl is required for library preparation.
12. Import the concentration and A260/A280 ratio of each sample into the sample tracking sheet generated by the *Solexa* database.

14. Prepare an **0.7 % agarose gel and run appr. 100 ng** per sample to evaluate that the genomic DNA is of high quality.

0.7 % Agarose Gel
50-100 ng Probenauftrag



Step 2. Shear genomic DNA

1. Use the NanoDrop system to determine the concentration of your gDNA sample. Make sure the gDNA is of high quality (non-degraded, A260/A280 is 1.8 to 2.0). Follow the instructions for the instrument.
2. Dilute **3 µg of high-quality gDNA** with 1X Low TE Buffer in a 1.5-mL LoBind tube to a **total volume of 50 µL in a clean area (pre-PCR)**.
3. Set up the Covaris instrument.
 - Check that the Covaris tank is filled with fresh deionized water to water fill **level 6**.
 - Set the chiller temperature to 4°C to ensure that the temperature reading in the water bath displays 7°C.
 - Degas the instrument for at least 45 minutes before use.

Refer to the Covaris instrument user guide.

4. Use a tapered pipette tip to slowly transfer the 50 µL DNA sample through the pre-split septa.
5. Be careful not to introduce a bubble into the bottom of the tube.
6. Secure the microTube in the tube holder and shear the DNA with the settings in below. For each DNA sample to be sequenced, prepare 1 library.

"150bp SureSelect" Covaris E210 shear settings

Setting	Value
Duty Factor	20%
PIP	175
Cycles per Burst	200
Time	165s
Temperature	4° to 7° C



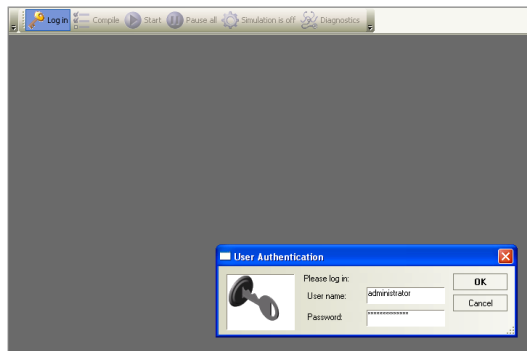
7. Put the Covaris microTube back into the loading and unloading station.
8. While **keeping the snap-cap on**, insert a pipette tip through the pre-split septa, then slowly remove the sheared DNA.
9. Transfer the sheared DNA into a new Eppendorf twin.tec 96 well PCR plate.
10. Proceed to the automated AMPure XP bead clean up.

BRAVO first Steps

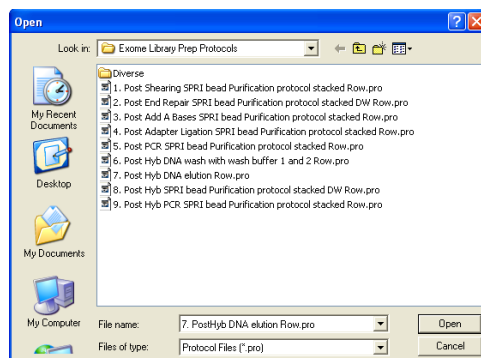
1. Switch on the computer and the BRAVO system.
2. open VWorks software.



3. Log in.



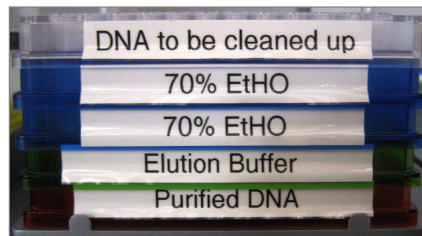
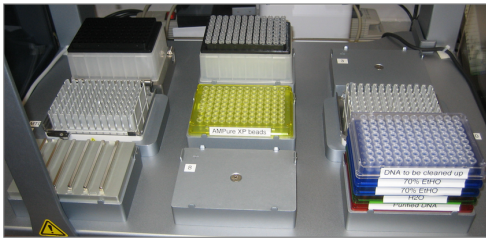
4. Open folder "Exome Library Prep Protocols".



5. Open Protocol "1. Post Shearing SPRI bead Purification protocol stacked Row.pro".
6. Agree to initialize the BRAVO system.
7. The BRAVO system is now ready to use.

Step 3. Purify the sheared sample using Agencourt AMPure XP beads 3.0x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **150 µl of homogenous Agencourt AMPure XP** beads in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **55µl of nuclease-free water** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110µl of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing sheared DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "1. Post Shearing SPRI bead Purification protocol stacked Row.pro" protocol and follow the instructions.

12. Enter parameters for the purification step as follows:

- | | |
|----------------------|-----|
| a. DNA to bead ratio | 3.0 |
| b. DNA_vol | 52 |
| c. elute_vol | 50 |
| d. samples | 48 |

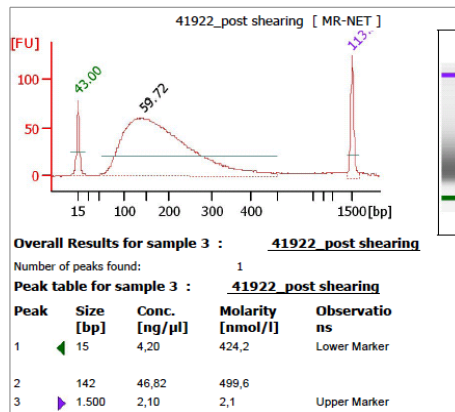


13. Once the protocol has finished assess quality with the Agilent 2100 Bioanalyzer.

SAFE STOPPING POINT (-20°C)

Step 4. Assess quality with the Agilent 2100 Bioanalyzer (DNA 1000 Kit)

1. Let the Bioanalyzer DNA 1000 Kit come to room temperature at least 30 minutes.
2. Check that the 2100 Bioanalyzer electrodes have been cleaned as instructed in the reagent kit guide.
3. Turn on the 2100 Bioanalyzer, open the Agilent 2100 Expert Software and check communication.
4. Within the instrument context, choose the DNA 1000 assay from the drop down list.
5. Enter sample names and comments in the Data and Assay context.
6. Mix the samples carefully.
7. Prepare the chip, samples and ladder as instructed in the reagent kit guide.
8. Load the prepared chip into the 2100 Bioanalyzer and start the run within five minutes after preparation.
9. Set the peak from the beginning of the curve to 500bp. Verify the results. Check that the electropherogram shows a distribution with a peak height between 50 to 500 nucleotides (average 135bp) and a minimal concentration of 20ng/μl (average 45ng/μl).



10. Import ng/μl, nM and bp of each sample into the sample tracking sheet generated by the *Solexa* database.
11. Proceed with the “end repair” protocol.

SAFE STOPPING POINT (-20°C)



Step 5. Repair the ends of fragmented DNA

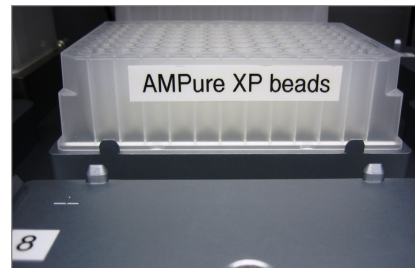
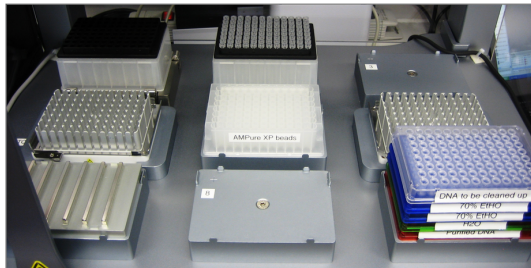
1. Master mix for multiple libraries (prepare on ice):
 - Prepare the reaction mix as follows (no excess included):

Reagent	Volume for 1 Library	Volume for 12 Libraries	Volume for 24 Libraries	Volume for 48 Libraries
Nuclease-free water	35,2	422,4	844,8	1689,6
10x End Repair Buffer (clear cap)	10	120	240	480
dNTPmix (green cap)	1,6	19,2	38,4	76,8
T4 DNA Polymerase (purple cap)	1	12	24	48
Klenow DNA Polymerase (yellow cap)	2	24	48	96
T4 Polynucleotide Kinase (orange cap)	2,2	26,4	52,8	105,6

- Add **51 μ L of the reaction mix** to each well or tube.
 - Add **49 μ L of each DNA sample** to each well or tube. Mix by pipetting. Change pipette tips between samples.
 - Incubate in a thermal cycler for **30 minutes at 20°C**. Do not use a heated lid.
2. Once the incubation step has completed, proceed to automated SPRI bead Purification protocol.

Step 6. Purify the end-repaired samples using Agencourt AMPure XP beads 2.2x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **220 μ l of homogenous Agencourt AMPure XP beads** in a 96 well Nunc 1ml Deep Well plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **37 μ l of nuclease-free water** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110 μ l of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing end-repaired DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "2. Post End Repair SPRI bead Purification protocol stacked DW Row.pro" protocol and follow the instructions.

12. Enter parameters for the purification step as follows:

- | | |
|----------------------|-----|
| a. DNA to bead ratio | 2.2 |
| b. DNA_vol | 102 |
| c. elute_vol | 32 |
| d. samples | 48 |



13. Proceed to "Add 'A' bases to 3' end of the DNA fragments" protocol.

SAFE STOPPING POINT (-20°C)

Step 7. Add 'A' bases to the 3' end of end-repaired DNA fragments

1. Master mix for multiple libraries (prepare on ice):

- Prepare the reaction mix as follows (no excess included):

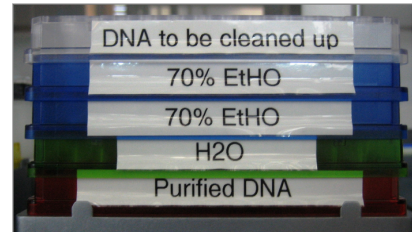
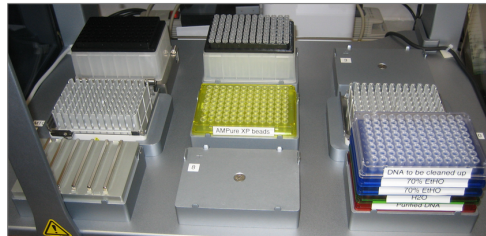
Reagent	Volume for 1 Library	Volume for 12 Libraries	Volume for 24 Libraries	Volume for 48 Libraries
Nuclease-free water	11	132	264	528
10x Klenow DNA Polymerase Buffer (blue cap)	5	60	120	240
dATP (green cap)	1	12	24	48
Exo (-) Klenow DNA Polymerase (red cap)	3	36	72	144

- Add **20 µL of the reaction mix** to each well or tube.
- Add **30 µL of each DNA sample** to each well or tube. Mix by pipetting. Change pipette tips between samples.
- Incubate in a thermal cycler for **30 minutes at 37°C**.
If you use a heated lid, make sure that the lid temperature does not exceed 50°C.

2. Once the incubation step has completed, proceed to automated SPRI bead Purification protocol.

Step 8. Purify the 'A' bases added samples using Agencourt AMPure XP beads 2.2x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **110 µl of homogenous Agencourt AMPure XP beads** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **35µl of nuclease-free water** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110µl of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing 'A' base added DNA at deck position 9 on top of the ethanol plates.



St
rt

Bravo automated "3. Post Add A Bases SPRI Bead Purification protocol stacked Row.
pro" protocol and follow the instructions.

12. Enter parameters for the purification step as follows:

- | | |
|----------------------|-----|
| a. DNA to bead ratio | 2.2 |
| b. DNA_vol | 52 |
| c. elute_vol | 30 |
| d. samples | 48 |



13. Proceed to "ligate the indexing-specific paired-end adapter" protocol.

Step 9. Ligate indexing-specific paired-end adapter

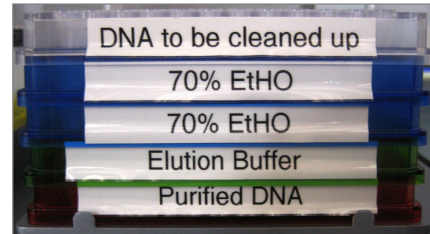
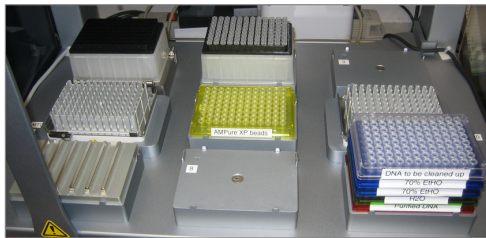
1. Master mix for multiple libraries (prepare on ice):
 - Prepare the reaction mix as follows (no excess included):

Reagent	Volume for 1 Library	Volume for 12 Libraries	Volume for 24 Libraries	Volume for 48 Libraries
5x T4 DNA ligase buffer (green cap)	10	120	240	480
InPE Adapter oligo mix (brown cap)	10	120	240	480
T4 DNA ligase (red cap)	1,5	18	36	72

- Add **21,5 µL of the reaction mix** to each well or tube.
 - Add **30 µL of each DNA sample** to each well or tube. Mix by pipetting. Change pipette tips between samples
 - Incubate for **15 minutes at 20°C** on a thermal cycler. Do not use a heated lid.
2. Once the incubation step has completed, proceed to automated SPRI bead Purification protocol.

Step 10. Purify the adapter ligated samples using Agencourt AMPure XP beads 0.7x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **35 µl of homogenous Agencourt AMPure XP beads** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **37µl of nuclease-free water** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110µl of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing adapter ligated DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "4. Post Ligation SPRI bead Purification protocol stacked Row. pro" protocol and follow the instructions.

12. Enter parameters for the purification step as follows:

- | | |
|----------------------|-----|
| a. DNA to bead ratio | 0.7 |
| b. DNA_vol | 52 |
| c. elute_vol | 32 |
| d. samples | 48 |



13. Once the protocol has finished proceed to amplify adapter-ligated library protocol.

SAFE STOPPING POINT (-20°C)

Step 11. Amplify adapter-ligated library

1. Master mix for multiple libraries (prepare on ice):

- Prepare the reaction mix as follows (excess included):

Reagent	Volume for 1 Library	Volume for 14 Libraries	Volume for 28 Libraries	Volume for 51 Libraries
Nuclease-free water	6	84	150	306
InPE primer 1.0 (brown cap)	1,25	17,5	35	63,75
SureSelect Indexing Pre-Capture PCR Reverse Primer (clear cap)	1,25	17,5	35	63,75
Herculase II 5x Reaction Buffer	10	140	280	510
dNTP mix (green cap)	0,5	7	14	25,5
Herculase II Fusion DNA Polymerase (red cap)	1	14	28	51

- Add **20 µL of the reaction mix** to each well or tube.
- Add **30 µL of each DNA sample** to each well or tube. Mix by pipetting. Change pipette tips between samples

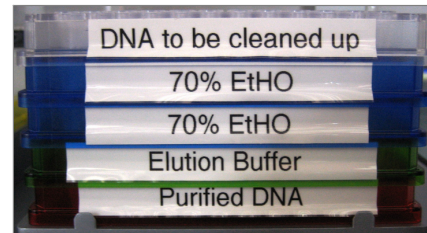
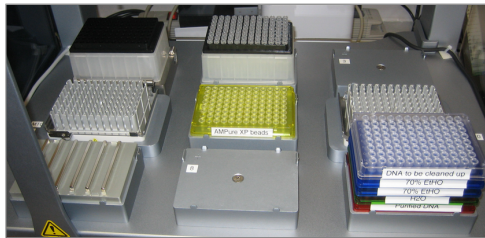
2. Run the program in a thermal cycler as follows:

Step	Temperature	Time
Step 1	98°C	2 minutes
Step 2	98°C	30 seconds
Step 3	65°C	30 seconds
Step 4	72°C	1 minute
Step 5		Repeat Step 2 through 4 for a total of 5 times
Step 6	72°C	10 minutes
Step7	4°C	Hold

3. Once the incubation step has completed, proceed to automated SPRI bead Purification protocol.

Step 12. Purify the amplified samples using Agencourt AMPure XP beads 1.8x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **90µl of homogenous Agencourt AMPure XP** beads in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **35µl of nuclease-free water** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110µl of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing amplified DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "5. Post PCR SPRI bead Purification protocol stacked Row.pro" protocol and follow the instructions.
12. Enter parameters for the purification step as follows:

- | | |
|----------------------|-----|
| a. DNA to bead ratio | 1.8 |
| b. DNA_vol | 52 |
| c. elute_vol | 30 |
| d. samples | 48 |

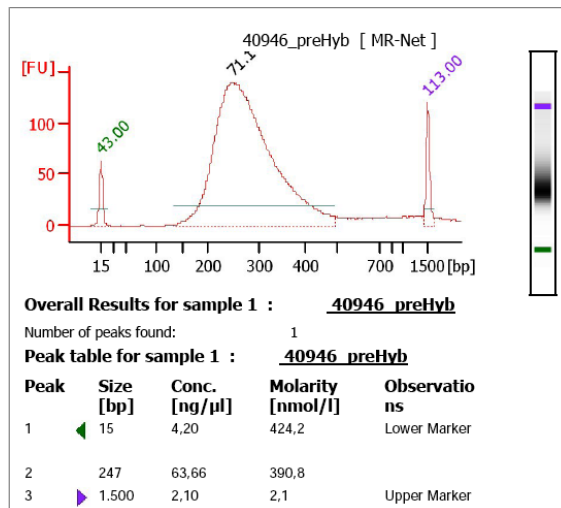


13. Once the protocol has finished assess quality and quantity with the Agilent 2100 Bioanalyzer.
- 14.

SAFE STOPPING POINT (-20°C)

Step 13. Assess quality and quantity with the Agilent 2100 Bioanalyzer (DNA 1000 Kit)

1. Let the Bioanalyzer DNA 1000 Kit come to room temperature at least 30 minutes.
2. Check that the 2100 Bioanalyzer electrodes have been cleaned as instructed in the reagent kit guide.
3. Turn on the 2100 Bioanalyzer, open the Agilent 2100 Expert Software and check communication.
4. Within the instrument context, choose the DNA 1000 assay from the drop down list.
5. Enter sample names and comments in the Data and Assay context.
6. Mix the samples carefully.
7. Prepare the chip, samples and ladder as instructed in the reagent kit guide.
8. Load the prepared chip into the 2100 Bioanalyzer and start the run within five minutes after preparation.
9. **Set the peak from the beginning of the curve to 700bp.** Verify the results. Check that the electropherogram shows a distribution with a peak height between 150 to 500 nucleotides (average 245bp) and a average concentration of 60ng/μl.



10. A minimum of 500ng of library is required for hybridization.
11. Import ng/μl, nM and bp of each sample into the sample tracking sheet generated by the *Solexa* database.
12. Proceed with the "Hybridize the libraries" protocol.

SAFE STOPPING POINT (-20°C)

Step 14. Hybridize the library

1. For each sample library prepared, do one hybridization and capture. Do not pool samples at this stage.
2. For each sample concentrate a 500ng aliquot at < 45°C down and resuspend in 3,4µl water to a concentration of 147ng/µl. Mix on a vortex mixer.
3. Prepare the following mixtures:
 - a. **SureSelect Hybridization buffer** (prepare at RT):
 - Prepare the reaction mix as follows (excess included)

Reagent	Volume for 1 Library	Volume for 12 Libraries	Volume for 24 Libraries	Volume for 48 Libraries
SureSelect Hyb #1 (orange cap)	25	300	600	1200
SureSelect Hyb #2 (red cap)	1	12	24	48
SureSelect Hyb #3 (yellow cap)	10	120	240	480
SureSelect Hyb #4 (black cap)	13	156	312	624

- b. **SureSelect RNase blocker dilution 1:3** (prepare on ice):
 - Prepare the reaction mix as follows (excess included)

Reagent	Volume for 1 Library	Volume for 12 Libraries	Volume for 24 Libraries	Volume for 48 Libraries
RNase blocker (purple cap)	1	12	24	48
Nuclease free water	2	24	48	96

- c. **SureSelect Capture library** (prepare on ice):
 - Prepare the reaction mix as follows (no excess included)

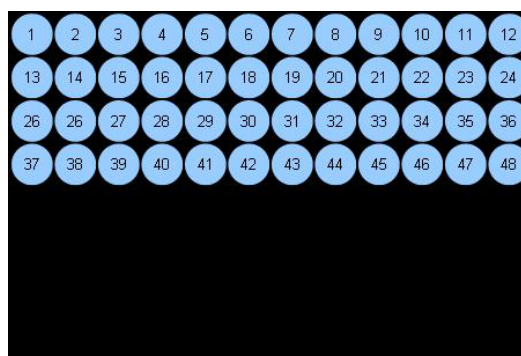
Reagent	Volume for 1 Library	Volume for 12 Libraries	Volume for 24 Libraries	Volume for 48 Libraries
SureSelect Capture Library (red cap)	5	60	120	240
RNase blocker dilution	2	24	48	96

- d. **SureSelect Block mix** (prepare on ice):

- Prepare the reaction mix as follows (no excess included)

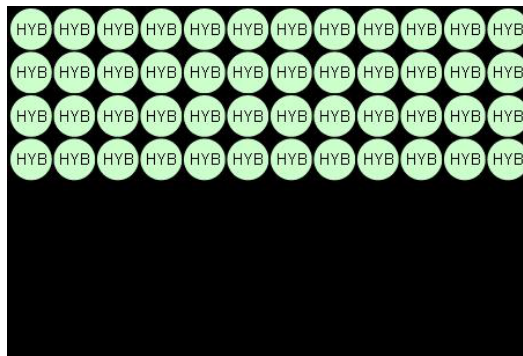
Reagent	Volume for 1 Library	Volume for 12 Libraries	Volume for 24 Libraries	Volume for 48 Libraries
SureSelect Indexing Block #1 (green cap)	2,5	30	60	120
SureSelect Block #2 (blue cap)	2,5	30	60	120
SureSelect Indexing Block #3 (brown cap)	0,6	7,2	14,4	28,8

4. On ice: Add 5,6µl SureSelect Block mix into an Eppendorf twin.tec plate ("library plate").
Add 3,4µl of 147ng/µl prepped library to each well and mix well by pipetting up and down.



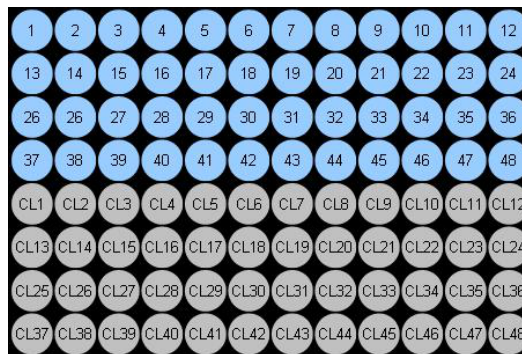
„library plate“ containing DNA libraries

5. Seal each well of the "library plate" with the appropriate Eppendorf cap strips horizontally.
6. Label the caps of the wells containing the DNA library with continuous numbers (i.e. 1-48).
7. Label the caps of the wells designated for Capture library with "CL" and continuous numbers (i.e. CL1-CL48).
8. Incubate the "library plate" 5 minutes at 95°C in a thermal cycler (PeqLab) to denature the double stranded DNA library. Use a heated lid at 105°C.
9. After the denature protocol has finished cool down the thermal cycler to 65°C. Use a heated lid at 105°C.
10. Simultaneously heat up the second thermal cycler to 65°C with a lid temperature at 105°C.
11. In a second Eppendorf twin.tec plate ("buffer plate") aliquot 40µl hybridization buffer and seal the wells with the appropriate Eppendorf cap strips horizontally.



„buffer plate“ containing hybridization buffer

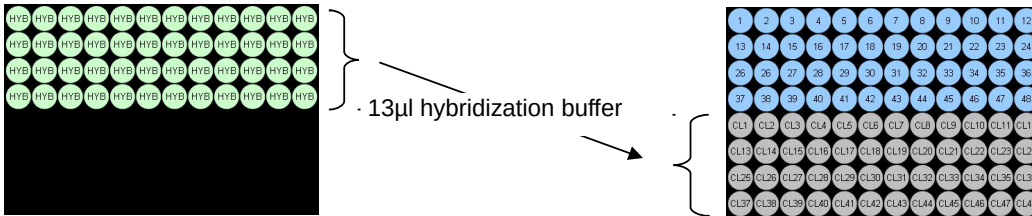
12. Label the caps of the wells containing the hybridization buffer with “Hyb”.
13. Incubate the “buffer plate” 5 minutes at 65°C.
14. Maintain the “library plate” at 65°C while you open the caps of the wells designated for the SureSelect capture library.
15. Aliquot 7µl SureSelect capture library into the empty wells of the “library plate”.



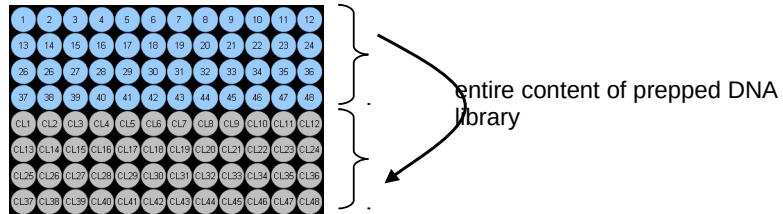
„library plate“ containing DNA libraries and SureSelect capture libraries

16. Seal the wells with fresh appropriate Eppendorf cap strips horizontally.
17. Incubate the “library plate” 2 minutes at 65°C.

18. Maintain the plates at 65°C while you use a multichannel pipette to take 13µl of hybridization buffer from the “buffer plate” and add it to the SureSelect capture library contained in the “library plate” for each sample.



19. Maintain the “library plate” at 65°C while you use a multichannel pipette to transfer the entire contents of each prepped DNA library to the SureSelect capture library mix.



20. Mix the hybridization mix containing hybridization buffer, capture library and DNA library well by slowly pipetting up and down 8 to 10 times.
21. Close the caps of the wells immediately after mixing to avoid excessive evaporation. Make sure all wells are completely sealed.
22. The hybridization mix is now 27 to 29µl, depending on degree of evaporating during preincubations and pipetting steps.
23. Incubate the hybridization mix for **48h at 65°C** with a heated lid at 105°C.
24. Check in between that all wells are closed during incubation.

Step 15. Prepare Dynal MyOne Streptavidin T1 magnetic beads

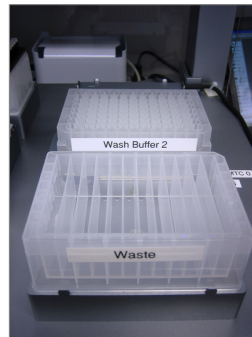
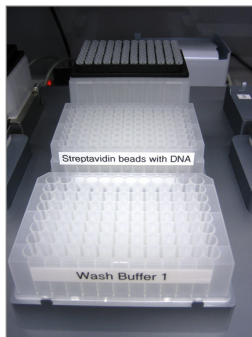
1. Vigorously resuspend the Dynal MyOne Streptavidin T1 magnetic beads on a vortex mixer.
2. For each hybridization, add 50 μ l Dynal magnetic beads to a 15ml tube.

Reagent	Volume for 1 Library	Volume for 13 Libraries	Volume for 25 Libraries	Volume for 50 Libraries
Dynal MyOne Streptavidin T1 magnetic beads	50	650	1250	2500
SureSelect Binding buffer	200	2600	5000	10.000

3. Wash the beads:
 - a. For each hybridization, add 200 μ l SureSelect Binding Buffer.
 - b. Mix the beads on a vortex mixer for 5 seconds.
 - c. Put the tube in the DynaMag 15 magnetic device.
 - d. Carefully remove and discard the supernatant.
 - e. Repeat step a through step d for a total of three washes.
4. For each hybridization, resuspend the beads in 200 μ l SureSelect binding buffer.
5. Aliquot 200 μ l of the homogenized beads for each sample to a 1,2ml deep well plate (NUNC).

Step 16. Select hybrid capture with SureSelect

1. Wipe down Agilent Bravo deck with 70% Ethanol.
2. Preheat the positions **4 and 6 to 65°C**.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **510 µl of SureSelect wash buffer 1** in NUNC 1,2ml deep well plate, then place plate at deck position 8.
6. Aliquot **1900 µl of SureSelect wash buffer 2** in NUNC 2,0ml deep well plate, then place plate at deck position 6 at 65°C.
7. Place Dynal MPC-92S magnet at deck position 7.
8. Place the waste reservoir on position 9.
9. Add the hybridized **DNA to the washed streptavidin beads** and place plate at position 5.
10. Start Bravo automated "6.Post Hyb DNA wash with wash buffer 1 and 2 Row.pro" protocol and follow the instructions.



11. Once the protocol has finished, continue with the "4. PostHyb DNA elution" protocol immediately.

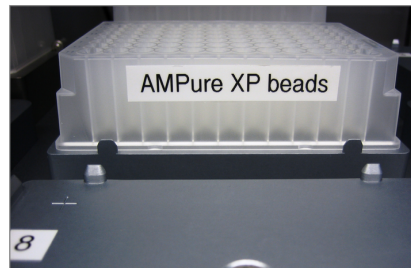
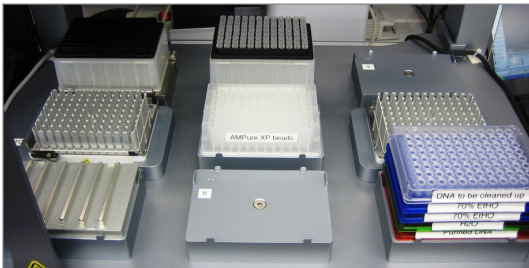
Step 17. DNA elution

1. Place an empty tip box on position 1.
2. Place a full tip box on position 2.
3. Place a Eppendorf twin.tec plate containing **55µl SureSelect Elution Buffer** of deck position 8.
4. Place a Eppendorf twin.tec plate containing **50µl SureSelect Neutralization Buffer** on deck position 9.
5. Start Bravo automated "7. Post Hyb DNA elution Row .pro" protocol.
6. Once the protocol has finished proceed to amplify enriched exome library protocol.



Step 18. Purify the captured libraries using Agencourt AMPure XP beads 1.8x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **180 µl of homogenous Agencourt AMPure XP beads** in a 96 well Nunc 1ml Deep Well plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **35µl of nuclease-free water buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110µl of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing captured DNA library at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "8. Post Hyb SPRI bead Purification protocol DW Row.pro" protocol and follow the instructions.

12. Enter parameters for the purification step as follows:

e. DNA to bead ratio	1.8
f. DNA_vol	102
g. elute_vol	30
h. samples	48



13. Proceed to "Amplify enriched exome library and add index tags" protocol.

SAFE STOPPING POINT (-20°C)

Step 19. Amplify enriched exome library and add index tags

1. Master mix for multiple libraries (prepare on ice):
 - Prepare the reaction mix as follows (excess included)

Reagent	Volume for 1 Library	Volume for 15 Libraries	Volume for 28 Libraries	Volume for 52 Libraries
Nuclease-free water	22,5	337,5	630	1170
Herculase II 5x Reaction Buffer (clear cap)	10	150	280	520
100 nM dNTP mix (green cap)	0,5	7,5	14	26
Herculase II Fusion DNA Polymerase (red cap)	1	15	28	52
SureSelect Indexing Post-Capture PCR forward primer (orange cap)	1	15	28	52
PCR PrimerIndex 1 through Index 12 (clear caps)	1			

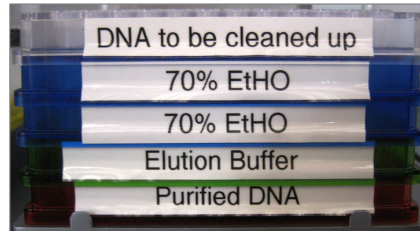
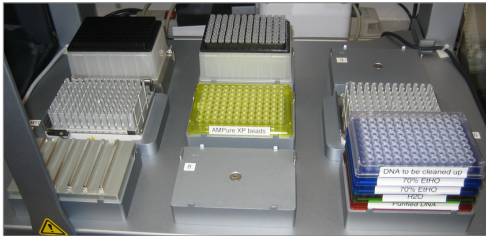
- Add **35 μ L of the reaction mix** to each well of an Eppendorf twin.tec plate..
 - Add **15 μ L of each DNA library** to each well or tube. Mix by pipetting. Change pipette tips between samples
2. Run the program in a thermal cycler as follows:

Step	Temperature	Time
Step 1	98°C	2 minutes
Step 2	98°C	30 seconds
Step 3	57°C	30 seconds
Step 4	72°C	1 minute
Step 5		Repeat Step 2 through Step 4 for a total of 12 times
Step 6	72°C	10 minutes
Step 7	4°C	Hold

3. Once the incubation step has completed, proceed to automated SPRI bead Purification protocol.

Step 20. Purify the amplified DNA libraries using Agencourt AMPure XP beads 1.8x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **90 µl of homogenous Agencourt AMPure XP beads** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **35µl of nuclease-free water buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **110µl of 70% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the water plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing adapter ligated DNA at deck position 9 on top of the ethanol plates.



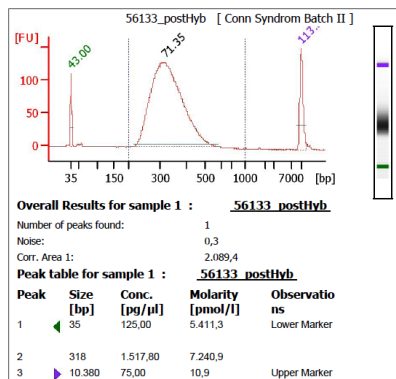
11. Start Bravo automated "9. Post Hyb SPRI bead Purification protocol stacked Row.pro" protocol and follow the instructions.
12. Enter parameters for the purification step as follows:

e. DNA to bead ratio	1,8
f. DNA_vol	52
g. elute_vol	32
h. samples	48
14. Once the protocol has finished proceed to quality measurement with the Bioanalyzer High Sensitivity DNA Kit.

SAFE STOPPING POINT (-20°C)

Step 21. Assess quality and quantity with the Agilent 2100 Bioanalyzer (DNA HS Kit)

1. Let the Bioanalyzer DNA High Sensitivity Kit come to room temperature at least 30 minutes.
2. Check that the 2100 Bioanalyzer electrodes have been cleaned as instructed in the reagent kit guide.
3. Turn on the 2100 Bioanalyzer, open the Agilent 2100 Expert Software and check communication.
4. Within the instrument context, choose the DNA High Sensitivity assay from the drop down list.
5. Enter sample names and comments in the Data and Assay context.
6. Mix the samples carefully.
7. Prepare the chip, samples and ladder as instructed in the reagent kit guide.
8. Load the prepared chip into the 2100 Bioanalyzer and start the run within five minutes after preparation.
9. Set the peak from the beginning of the curve to the end of the curve. Verify the results. Check that the electropherogram shows a distribution with a peak height between 200 to 700 nucleotides (average 300bp) and an average molarity of 10nM.



10. Import ng/µl, nM and bp of each sample into the sample tracking sheet generated by the *Solexa* database.
11. Dilute the samples to 10 nM stocks if sample concentration is above 15 nM and transfer them in 2D barcoded tubes and label them.
12. Once the quality measurement has finished proceed with the PicoGreen Quantification and import the final sample tracking sheet into the database.

5.2 Protokoll zur automatisierten Genom-*Library*-Präparation
auf der *Bravo Liquid Handling Station*

BRAVO Automated Protocol for Whole Genome Library Preparation – From DNA to Whole Genome Library*



Elisabeth Graf
Version 2.0
10.05.2012
12 to 48 sample version

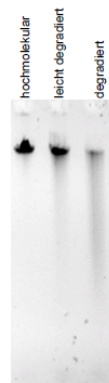
*adapted from TruSeq DNA Sample Preparation Guide (Illumina) and the SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library V1.1, Jan 2011 (Agilent)

Step 1. Measurement of DNA quality and quantity with the NanoDrop System

1. Use the **NanoDrop system** to determine the concentration and quality of your gDNA sample.
2. Before starting the measurement clean the NanoDrop pedestals by adding 2µl H₂O. Leave the solution on for 2 minutes.
3. Start the NanoDrop program using the password **XXX**.
4. Between each sample, wipe down the pedestals with H₂O followed by a dry tissue.
5. Follow the instructions to initialize the instrument.
6. Perform a blank measurement with 1.5µl low TE buffer.
7. Check the accuracy of the instrument by measurement of **two internal controls** (110ng/µl and 55ng/µl).
8. Proceed with the measurement of your DNA samples.
9. Therefore add **1.5µl of each sample** and add the sample ID.
10. Evaluate the output graph and quality data.
11. The gDNA must be of high quality with an A260/A280 ratio between 1.8 and 2.0.
12. Import the concentration and A260/A280 ratio of each sample into the sample tracking sheet generated by the *Solexa* database.

14. Prepare an **0.7 % agarose gel and run appr. 100 ng** per sample to evaluate that the genomic DNA is of high quality.

0.7 % Agarose Gel
50-100 ng Probenauftrag



Step 2. Shear genomic DNA

1. Use the NanoDrop system to determine the concentration of your gDNA sample. Make sure the gDNA is of high quality (non-degraded, A260/A280 is 1.8 to 2.0). Follow the instructions for the instrument.
2. Dilute **1 µg of high-quality gDNA** with TE Buffer in a 1.5-mL LoBind tube to a **total volume of 120 µL**.
3. Set up the Covaris instrument.
 - Check that the Covaris tank is filled with fresh deionized water to **water fill level 12**.
 - Check that the water covers half of the visible glass part of the tube.
 - Set the chiller temperature to 4°C to ensure that the temperature reading in the water bath displays 7°C.
 - Degas the instrument for least 45 minutes before use.
 - Refer to the Covaris instrument user guide.
4. Use a tapered pipette tip to slowly transfer the 120 µL DNA sample through the pre-split septa.
5. Be careful not to introduce a bubble into the bottom of the tube.
6. Secure the microTube in the tube holder and shear the DNA with the settings in below. For each DNA sample to be sequenced, prepare 1 library.

"Illumina 350bp" Covaris E210 shear settings

Setting	Value
Duty Factor	5%
PIP	175
Cycles per Burst	1000
Time	225s
Temperature	4° to 7° C



7. Put the Covaris microTube back into the loading and unloading station.
8. While keeping the snap-cap on, insert a pipette tip through the pre-split septa, then slowly remove the sheared DNA.
9. Transfer the sheared DNA into a new Eppendorf twin.tec 96 well PCR plate.
10. Proceed immediately to the automated AMPure XP bead clean up.

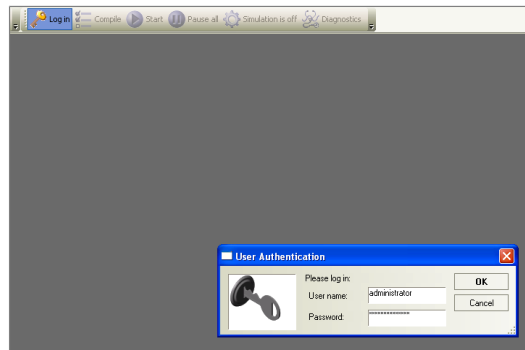
BRAVO first Steps

1. Switch on the computer and the BRAVO system.

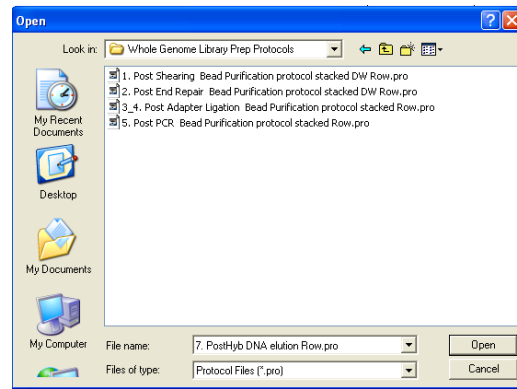
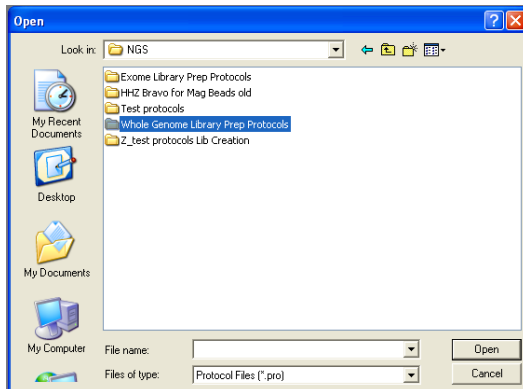
2. open VWorks software.



3. Log in.



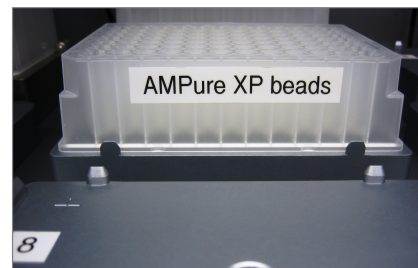
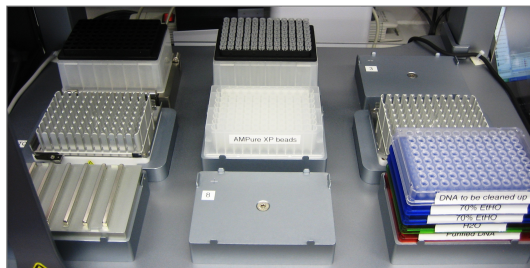
4. Open folder "**Whole Genome Library Prep Protocols**".
5. Open Protocol "1. Post Shearing Bead Purification stacked DW Row.pro".



6. Agree to initialize the BRAVO system.
7. The BRAVO system is now ready to use.

Step 3. Purify the sheared samples using Agencourt AMPure XP beads 0.75x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **90 µl of homogenous Agencourt AMPure XP beads** in a 96 well Nunc 1ml Deep Well plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **53µl of Illumina Resuspension Buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **180µl of 80% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the buffer plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing end-repaired DNA at deck position 9 on top of the ethanol plates.



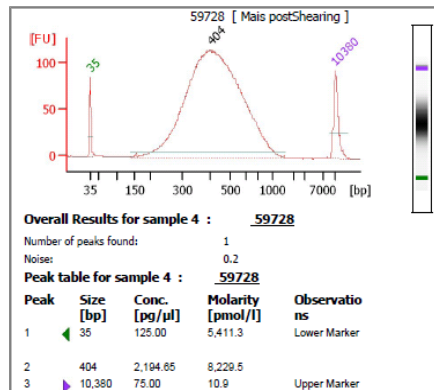
11. Start Bravo automated "1. Post Shearing Bead Purification stacked DW Row.pro" protocol and follow the instructions.
12. Enter parameters for the purification step as follows:
 - a. **DNA to bead ratio** **0.75**
 - b. **DNA_vol** **122**
 - c. **elute_vol** **50**
 - d.
13. samples 12 (or multiple of 12) Import ng/µl, nM and bp of each sample into the sample tracking sheet generated by the *So/lexa* database.
14. Proceed to "End repair of the DNA fragments" protocol.



SAFE STOPPING POINT (-20°C)

Step 4. Assess quality with the Agilent 2100 Bioanalyzer (DNA HS Kit)

2. Let the Bioanalyzer DNA 1000 Kit come to room temperature at least 30 minutes.
3. Check that the 2100 Bioanalyzer electrodes have been cleaned as instructed in the reagent kit guide.
4. Turn on the 2100 Bioanalyzer, open the Agilent 2100 Expert Software and check communication.
5. Within the instrument context, choose the DNA HS assay from the drop down list.
6. Enter sample names and comments in the Data and Assay context.
7. Mix the samples carefully.
8. Prepare the chip, samples and ladder as instructed in the reagent kit guide.
9. Load the prepared chip into the 2100 Bioanalyzer and start the run within five minutes after preparation.
10. Set the peak from the beginning of the curve to the end. Verify the results. Check that the electropherogram shows a distribution with a peak height between 200 to 1000 nucleotides (average 380bp).



11. Import ng/μl, nM and bp of each sample into the sample tracking sheet generated by the Solexa database.

SAFE STOPPING POINT (-20°C)

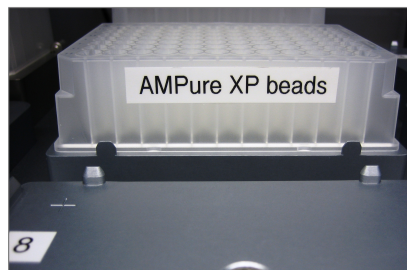
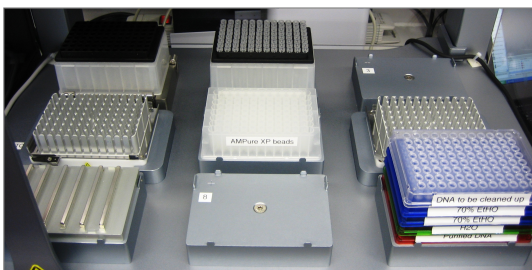


Step 5. Repair the ends of fragmented DNA

1. Add **40 µl of End-Repair Mix** to each well of the Eppendorf twin.tec plate.
2. Add **10µl Resuspention Buffer** to each well well of the Eppendorf twin.tec plate.
3. Gently mix up and down to mix the suspension.
4. Seal the plate with adhesive film.
5. Incubate in a thermal cycler for **30 minutes at 30°C**. Do not use a heated lid.
6. Once the incubation step has completed, proceed to automated SPRI bead Purification protocol.

Step 6. Purify the end-repaired samples using Agencourt AMPure XP beads 1.6x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **160 µl of homogenous Agencourt AMPure XP beads** in a 96 well Nunc 1ml Deep Well plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **25µl of Illumina Resuspension Buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **180µl of 80% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the buffer plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing end-repaired DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated “2. End Repair Bead Purification stacked DW Row.pro” protocol and follow the instructions.

12. Enter parameters for the purification step as follows:

- | | |
|----------------------|------------------------|
| a. DNA to bead ratio | 1.6 |
| b. DNA_vol | 102 |
| c. elute_vol | 20 |
| d. samples | 12 (or multiple of 12) |



13. Import ng/μl, nM and bp of each sample into the sample tracking sheet generated by the *Solexa* database.

14. Proceed to “Add ‘A’ bases to 3’ end of the DNA fragments” protocol.

SAFE STOPPING POINT (-20°C)

Step 7. Add ‘A’ bases to the 3’ end of end-repaired DNA fragments

13. Add **12.5 μl of A-Tailing Mix** to each well of the Eppendorf twin.tec plate.

14. Gently mix up and down to mix the suspension.

15. Seal the plate with adhesive film.

16. Incubate in a thermal cycler for **30 minutes at 37°C**. Do not use a heated lid.

17. Once the incubation step has completed, proceed immediately to the step “Ligate Adapters”. **No SPRI bead purification is required.**



Step 8. Ligate indexing-specific paired-end adapter

1. Add **2.5 μl of Resuspension Buffer** to each well of the Eppendorf twin.tec plate.

1. Add **2.5 μl of thawd DNA Adapter Index** to each well of the Eppendorf twin.tec plate.

1. Add **2.5 μl of DNA Ligase Mix** to each well of the Eppendorf twin.tec plate.

2. Gently mix up and down to mix the suspension.

3. Seal the plate with adhesive film.

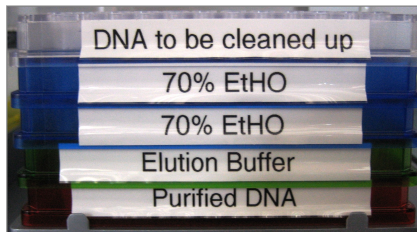
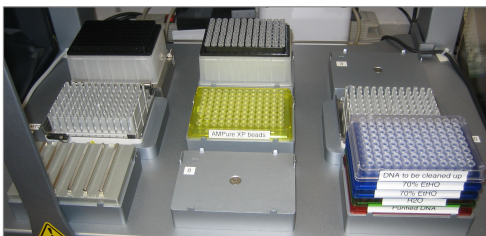
4. Incubate in a thermal cycler for **10 minutes at 30°C**. Do not use a heated lid.

5. Once the incubation step has completed, add **5μl Stop Ligase Mix** to each well of the Eppendorf twin.tec plate.

6. Once the protocol has completed, proceed to automated SPRI bead Purification protocol.

Step 9. Purify the adapter ligated samples using Agencourt AMPure XP beads 0.95x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **42.5µl of homogenous Agencourt AMPure XP beads** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **55µl of Illumina Resuspension Buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **180µl of 80% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the buffer plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing adapter ligated DNA at deck position 9 on top of the ethanol plates.

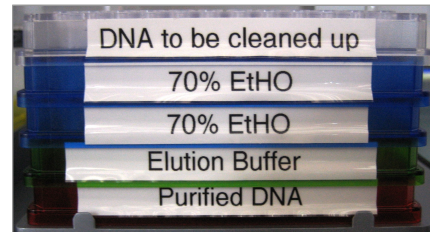
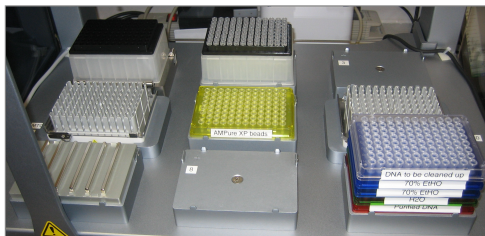


11. Start Bravo automated "3_4. Post Adapter Ligation Bead Purification stacked Row.pro" protocol and follow the instructions.
12. Enter parameters for the purification step as follows:
 - a. **DNA to bead ratio** **0.95**
 - b. **DNA_vol** **47**
 - c. **elute_vol** **52.5**
 - d. **samples** **12 (or multiple of 12)**

13. Once the protocol has finished proceed to the **second SPRI bead purification** step.

Step 10. Second Purification of adapter ligated samples using Agencourt AMPure XP beads 1.0x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **50µl of homogenous Agencourt AMPure XP beads** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **35µl of Illumina Resuspension Buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **180µl of 80% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the buffer plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing adapter ligated DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "3_4. Post Adapter Ligation Bead Purification stacked Row.pro" protocol and follow the instructions.
12. Enter parameters for the purification step as follows:
 - e. **DNA to bead ratio** **1.0**
 - f. **DNA_vol** **54**
 - g. **elute_vol** **30**
 - h. **samples** **12 (or multiple of 12)**

13. Once the protocol has finished proceed to purification of the DNA library by Pippin Prep.

SAFE STOPPING POINT (-20°C)

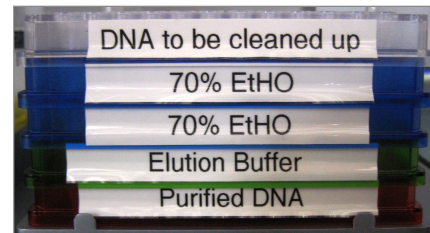
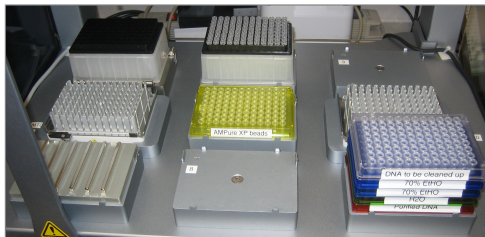
Step 11. Size Select Ligation Products

1. Set up the Pippin Prep by following the manufacturers' instructions.
2. Use the program "tight cut 500".
3. After size selecting the DNA libraries continue with an AMPure XP clean up.

SAFE STOPPING POINT (-20°C)

Step 12. Purify the size selected samples using Agencourt AMPure XP beads 1.0x

13. Let the AMPure XP beads come to room temperature for at least 30 minutes.
14. Wipe down Agilent Bravo deck with 70% Ethanol.
15. Place an empty tip box on position 1.
16. Place a full tip box on position 2.
17. Aliquot **70µl of homogenous Agencourt AMPure XP beads** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
18. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
19. Aliquot **25µl of Illumina Resuspension Buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
20. Aliquot **180µl of 80% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the buffer plate.
21. Place Dynal MPC-92S magnet at deck position 7.
22. Place the Eppendorf twin.tec plate containing adapter ligated DNA at deck position 9 on top of the ethanol plates.



23. Start Bravo automated “Post Pippin Bead Purification stacked Row.pro” protocol and follow the instructions.
24. Enter parameters for the purification step as follows:
 - a. **DNA to bead ratio** **1.0**
 - b. **DNA_vol** **70**
 - c. **elute_vol** **20**
 - d. **samples** **12 (or multiple of 12)**

e.

SAFE STOPPING POINT (-20°C)



Step 13. Amplify the Size-Selected Library

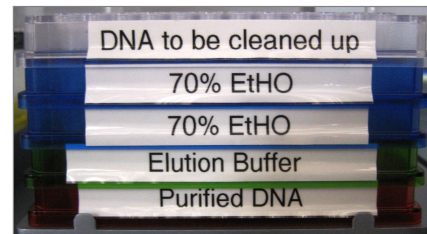
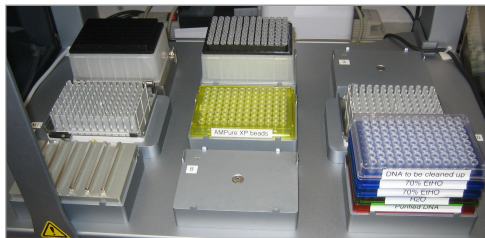
1. Add 5µl of thawed Illumina PCR Primer Cocktail to each well of the Eppendorf twin.tec plate.
2. Add 25µl thawed Illumina PCR Master Mix to each well of the PCR plate.
3. Gently mix up and down thoroughly.
4. Seal the plate with adhesive film.
5. Run the program in a thermal cycler as follows:

<u>Step</u>	<u>Temperature</u>	<u>Time</u>
Step 1	98°C	30seconds
Step 2	98°C	10seconds
Step 3	60°C	30seconds
Step 4	72°C	30seconds
Step 5		Repeat Step 2 through 4 for a total of 10 times
Step 6	72°C	5minutes
Step7	4°C	Hold

3. Once the incubation step has completed, proceed to automated SPRI bead Purification protocol.

Step 14. Purify the amplified samples using Agencourt AMPure XP beads 1.0x

1. Let the AMPure XP beads come to room temperature for at least 30 minutes.
2. Wipe down Agilent Bravo deck with 70% Ethanol.
3. Place an empty tip box on position 1.
4. Place a full tip box on position 2.
5. Aliquot **50µl of homogenous Agencourt AMPure XP** beads in a 96 well Eppendorf twin.tec plate, then place plate at deck position 5.
6. Place an empty Eppendorf twin.tec plate on deck position 9 for final sample collection.
7. Aliquot **35µl of Illumina Resuspension Buffer** in a 96 well Eppendorf twin.tec plate, then place plate at deck position 9 on top of the sample collection plate.
8. Aliquot **180µl of 80% ethanol** in two 96 well Eppendorf twin.tec plates, then place plates at deck position 9 on top of the buffer plate.
9. Place Dynal MPC-92S magnet at deck position 7.
10. Place the Eppendorf twin.tec plate containing amplified DNA at deck position 9 on top of the ethanol plates.



11. Start Bravo automated "5. Post PCR Bead Purification stacked Row.pro" protocol and follow the instructions.
12. Enter parameters for the purification step as follows:
 - a. **DNA to bead ratio** **1.0**
 - b. **DNA_vol** **52**
 - c. **elute_vol** **32.5**
 - d. **samples** **12 (or multiple of 12)**

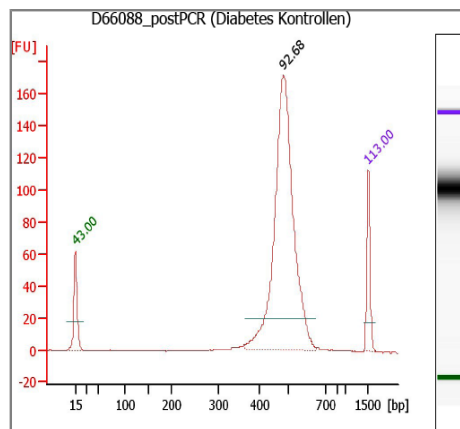
-
13. Once the protocol has finished assess quality and quantity with the Agilent 2100 Bioanalyzer.

SAFE STOPPING POINT (-20°C)



Step 14. Assess quality and quantity with the Agilent 2100 Bioanalyzer (DNA 1000 Kit)

1. Let the Bioanalyzer DNA 1000 Kit come to room temperature at least 30 minutes.
2. Check that the 2100 Bioanalyzer electrodes have been cleaned as instructed in the reagent kit guide.
3. Turn on the 2100 Bioanalyzer, open the Agilent 2100 Expert Software and check communication.
4. Within the instrument context, choose the DNA 1000 assay from the drop down list.
5. Enter sample names and comments in the Data and Assay context.
6. Mix the samples carefully.
7. Prepare the chip, samples and ladder as instructed in the reagent kit guide.
8. Load the prepared chip into the 2100 Bioanalyzer and start the run within five minutes after preparation.
9. Set the peak from the beginning of the curve to the end. Verify the results. Check that the electropherogram shows a distribution with a peak height between 350 to 700 nucleotides (average 430bp).



-
10. Import ng/ μ l, nM and bp of each sample into the sample tracking sheet generated by the *Solexa* database.
 11. Dilute the samples to 10 nM stocks if sample concentration is above 15 nM and transfer them in 2D barcoded tubes and label them.
 12. Once the quality measurement has finished proceed with the PicoGreen Quantification and import the final sample tracking sheet into the database.

5.3 *Lab Tracking*-Formulare für die automatisierten Exom-
Library-Präparation auf der *Bravo Liquid Handling*
Station

Exome Lab Tracking Form

DNA batch description: _____

Date: _____

PLEASE MAKE A NOTE OF THE USED THERMOCYCLER AT EACH INCUBATION STEP

1 Measurement of DNA stocks

Date: _____

Operator: _____

Quality and Quantity control: Start: _____ Stop: _____

- NanoDrop cleaned with sodium hypochlorite
- measurement of 1µl DNA stock per sample
- verification of DNA quality and quantity
(high molecular DNA with a minimum of 60ng/µl with a A260/280 ratio between 1.8 and 2.0 is required)

2 Fragment DNA

Date: _____

Operator: _____

Fragment DNA: Start: _____ Stop: _____

- covaris instrument at 7°C and degased at least for 30min
- dilution of 3µg DNA with low TE to a final volume of 50µl
- sheared with Broad settings (Broad SureSelect)
- sheared DNA transferred into an Eppendorf twin.tec plate

3 Perform automated AMPure XP Purification

Date: _____

Operator: _____

Purify DNA: Start: _____ Stop: _____

- beads 30min at RT, homogenous
- Bravo instrument set up and cleaned
- twin.tec plate containing 150µl AMPure beads at position 5
- empty twin.tec plate on position 9 for purified DNA
- twin.tec plate containing 55µl water per sample on position 9 on top of the empty plate
- two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
- DynaMag magnetic plate on position 7
- empty tip box on position 1
- box with fresh tips on position 2
- plate containing sheared DNA on top of the stack on position 9
- Bravo protocol 1.SPRI bead purification protocol stacked started

4 Quality Check via Bioanalyzer (DNA 1000 Kit)

Date: _____

Operator: _____

Bioanalyzer Start: _____ Stop: _____

- DNA 1000 Kit 30min at RT
- 1µl sample per well
- verification of quality and quantity of sheared DNA
- peak from the beginning of the curve to 500bp
- average peak height 135bp
- average concentration 45ng/µl (20ng/µl minimum)
- result tables exported and sample tracking sheet for BA post shearing completed

5 Perform End-Repair

Date: _____
Operator: _____
 End-Repair Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 master mix for end-repair prepared
 51µl master mix added to 49µl sheared DNA sample
 incubated 30min at 20°C

6 Perform automated AMPure XP Purification

Date: _____
Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 Nunc 1.2ml deep well plate containing 220µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 37µl water per sample on position 9 on top of the empty plate
 two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing end-repaired DNA on top of the stack on position 9
 Bravo protocol 2.SPRI bead purification protocol deep well started

7 Add 'A' bases to the 3' end

Date: _____
Operator: _____
 Add 'A' base Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 master mix for end-repair prepared
 20µl master mix added to 30µl sheared DNA sample
 incubated 30min at 37°C

8 Perform automated AMPure XP Purification

Date: _____
Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 110µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 35µl water per sample on position 9 on top of the empty plate
 two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing 'A' base added DNA on top of the stack on position 9
 Bravo protocol 1.SPRI bead purification protocol stacked started

9 Ligate indexing-specific adapters

- Date: _____
Operator: _____
 Ligate Adapters Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 master mix for end-repair prepared
 21,5µl master mix added to 30µl sheared DNA sample
 incubated 15min at 20°C

10 Perform automated AMPure XP Purification

- Date: _____
Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 35µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 37µl water per sample on position 9 on top of the empty plate
 two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing adapter ligated DNA on top of the stack on position 9
 Bravo protocol 1.SPRI bead purification protocol stacked started

11 Amplify adapter-ligated library

- Date: _____
Operator: _____
 Amplify DNA Fragments Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 master mix for end-repair prepared
 20µl master mix added to 30µl sheared DNA sample
 PCR program started

12 Perform automated AMPure XP Purification

- Date: _____
Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 90µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 35µl water per sample on position 9 on top of the empty plate
 two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing amplified DNA on top of the stack on position 9
 Bravo protocol 1.SPRI bead purification protocol stacked started

13 Quality Check via Bioanalyzer

- Date: _____
Operator: _____
 Bioanalyzer Start: _____ Stop: _____
 DNA 1000 Kit 30min at RT
 1µl sample per well
 verification of quality and quantity of sheared DNA
 peak from the beginning of the curve to 700bp
 average peak height 245bp
 average concentration 60ng/µl (18ng/µl minimum)
 result tables exported and sample tracking sheet for BA preHyb completed

14 Hybridize the library

- Date: _____
Operator: _____
 Concentrate the Library: Start: _____ Stop: _____
 500ng library concentrated via speed vac centrifuge
 library pellet in 3,4µl water diluted to 147ng/µl

 Prepare Master Mixes: Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 master mix for hybridization buffer prepared
 master mix for capture library prepared
 master mix for blocking buffer prepared

 Set up Hybridization: Start: _____ Stop: _____
 5.6µl blocking mix in an Eppendorf twin.tec plate ("library plate")
 3,4µl diluted DNA library added to blocking mix
 cycler 1: "library plate" 5min at 95°C with a heated lid at 105°C
 cycler 1: after 95°C step cycler temperature at 65°C
 cycler 2: cycler temperature at 65°C
 40µl hybridization buffer in a new Eppendorf twin.tec plate ("buffer plate")
 cycler 2: "buffer plate" 5min at 65°C with a heated lid at 105°C
 7µl capture library in the empty wells of the "library plate"
 "library plate" 2min at 65°C with a heated lid at 105°C
 13µl hybridization buffer into the capture libraries contained in the "library plate"
 10µl DNA library into the capture libraries contained in the "library plate"
 hybridization mixture mixed by pipetting up and down 10 times
 48h at 65°C with a heated lid at 105°C

15 Perform automated Hybrid Capture

- Date: _____
Operator: _____
- Select Hybrid Capture: Start: _____ Stop: _____
 Dynal MyOne Streptavidin T1 magnetic beads washed
 200µl washed Streptavidin beads in a Nunc 1,2ml deep well plate
- Bravo instrument set up and cleaned
 Bravo positions 4 and 6 to 65°C
 510µl SureSelect wash buffer into a Nunc 1,2ml deep well plate
 plate containing wash buffer 1 on position 8
 1900µl SureSelect wash buffer 2 into a Nunc 2,0ml deep well plate
 plate containing wash buffer 2 on position 6
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 hybridization mix transferred into the Streptavidin beads containing plate
 Bravo protocol 3. PostHyb DNA wash with wash buffer 1 and 2 started
 replacement of used tip boxes during automated capture (time after run start)
1: _____ 2: _____ 3: _____

16 DNA Elution

- Date: _____
Operator: _____
- DNA Elution: Start: _____ Stop: _____
 plates containing wash buffer 1 and wash buffer 2 removed
 waste container removed
 twin.tec plate containing 55µl water per sample on position 8
 twin.tec plate containing 50µl SureSelect Neutralization buffer on deck position 9
 empty tip box on position 1
 box with fresh tips on position 2
 Bravo protocol 4. PostHyb DNA elution started
 replacement of used tip boxes during automated capture (time after run start)
1: _____

17 Perform automated AMPure XP Purification

- Date: _____
Operator: _____
- Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 Nunc 1.2ml deep well plate containing 180µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 35µl water per sample on position 9 on top of the empty plate
 two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing amplified and enriched DNA on top of the stack on position 9
 Bravo protocol 2.SPRI bead purification protocol deep well started

18 Amplify enrichend exome library and add index-tags

- Date: _____
Operator: _____
 Amplify Enriched Library Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 master mix for end-repair prepared
 1µl specific index primer added
 35µl master mix added to 15µl sheared DNA sample
 PCR program started

19 Perform automated AMPure XP Purification

- Date: _____
Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 90µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 35µl water per sample on position 9 on top of the empty plate
 two twin.tec plates containing 110µl 70% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing amplified DNA on top of the stack on position 9
 Bravo protocol 1.SPRI bead purification protocol stacked started

20 Quality Check via Bioanalyzer (DNA HS Kit)

- Date: _____
Operator: _____
 Bioanalyzer Start: _____ Stop: _____
 DNA HS Kit 30min at RT
 1µl sample per well
 verification of quality and quantity of sheared DNA
 peak from the beginning of the curve to the end of the curve
 average peak height 300bp
 average molarity of 10nM (minimum 5nM)
 result tables exported and sample tracking sheet for BA postHyb completed
 dilution to 10nM necessary? _____

21 Library Quantification with picoGreen

- Date: _____
Operator: _____
 picoGreen measurement Start: _____ Stop: _____
 pico green thawed, mixed and centrifuged
 lambda DNA diluted (1:1000)
 8 points serial 1:2 dilution of lambda DNA
 DNA libraries added to 100µl TE buffer
 picoGreen dilution 1:200
 100µl picoGreen added to standard and library
 final xls. Sheet saved and printed
 results exported and sample tracking sheet completed

5.4 *Lab Tracking*-Formulare für die automatisierten Genom-
Library-Präparation auf der *Bravo Liquid Handling*
Station

Whole Genome Lab Tracking Form

DNA batch description: _____

Date: _____

1 Measurement of DNA stocks

Date: _____
Operator: _____
Quality and Quantity control: _____ Start: _____ Stop: _____
 NanoDrop cleaned with sodium hypochlorite
 measurement of 1 µl DNA stock per sample
 verification of DNA quality and quantity
(high molecular DNA with an A260/280 ratio between 1.8 and 2.0 is required)

2 Fragment DNA

Date: _____
Operator: _____
 Fragment DNA: _____ Start: _____ Stop: _____
 Covaris instrument at 7°C and degased at least for 30min
 dilution of 1 µg DNA with low TE to a final volume of 120 µl
 shearing protocol "350bp"
 sheared DNA transferred into an Eppendorf twin.tec plate

3 Perform automated AMPure XP Purification

Date: _____
Operator: _____
 Purify DNA: _____ Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 90 µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 53 µl Illumina Resuspension buffer per sample on position 9 on top of the empty plate
 two twin.tec plates containing 180 µl 80% EtOH per sample on position 9 on top of the buffer plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing sheared DNA on top of the stack on position 9
 Bravo protocol 2.SPRI bead purification protocol deep well started

4 Quality Check via Bioanalyzer (DNA High Sensitivity Kit)

Date: _____
Operator: _____
 Bioanalyzer _____ Start: _____ Stop: _____
 DNA High Sensitivity Kit 30min at RT
 1 µl sample per well
 verification of quality and quantity of sheared DNA
 peak from the beginning of the curve to the end
 average peak height 380bp
 result tables exported and sample tracking sheet for BA post shearing completed

5 Perform End-Repair

- Date: _____
Operator: _____
 End-Repair Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 40µl End-Repair Mix added
 10µl Illumina Resuspension Buffer added
 incubated 30min at 30°C

6 Perform automated AMPure XP Purification

- Date: _____
Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 160µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 25µl Illumina Resuspension buffer per sample on position 9 on top of the empty plate
 two twin.tec plates containing 180µl 80% EtOH per sample on position 9 on top of the buffer
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing end-repaired DNA on top of the stack on position 9
 Bravo protocol 2.SPRI bead purification protocol deep well started

7 Add 'A' bases to the 3' end

- Date: _____
Operator: _____
 Add 'A' base Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 12.5µl A-Tailing Mix added
 incubated 30min at 37°C

8 Ligate indexing-specific adapters

- Date: _____
Operator: _____
 Ligate Adapters Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 master mix for end-repair prepared
 21,5µl master mix added to 30µl sheared DNA sample
 incubated 10min at 30°C
 5µl Stop Ligase Mix added

9 Perform automated AMPure XP Purification

Date: _____

Operator: _____

- Purify DNA:
- beads 30min at RT, homogenous
- Bravo instrument set up and cleaned
- twin.tec plate containing 42.5µl AMPure beads at position 5
- empty twin.tec plate on position 9 for purified DNA
- twin.tec plate containing 55µl Illumina Resuspension buffer per sample on position 9 on top of the empty plate
- two twin.tec plates containing 180µl 80% EtOH per sample on position 9 on top of the buffer
- DynaMag magnetic plate on position 7
- empty tip box on position 1
- box with fresh tips on position 2
- plate containing adapter-ligated DNA on top of the stack on position 9
- Bravo protocol 1.SPRI bead purification protocol stacked started

10 Perform automated AMPure XP Purification

Date: _____

Operator: _____

- Purify DNA:
- beads 30min at RT, homogenous
- Bravo instrument set up and cleaned
- twin.tec plate containing 50µl AMPure beads at position 5
- empty twin.tec plate on position 9 for purified DNA
- twin.tec plate containing 35µl Illumina Resuspension buffer per sample on position 9 on top of the empty plate
- two twin.tec plates containing 180µl 80% EtOH per sample on position 9 on top of the buffer
- DynaMag magnetic plate on position 7
- empty tip box on position 1
- box with fresh tips on position 2
- plate containing purified DNA on top of the stack on position 9
- Bravo protocol 1.SPRI bead purification protocol stacked started

11 Size Selection of Adapter Ligated Fragments

Date: _____

Operator: _____

- Size Selection:
- 1.5% Size Select Pippin Gel
- fill up all buffer wells
- volume of elution wells replaced with 40ul fresh buffer
- test run performed
- 40ul buffer removed from the sample wells
- 40ul DNA-Marker added to one sample well
- 10ul loading solution mixed with 30ul of the library
- 40ul library mixed with loading solution added to the other lanes
- start run 500bp tight cut
- elution wells sealed with adhesive film
- remove size selected library

12 Perform automated AMPure XP Purification

Date: _____

Operator: _____

- Purify DNA:
- beads 30min at RT, homogenous
- Bravo instrument set up and cleaned
- NUNC deep well plate 1.0ml plate containing 70µl AMPure beads at position 5
- empty twin.tec plate on position 9 for purified DNA
- twin.tec plate containing 25µl Illumina Resuspension buffer per sample on position 9 on top of the empty plate
- two twin.tec plates containing 180µl 80% EtOH per sample on position 9 on top of the buffer
- DynaMag magnetic plate on position 7
- empty tip box on position 1
- box with fresh tips on position 2
- plate containing purified DNA on top of the stack on position 9
- Bravo protocol post pippin prep purification protocol stacked DW started

13 Amplify adapter-ligated library

Date: _____
Operator: _____
 Amplification Start: _____ Stop: _____
 thawed reagents mixed and briefly centrifugated
 5µl PCR Primer Cocktail added
 25µl PCR master mix added
 PCR program started

14 Perform automated AMPure XP Purification

Date: _____
Operator: _____
 Purify DNA: Start: _____ Stop: _____
 beads 30min at RT, homogenous
 Bravo instrument set up and cleaned
 twin.tec plate containing 50µl AMPure beads at position 5
 empty twin.tec plate on position 9 for purified DNA
 twin.tec plate containing 35µl Illumina Resuspension buffer per sample on position 9 on top of the empty plate
 two twin.tec plates containing 180µl 80% EtOH per sample on position 9 on top of the water plate
 DynaMag magnetic plate on position 7
 empty tip box on position 1
 box with fresh tips on position 2
 plate containing amplified DNA on top of the stack on position 9
 Bravo protocol 1.SPRI bead purification protocol stacked started

15 Quality Check via Bioanalyzer

Date: _____
Operator: _____
 Bioanalyzer Start: _____ Stop: _____
 DNA 1000 Kit 30min at RT
 1µl sample per well
 verification of quality and quantity of sheared DNA
 peak from the beginning of the curve to the end
 average peak height 430bp
 result tables exported and sample tracking sheet for BA preHyb completed
 dilute samples to 10nM stocks

16 Library Quantification with picoGreen Post Dilution

Date: _____
Operator: _____
 picoGreen measurement Start: _____ Stop: _____
 pico green thawed, mixed and centrifuged
 lambda DNA diluted (1:1000)
 8 points serial 1:2 dilution of lambda DNA
 DNA libraries added to 100µl TE buffer
 picoGreen dilution 1:200
 100µl picoGreen added to standard and library
 final xls. Sheet saved and printed
 results exported and sample tracking sheet completed

5.5 Primer

5.5.1 Primer für *VPS35*

Tabelle 19: Intronsche Primer für die Amplifikation der kodierenden Sequenz von *VPS35*.

Exon	Vorwärtsprimer	Rückwärtsprimer	Amplikon-Größe in bp
1	AGGCCAATCAATGAGCCG	AATCGGGCTGGTCTTAATCC	270
2	CTGGACATTGTTGGAGAAGG	TGAAGATTAGATTTACCAGAAACACC	301
3	GCAAGGTGAACCGATCTAAG	CAAACAAAATAGGCCTTATCAAGAA	332
4	CGGCTAAAATTCATCCTTCG	ACAGACACAAAAGTTCTCTGATTTG	308
5	AAATGAGATATTTTCCTTGTCATGC	GGCTGACTGGGTGGAAGAAT	440
6	AAGTAGGAAATACACAAAGGAGAAGG	AACACTGCATATATATCTTTCCGTTTA	453
7	AGTTAGTAGTTCAAATGCGTTACTAGA	TCACCACACCCCGCAAG	305
8	CAAAAGTAATGTGTTTGAATTTAGTTG	TTAAGCCTCTTTCATGTCTAACCT	433
9	TTGAGTGGAGAGTGCGTGAC	GGCAAGATTAATCAGAGAGACTTG	467
10	CATGCCTTCAGAGGATGTTG	CAAATGCCAGCAGTTCAAAA	232
11	CCACAGAAGACTATAACAACCTCC	AGATGGGAGGGGAGGGTAAT	451
12	GCCAATGGTTTTACCTCAGT	GCAATTCCTTCAAACCATCT	316
13	CTGTCTCTGGGTTGTTGTGAC	TAGAGAGTTGGCGAAAATGG	447
14	AAATGGATATCCTGGAACAAG	TGCAAACAAGCAACTGAACC	309
15	TGCTCAACTAGAGGATGGTTG	ACACAAGGCCATGACAACCTG	366
16	TGTTGTGTTTCTTTGTATGAAAACG	CAAATCTCCTAAGAGTAGGAAGGG	281
17	GGAATTGTATCCCCATCTTTC	TATGGAAGGGAAACCTAGCG	400

Danksagung

Zunächst möchte ich mich bei Herrn Prof. Dr. Thomas Meitinger und Herrn PD Dr. Tim-Matthias Strom für die Möglichkeit bedanken, meine Doktorarbeit am Institut für Humangenetik des Helmholtz Zentrum München durchführen zu können. Besonderer Dank gilt dabei Herrn Prof. Dr. Thomas Meitinger für seine Unterstützung bei dieser Arbeit vor allem in der Finalisierungsphase durch konstruktive Kritik und anregende Diskussionen.

Weiterhin danke ich Herrn Prof. Dr. Fries, Leiter des Lehrstuhls für Tierzucht des WZW Weihenstephan, für die Übernahme der Vertretung meiner Doktorarbeit gegenüber der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München.

Ich bedanke mich bei Frau Prof. Dr. Angelika Schnieke, Leiterin des Lehrstuhls für Biotechnologie der Nutztiere, für die Übernahme des Vorsitzes der Prüfungskommission.

Weiterhin möchte ich den Kollegen aus meiner Arbeitsgruppe, die mich über die vergangenen Jahre hinweg unterstützt haben, danken. Allen voran Sandy Lösecke, Jasmin Gundlach, Carola Fischer, Thomas Wieland und Thomas Schwarzmayr. Ein besonderer Dank geht an Barbara Schormair für ihre Motivation und Unterstützung in der finalen Phase meiner Arbeit.

Zum Schluss möchte ich mich von ganzem Herzen bei meinem Ehemann Daniel für seine Unterstützung und seinen Glauben an mich bedanken. Danke, das du mir zuhörst und mit mir und unseren drei wundervollen Kindern Max, Niels und Lina Charlotte die Herausforderung eines gemeinsamen Lebens angenommen hast. Ihr seid mein größtes Glück.