

Masterarbeit

Implementierung einer Probenidentifikations-Pipeline zur
Vermeidung von Probenverwechslungen bei NGS-Analysen
mittels SNP-ID Analyse: ein Methodenvergleich

Eingereicht von

Isabel Menrath

zur Erlangung des akademischen Grades

Master of Science (MSc)

an der

Medizinischen Universität Graz

ausgeführt am

Diagnostik & Forschungs- (D&F) Institut für Humangenetik

unter der Anleitung von Betreuer

ao.Univ.-Prof. Mag. DDr. Erwin Petek

Unterhaching, 03.08.2023

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Unterhaching, am 03.08.2023

Isabel Menrath

Inhalt

Zusammenfassung.....	1
Abstract	2
Abkürzungsverzeichnis.....	3
Abbildungsverzeichnis.....	4
Tabellenverzeichnis.....	4
1. Einleitung.....	6
1.1. Mögliche Zeitpunkte der Probenverwechslung	6
1.1.1. Blutabnahme	6
1.1.2. DNA-Extraktion.....	7
1.1.3. NGS-Analyse	8
1.2. Bisherige Möglichkeiten zur Identifikation von Probenverwechslung	9
1.2.1. Probenidentifikation mittels STR-Analyse.....	11
1.3. Probenidentifikation mittels SNP-ID Analyse.....	12
2. Methoden.....	15
2.1. Patient*innenproben	15
2.2. Sanger-Sequenzierung	16
2.2.1. SNP ID-Analyse mit Sanger-Sequenzierung.....	17
2.3. NGS - Next-Generation-Sequencing.....	21
2.3.1. Library Preparation.....	22
2.3.2. Klonale Amplifikation	23
2.3.3. SBS - Sequenzierung	24
2.3.4. NGS-Applikationen	25
2.3.5. SNP-ID Analyse mittels NGS	27
3. Ergebnisse	32
3.1. Sanger-Sequenzierung	32
3.2. SureDirect Blood PCR Kit.....	38

3.3.	Twist Human Sample ID Kit	49
3.3.1.	Beispiel einer Probenverwechslung/Kontamination	53
3.3.2.	Beispiel eines unklaren Ergebnisses	55
4.	Diskussion.....	57
4.1.	SureDirect Blood PCR Kit	60
4.2.	TWIST Human Sample ID Kit	62
5.	Fazit	65
6.	Anhang	67
7.	Literatur.....	84

Zusammenfassung

In den letzten Jahren wurden neue Methoden in der Routinediagnostik der Molekulargenetik etabliert, die es erlauben, große Anzahlen an Proben auf einmal zu sequenzieren und zu analysieren. Der Vorteil dieser neuen Methoden der Hochdurchsatzsequenzierung besteht in den geringeren Kosten und der schnelleren Abarbeitung. Zudem erlaubt die Qualität der Daten eine umfassendere Auswertung. Mit dem höheren Probenumfang der Analysen steigt jedoch das Risiko für Probenverwechslungen oder -kontaminationen, vor allem, wenn die Analyse nicht oder nur teilautomatisiert abläuft. Die bisher in unserem Labor durchgeführten Methoden, um eine solche Verwechslung oder Kontamination zu identifizieren, entsprechen nicht mehr den neusten Standards. Im Rahmen dieser Masterarbeit soll deshalb eine robuste und sichere Methode etabliert werden, die gut mit den Routine-Abläufen der molekulargenetischen Diagnostik vereinbar und kostensparend ist, und eine möglichst einfache Auswertung ermöglicht. Diese Methode soll den Patient*innen und auch dem Labor eine hohe Sicherheit und einen hohen Qualitätsstandard bieten.

Es wird eine Single Nucleotid Polymorphism (SNP)-ID-Analyse aus EDTA-Blut implementiert, durch die Probenverwechslungen oder -kontaminationen im Rahmen von NGS-Analysen ab der DNA-Isolierung sicher identifiziert werden können. Bei den SNP handelt es sich um hochpolymorphe Varianten, die es erlauben, Individuen eindeutig voneinander zu unterscheiden, wodurch eine sehr zuverlässige Identifikation der Probe gewährleistet werden kann. Diese SNP werden parallel und idealerweise aus der Primärprobe (EDTA-Blut) und der daraus isolierten DNA bestimmt. Durch den Vergleich der SNP-Profile aus Blut- und zugehöriger DNA-Probe kann eine Probenverwechslung nahezu sicher ausgeschlossen werden (mit Ausnahme von eineiigen Zwillingen).

In der vorliegenden Arbeit werden die möglichen Zeitpunkte der Probenverwechslung beschrieben und bisher verwendete Methoden zur Überprüfung einer Probenverwechslung erläutert. Die SNP-ID-Analyse wird mittels Sanger-Sequenzierung und dem SureDirect Blood PCR Kit (Agilent) und einer anschließenden Library Preparation durchgeführt. In Zusammenarbeit mit der Firma

Twist Bioscience werden außerdem zur Etablierung des Produkts 'Twist Human Sample ID Kit' sowohl das α - als auch das β -Kit auf deren Durchführbarkeit getestet.

Abstract

With the introduction of NGS analysis into routine diagnostics in molecular human genetics, it is possible to sequence and analyze large numbers of samples simultaneously.

The advantages besides the high throughput are lower costs and faster processing, but also the quality of the data allows a more profound analyses. However, the higher sample turnover can also lead to an increased risk for sample mix-ups or contamination, especially if the analysis is not or only partially automated.

The methods used so far in the laboratory to identify potential mix-ups or contamination no longer meet the newest standards. In the context of this master thesis, therefore, a new, robust method is to be established, which is well compatible with the routine diagnostics and cost-effective. In addition, a simple evaluation is desired. This method should offer the patient and the laboratory a high level of safety and a high-quality standard.

A single nucleotide polymorphism (SNP) ID analysis from EDTA blood will be implemented, by which sample mix-ups or contaminations can be reliably identified in the context of NGS analyses from the point of DNA isolation onwards. SNPs are highly polymorphic variants that allow individuals to be uniquely identified, making samples distinguishable from one another. These SNPs are determined in parallel from the primary sample (EDTA blood) and the isolated DNA. By comparing the SNP profiles from the blood sample and the corresponding DNA sample, sample mix-ups can be ruled out with high certainty.

In this thesis, the possible time points of sample mix-up are described, and previously used methods to detect sample mix-ups are illustrated. SNP-ID analysis is performed using Sanger sequencing and the Agilent SureDirect Blood PCR Kit (NGS). Furthermore, in collaboration with Twist Bioscience, both the α - and β -version of the kit 'Twist Human Sample ID Kit' will be tested.

Abkürzungsverzeichnis

AF	Allelfrequenz
altAF	Alternative Allelfrequenz
AZF	Azoospermiefaktor
bp	Basenpaar
dNTP	Desoxyribonukleosidtriphosphat
ddNTP	Didesoxyribonukleosidtriphosphat
DNA	Desoxyribonukleinsäure
EDTA	Ethylendiamintetraessigsäure
FISH	Fluoreszenz-InSitu-Hybridisierung
GfH	Gesellschaft für Humangenetik
HPO	Human Phenotype Ontology
ID	Identity
MAF	Minor allele frequency
Mb	Megabasen
mL	Milliliter
µL	Mikroliter
ng	Nanogramm
NGS	Next Generation Sequencing
nM	Nanomolar
PCR	Polymerasekettenreaktion
pM	Picomolar
RBS	Resuspension Buffer
Ref	Referenzbase
rpm	revolutions per minute
SBS	Sequencing by synthesis
SHOT	Serious Hazards of Transfusion
SNP	Single Nucleotide Polymorphism
STR	Short tandem repeats

WES	Whole exome sequencing
WGS	Whole genome sequencing

Abbildungsverzeichnis

Abbildung 1: Workflow einer NGS-Analyse ab der Blutabnahme bis zur Auswertung der Daten.....	9
Abbildung 2: Statement 20 der S1-Leitlinie:	12
Abbildung 3: Die drei Schritte der Sanger-Sequenzierung	16
Abbildung 4: Überblick der Library Preparation mittels Ligation	23
Abbildung 5: Ablauf der einzelnen Schritte der Bridge-PCR	24
Abbildung 6: Ablauf der Sequenzierung mittels SBS.....	25
Abbildung 7: Mittelwerte der SNP ID Tiefe Lauf 1, logarithmische Darstellung	40
Abbildung 8: Mittelwerte der SNP ID Tiefe Lauf 2, logarithmische Darstellung	42
Abbildung 9: Mittelwerte SNP ID Tiefe Lauf 3, logarithmische Darstellung	44
Abbildung 10: Mittelwerte SNP ID Tiefe Lauf 4, logarithmische Darstellung	46
Abbildung 11: Mittelwerte SNP ID Tiefe Lauf 5 – 8, logarithmische Darstellung ..	48
Abbildung 12: Mittelwerte der SNP ID Tiefe Lauf 9, logarithmische Darstellung ..	51
Abbildung 13: Mittelwerte der SNP ID Tiefe Lauf 10, logarithmische Darstellung	52
Abbildung 14: Rohdaten IGV-Viewer Probe 13	56

Tabellenverzeichnis

Tabelle 1: SNP nach Pengelly et.al mit Primer-Sequenzen, Basenpaargröße und MAF.....	18
Tabelle 2: Primerverdünnung für SureDirect Blood PCR Kit	28
Tabelle 3: SNP im TWIST Custom Panel ID2020 und Exom	30

Tabelle 4: Beispiel Probe 1 SureDirect Blood PCR Kit (Lauf 1).....	38
Tabelle 5: Beispiel Probe 2 SureDirect Blood PCR Kit (Lauf 2).....	41
Tabelle 6: Beispiel Probe 3 SureDirect Blood PCR Kit (Lauf 3).....	42
Tabelle 7: Beispiel Probe 4 SureDirect Blood PCR Kit (Lauf 4).....	44
Tabelle 8: Beispiel Probe 9 TWIST Human Sample ID Kit	49
Tabelle 9: Beispiel Probe 10 TWIST Human Sample ID Kit	51
Tabelle 10: Verdünnung der SNP ID-Library je nach Probenanzahl	53
Tabelle 11: Übersicht Kriterien Probenverwechslung/Kontamination	55

1. Einleitung

Durch den Einzug von Next-Generation-Sequencing-Methoden in die molekulargenetische Routinediagnostik der Humangenetik konnten viele Erschwernisse, welche die bis dato üblichen Methoden mit sich brachten, verringert werden. Neben der Verbesserung der Genauigkeit der Daten, der schnelleren Abarbeitung der Proben und der Senkung der Kosten ist vor allem der hohe Durchsatz ein Vorteil der neuen Technologien (siehe auch Methodenteil). Bei der Hochdurchsatzsequenzierung im Rahmen von NGS werden die Proben gepoolt, sodass eine höhere Anzahl an Proben auf einmal bearbeitet und analysiert werden kann.

Mit dem höheren Probendurchsatz steigt allerdings auch das Risiko einer Probenverwechslung zu Beginn und während der Analyse, vor allem bei semiautomatischen Workflows wie sie im SYNLAB MVZ Humane Genetik in München durchgeführt werden. Da die Library Preparation nicht automatisiert stattfindet, können theoretisch jederzeit Probenverwechslungen und -kontaminationen auftreten. Dadurch besteht die Gefahr, dass Patient*innen falsche Analyseergebnisse erhalten.

1.1. Mögliche Zeitpunkte der Probenverwechslung

1.1.1. Blutabnahme

Eine Probenverwechslung kann zu unterschiedlichen Zeitpunkten stattfinden. Der erste mögliche Zeitpunkt stellt dabei die Blutabnahme dar. In der Standard-Arbeitsanleitung zur peripher venösen Blutentnahme für die labormedizinische Diagnostik ist deshalb genau vorgegeben, wie eine korrekte Blutabnahme erfolgen sollte. Hier wird angegeben, dass in Anwesenheit der Patient*innen vor oder direkt nach der Blutabnahme die Probengefäße mit dem richtigen Probenetikett, passend zum Auftrag, beklebt werden müssen. Die Identität der Patient*innen sollte durch Erfragen des Namens bestätigt werden. Zusätzlich kann auch das Geburtsdatum erfragt werden, um die korrekte Zuordnung zu erleichtern (von Meyer et al. 2017). Trotz dieser Maßnahmen kann es dazu kommen, dass Probenröhrchen gar nicht

oder falsch beschriftet werden und an das Analyselabor versendet werden. Das UK SHOT definiert das „wrong blood in tube“ so, dass entweder das Blut des*der falschen Patient*in in das richtig beschrifteten Röhrchen gelangt oder das Blut des*der richtigen Patient*in in ein falsch beschriftetes Röhrchen. In einer Studie werden als mögliche Gründe für die Verwechslung Unterbrechung während des Vorgangs, Ablenkung oder Überarbeitung des Personals angegeben. Da viele Fehler nicht erkannt werden, weil keine zusätzlichen Maßnahmen zur Überprüfung der korrekten Zuordnung der Proben stattfinden, liegen hier keine genauen Daten vor. Der Prozentsatz von falschen Etikettierungen wird auf ca. 0,1 % geschätzt (Huijsmans et al. 2007), die Dunkelziffer liegt aber wahrscheinlich deutlich höher (Bolton-Maggs et al. 2015).

Probenröhrchen, die nicht beschriftet sind und nicht eindeutig einem Auftrag zugeordnet werden können, sollen vom Analyselabor nicht angenommen werden. In diesem Fall wird der Einsender informiert und neues Material angefordert, das nicht verwendbare Material wird vernichtet. Ein falsch beschriftetes Probenröhrchen kann vom Analyselabor als solches nicht identifiziert werden. Eine Probenverwechslung, die so stattgefunden hat, kann demnach in der Analyse nicht erkannt werden, sodass Patient*innen ein falsches Ergebnis und damit einen falschen Befund erhalten.

1.1.2.DNA-Extraktion

Ein weiterer Zeitpunkt zur Möglichkeit einer Probenverwechslung ist die DNA-Extraktion. Im SYNLAB MVZ Humane Genetik München werden verschiedene Methoden zur DNA-Extraktion verwendet, darunter vollautomatische, halbautomatische und manuelle. Bei vollautomatischen Methoden werden sowohl die barcodierten Blutröhrchen als auch die barcodierten DNA-Röhrchen durch einen Scanner erkannt und dementsprechend zugeordnet. Dadurch ist eine Probenverwechslung nahezu ausgeschlossen, da das Gerät bei Nicht-Übereinstimmung der Barcodes einen Fehler anzeigt und die Extraktion nicht startet. Bei semiautomatischen und manuellen Methoden werden die Barcodes nicht gescannt, sodass eine Möglichkeit der Probenverwechslung besteht. Dieses

Risiko kann minimiert werden, indem ein Vieraugenprinzip während der Überführung des Blutes in die Extraktionsgefäße angewendet wird. Dadurch besteht auch hier ein sehr geringes Risiko einer Probenverwechslung, restlos ausgeschlossen werden kann sie aber nicht.

1.1.3.NGS-Analyse

Da im SYNLAB MVZ Humane Genetik München die Library Preparation manuell geschieht, kann es an vielen unterschiedlichen Zeitpunkten während der Analyse zu Probenverwechslungen und -kontaminationen kommen. Als erstes kann bereits zu Beginn die falsche DNA in die Analyse eingesetzt werden. Dadurch wird Patient*innen die falsche DNA zugeordnet, was weitreichende Folgen haben kann. Diese Möglichkeit der Probenverwechslung wird damit umgangen, dass die Normalisierung der Proben zum Erreichen einer bestimmten, für die Analyse notwendigen DNA-Konzentration hauptsächlich automatisch passiert. Dabei werden die Barcodes der DNA-Röhrchen gescannt und mit der Liste der zu bearbeitenden Proben verglichen. Auch hier wird ein Fehler angezeigt, wenn Unstimmigkeiten zwischen Barcode und Liste vorliegen.

Die größte Gefahr einer Verwechslung oder Kontamination besteht aber während der Library Preparation, da diese, wie bereits erwähnt, manuell abläuft. Während der komplexen und zeitintensiven Durchführung der Analyse vieler Proben kann es vorkommen, dass unbewusst Pipettierfehler auftreten oder kleine DNA-Mengen in andere Kavitäten gelangen. So können Proben vertauscht oder DNA durch DNA anderer Proben kontaminiert werden.

In Abbildung 1 wird der routinemäßige Workflow ab der Blutabnahme bis zur NGS-Analyse dargestellt. Die roten Pfeile markieren mögliche Zeitpunkte, an denen eine Probenverwechslung stattfinden kann.

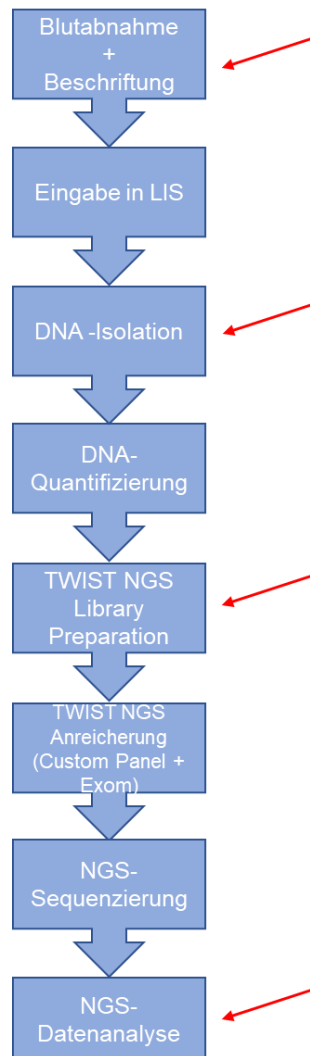


Abbildung 1: Workflow einer NGS-Analyse ab der Blutabnahme bis zur Auswertung der Daten.

1.2. Bisherige Möglichkeiten zur Identifikation von Probenverwechslung

Um eine mögliche Probenverwechslung auszuschließen, wurden bereits einige Verfahren eingesetzt. Ein Verfahren dabei ist der Geschlechtsabgleich. Die NGS-Daten, die im SYNLAB MVZ Humane Genetik München entstehen, werden mit der Plattform varvis® der Firma Limbus Medical Technologies GmbH ausgewertet. Für alle Patient*innen berechnet varvis® anhand verschiedener Parameter ein Geschlecht, welches mit dem eingegebenen Patient*innen-Geschlecht verglichen wird. Falls eine Nicht-Übereinstimmung der Geschlechter vorliegt, wird dies angezeigt und es werden weitere Schritte der Überprüfung eingeleitet. Der

Geschlechtsabgleich hat den Nachteil, dass nur bei Vorliegen unterschiedlicher Geschlechter eine Probenverwechslung aufgedeckt werden kann. Liegt die Verwechslung beispielsweise bei zwei weiblichen Proben vor, kann diese nicht erkannt werden. Weitere Probleme treten auf, wenn Patient*innen im Vorfeld eine Knochenmarkstransplantation oder eine Blut-Transfusion des anderen Geschlechts erhalten haben. Hier würde der Abgleich eine Auffälligkeit anzeigen, obwohl keine Probenverwechslung vorliegt. Gonosomale Veränderungen wie beispielsweise das Klinefelter-Syndrom werden ebenfalls nicht korrekt erkannt, sodass hier der alleinige Geschlechtsabgleich nicht ausreicht.

Ein weiteres Tool der varvis® -Plattform ist der Mix-up Check. Varvis® vergleicht bei der Auswertung der Daten jede Probe mit jeder Probe eines NGS-Laufs und berechnet auf Grundlage mehrerer, nicht genauer definierten Parameter einen Similarity-score. Dieser Score gibt an, wie ähnlich sich zwei Proben sind. Der Score reicht von 0,001 bis 0,1, wobei die Werte in Abhängigkeit vom verwendeten Enrichment variieren. Im hier verwendeten Custom-Panel TWIST Custom Panel ID2020 liegen beispielsweise bei einem Wert von $< 0,03$ identische Proben vor. Die Beziehung zwischen Eltern und Kind bei einer Trio-Exom-Analyse spiegelt sich in einem Wert von 0,07 – 0,09 wider. Der Score bei zwei unverwandten Proben liegt bei $> 0,1$. Je niedriger der Wert ist, desto ähnlicher sind sich also zwei Proben, wodurch eine Probenverwechslung oder Kontamination mit Fremd-DNA ausgeschlossen werden kann. Zu beachten ist, dass es sich hierbei lediglich um einen berechneten Score handelt, und dass oftmals keine Informationen zu Verwandtschaftskonstellationen vorliegen.

Diese Methoden sind gute Möglichkeiten zur Identifikation einer Probenverwechslung, allerdings sind weitere Methoden nötig, um diese zu überprüfen, da sie nicht eindeutig genug sind und somit eine Probenverwechslung nicht zu 100 % ausschließen können. Lag bisher der Verdacht einer Probenverwechslung durch fehlende Übereinstimmung zwischen den Geschlechtern vor, wurde eine AZF-Locusanalyse durchgeführt. Der Locus des Azoospermiefaktors liegt auf dem Y-Chromosom, sodass mit dieser schnellen und einfachen Analyse herausgefunden werden kann, ob es sich bei der Probe um eine männliche oder weibliche handelt (Vogt et al. 2008). So kann der Verdacht einer

Probenverwechslung aufgrund des Geschlechts bestätigt oder ausgeschlossen werden. Um Gonosomenstörungen (Turner-Syndrom, Klinefelter-Syndrom) zu analysieren, wurde eine FISH-Analyse mit X- und Y-chromosomalen Sonden angeschlossen. Falls nach diesen zusätzlichen Analysen der Verdacht einer Probenverwechslung nicht sicher ausgeschlossen werden konnte, mussten die Proben teilweise erneut analysiert werden, um eine sichere Diagnostik zu gewährleisten. Dadurch, dass die weiterführenden Analysen hauptsächlich aus der extrahierten DNA stattfanden, konnte eine Probenverwechslung vor der Extraktion nicht ausgeschlossen werden.

1.2.1. Probenidentifikation mittels STR-Analyse

Lag der Verdacht vor, dass es sich bei zwei Proben um identische Proben handelt, oder Proben, die in einem Verwandtschaftsverhältnis stehen, kann als zusätzliche Analyse eine STR-Analyse durchgeführt werden. Short tandem repeats, kurz STR, werden auch Mikrosatelliten genannt und kommen in der menschlichen DNA sehr häufig vor. Es handelt sich um DNA-Abschnitte, die aus vielen Wiederholungseinheiten von zwei bis sieben Nukleotiden bestehen, welche sich tandemartig wiederholen und insgesamt ca. 3 % des menschlichen Genoms ausmachen. Die STR liegen in den nicht-codierenden Regionen zwischen den Genen und sind hochvariabel (Butler 2007). Durch ihre Variabilität können sie vielseitig eingesetzt werden; in der forensischen DNA-Forschung ist die STR-Analyse zur Identifikation von Proben der Goldstandard. Bei der Analyse wird mittels Multiplex-PCR eine bestimmte Anzahl hochvariabler STR-Loci, meist zwischen 16 und 21, bestimmt. Dabei entstehen PCR-Produkte mit einer Länge zwischen 100 und 500 bp, die in der Kapillarelektrophorese aufgetrennt werden. Die Anzahl der STR-Wiederholungen variiert von Mensch zu Mensch innerhalb einer Population und dient damit als sehr gute Unterscheidungsmöglichkeit zwischen Individuen. Je mehr Loci dabei analysiert werden, desto höher ist die Unterscheidungskraft (Tytgat et al. 2021). In der Routinediagnostik wird die Analyse der STR-Loci zum Beispiel bei Vaterschaftstests angewendet. Hierfür sind kommerzielle Kits auf dem Markt erhältlich. Limitiert ist die Analyse, da STR eine hohe Mutationsrate aufweisen,

große Amplifikationsprodukte nötig sind und Hochdurchsatztechnologien fehlen (Du et al. 2017). Deshalb und wegen der fehlenden Vergleichbarkeit zwischen STR-Markern und NGS-Daten kann die STR-Analyse nicht routinemäßig als Möglichkeit zur Probenidentifikation eingesetzt werden.

Die bisher durchgeführten Methoden zur Probenidentifikation sind nicht ausreichend, um eine 100 % sichere Diagnostik für Patient*innen zu gewährleisten. Auch entsprechen sie nicht vollständig der S1 Leitlinie der GfH (deutsche Gesellschaft für Humangenetik e.V.). In dieser Leitlinie geht es darum, „die Anforderungen für eine Hochdurchsatz-basierte Diagnostik von angeborenen Veränderungen des Erbguts (Keimbahn) zu definieren“ (Bauer et al. 2018). Im Statement 20 wird empfohlen, eine Möglichkeit zur Identifikation von Probenverwechslungen mittels Barcodes zu etablieren, „die nicht für die Prozessierung eingesetzt“ werden (Bauer et al. 2018)(siehe Abbildung 2). Es war also nötig, ein neues Verfahren zur Detektion von Probenverwechslungen einzuführen, um Patient*innen eine sichere Diagnostik und einen hohen Qualitätsstandard zu gewährleisten.

Statement 20: Aspekte der Probenverfolgung und die Nutzung von molekularen Barcodes sollten während der Test-Validierung und Plattform-Validierung mitbetrachtet werden.

Hierbei ist insbesondere von Bedeutung, ob nicht-zugeordnete Sequenzinformationen im Sinne einer Kontamination (beispielsweise mit nicht-humaner DNA) oder eine mögliche Probenverwechslung (Detektion von Barcodes, die nicht für die Prozessierung eingesetzt waren) beobachtet wurden.

Abbildung 2: Statement 20 der S1-Leitlinie:

Molekulargenetische Diagnostik mit Hochdurchsatz-Verfahren der Keimbahn, beispielsweise mit Next-Generation Sequencing (Bauer et al. 2018)

1.3. Probenidentifikation mittels SNP-ID Analyse

Eine weitere Möglichkeit, Proben voneinander zu unterscheiden, ist die Analyse von Single Nucleotide Polymorphisms, kurz SNP. Dabei handelt es sich um einzelne Nukleotid-Substitutionen, die nach Definition häufiger als 1 % in der Bevölkerung vorkommen. Diese SNP zeichnen sich durch eine geringe Mutationsrate, eine

schnelle Genotypisierung und eine hohe Abundanz im Genom aus. Ein weiterer Vorteil gegenüber der STR-Analyse ist, dass SNP mit Hochdurchsatz-Sequenzieretechnologien nachgewiesen werden können. Durch die hohe Variabilität der SNP eignen sich diese zur genauen Identifizierung von Proben. Es gab bereits viele Studien zur Entwicklung eines SNP-Panels, welche eine individuelle Identifizierung von Proben und damit eine hohe Diskriminierung zwischen Proben erlaubt. Du et al. (2017) beispielsweise entwickelten eine Methode zur Identifizierung des Genoms durch Analyse einer kleineren Anzahl an SNP, indem sie die fünf häufigsten Populationen verglichen, die im 1000 Genome Project analysiert wurden. Sie sammelten 74 biallelische autosomale SNP, die sich in acht weit verbreiteten Genotypisierungs-Assays überschneiden. Zusätzlich fügten sie vier SNP, die zur ABO-Blutgruppen-Klassifikation verwendet werden und zwei SNP zur Geschlechtsbestimmung hinzu. Sie geben an, dass mit diesen 80 SNP jede einzelne Person der Weltbevölkerung durch ihre genetische Daten identifiziert werden kann. Kim et al. (2010) analysierten Genotypisierungsergebnisse von 2000 koreanischen Personen und fanden 56 hochgradig polymorphe SNP-Marker, welche die Anforderung erfüllten. 30 identifizierte Marker wurden zur weiteren Validierung ausgewählt. Zusätzlich beschrieben sie, dass sich die Unterscheidungskraft der SNP um das Zehnfache erhöht, wenn fünf zusätzliche SNP-Marker hinzugefügt werden. Auch wurde gezeigt, dass hochpolymorphe und geschlechtsspezifische SNP-Marker Individuen genauso gut diskriminieren wie STR-Marker, sofern genug SNP verwendet werden. Sie behaupten, dass die Identifizierung von Koreanern durch 40 SNP-Marker mit $MAF > 0,45$ genauso eindeutig ist wie mit 16 STR-Markern.

Alle bereits durchgeführte Studien wählten die SNP-Marker nach folgenden Kriterien aus:

- SNP mit hohem Heterozygotiegrad ($> 0,45$)
- Geringe Unterschiede in Allelfrequenzen in allen Populationen
- $MAF > 0,45$ in allen ethnischen Gruppen
- Biallelische Substitutionen ohne Transversionen
- Keine pathogene oder wahrscheinlich pathogene Varianten

Die meisten SNP-Panel, die bisher in der Routine vor allem in der Forensik verwendet werden, umfassen SNP, die in nicht-exonischen Bereichen liegen. Diese können bei NGS-Analysen im Rahmen des Whole-Exome-Sequencing (WES) nicht verwendet werden, da diese Regionen nicht angereichert werden. Pengelly et al. (2013) verglichen die drei häufigsten Exom-Anreicherungskits (Agilent SureSelect Human All Exon V4, Illumina TruSeq Exome Enrichment und Nimblegen SeqCap EZ Human Exome Library V3.0-Kits) und filterten die gemeinsamen SNP heraus. Diese sollten selbst in großen Datensätzen dazu führen, dass Proben eindeutig voneinander diskriminiert werden können.

Die SNP wurden nach den bereits genannten Kriterien ausgewählt. Zusätzlich wurden weitere Filter verwendet:

- SNP nicht in Repeat- oder Homopolymerregionen
- GC-Gehalt der angrenzenden 250 bp zwischen 40 – 55 %
- Keine weitere Variante im angrenzenden 50 bp Bereich mit AF > 0,01
- AF zwischen 0,2 und 0,8
- SNP darf kodiertes Protein nicht verändern oder OMIM-Datensatz aufweisen
- SNP mehr als 10 bp von Exongrenze entfernt
- Nicht in Regionen mit hoher Sequenzähnlichkeit zu anderen Regionen
- Außerhalb des Kopplungsungleichgewichts mit allen anderen SNP

Zur Erstellung des SNP-Panels wurden die Daten aus der 1. Phase des 1000 Genome Projects verwendet. Insgesamt konnten 1662 übereinstimmende SNP in den drei Anreicherungskits gefunden werden, von denen nach der Filterung 117 übrigblieben. Das optimierte Panel umfasst 24 SNP mit einer AF von ungefähr 0,5. Die Forschungsgruppe zeigte, dass 24 SNP 48 Allelvergleiche ermöglichen und durchschnittlich 18 unterschiedliche Allele zwischen zwei beliebigen Proben nötig sind, um diese klar voneinander zu unterscheiden. Umgekehrt bedeutet dies, dass bei weniger Allelen oder keiner Übereinstimmung eine Probenverwechslung oder -kontamination vorliegen kann.

Es konnte auch gezeigt werden, dass die Datensatzgröße eine Auswirkung auf die Möglichkeit von gleichen oder ähnlichen SNP-Profilen hat. Die

Unterscheidungskraft des SNP-Profiles durch die SNP von Pengelly et al. reicht bis zu einer Probenanzahl von ca. 10.000 aus, sodass die Proben noch eindeutig voneinander unterschieden werden können. Werden mehr Proben analysiert oder soll die Sicherheit weiter erhöht werden, können aus dem SNP-Pool weitere SNP hinzugefügt werden. Durch die Wahl der SNP können auch verschiedene Bevölkerungsgruppen analysiert werden. Limitiert ist das Verfahren bei Verwandtschaftsverhältnissen, vor allem bei Konsanguinität innerhalb der Familie. Durch die Integration weiterer SNP können diese Fälle gelöst werden.

Das entstandene SNP-Profil kann nach der Analyse mit einem unabhängigen durch WES-Analyse produzierten Genotyp-Profil verglichen werden. Wichtig für den Vergleich in der Routinediagnostik ist, dass die SNP in das Anreicherungspanel integriert sind.

2. Methoden

2.1. Patient*innenproben

Bei den EDTA-Proben, die für die Erstellung dieser Masterarbeit verwendet werden, handelt es sich um Blutproben von Patient*innen, bei denen nach Aufklärung und unterschriebener Einverständniserklärung gemäß Gendiagnostik-Gesetz (GenDG) eine NGS-Analyse angefordert und durchgeführt wurde. Da es sich bei der Analyse um qualitätssichernde Maßnahmen handelt, die der Sicherheit der Patient*innen dienen, ist kein Ethikvotum nötig. Daten von Patient*innen, die in der Masterarbeit veröffentlicht werden, sind vollständig anonymisiert, sodass keine Rückführbarkeit auf ursprüngliche Patient*innen möglich ist.

Die DNA-Extraktion aus EDTA-Blut erfolgt automatisiert mit dem Extraktionsautomaten chemagic Star (Typ Microlab Star) der Firma Hamilton Bonaduz AG, manuell mit dem FlexiGene DNA Kit (Qiagen) oder automatisch mit dem Maxwell® CSC Genomic DNA Kit auf dem Maxwell® CSC 48 IVD (Promega).

2.2. Sanger-Sequenzierung

Die Sanger-Sequenzierung galt lange als Goldstandard bei klinischen Tests zum Nachweis von krankheits-verursachenden genetischen Varianten und Bestimmung der genauen Basensequenz eines DNA-Bereichs. Bei dieser Sequenzier-Methode, auch Kettenabbruch-Methode genannt, wird als Ausgangsmaterial gereinigtes PCR-Produkt eingesetzt. Es werden veränderte Didesoxyribonukleotide (ddNTP) verwendet, welchen eine 3'OH-Gruppe fehlt, die für die Verlängerung des DNA-Strangs nötig ist. Die DNA-Polymerase kann so keine Phosphodiesterbindung mit dem 5'Phosphat des nächsten Nukleotid bilden, was dazu führt, dass der Strang nicht weiter verlängert werden kann. Dadurch, dass in der Sequenzierreaktion dNTP und zu einem deutlich geringeren Anteil fluoreszierende ddNTP in die Reaktion gegeben werden, kommt es dazu, dass Stränge jeder möglichen Länge entstehen, da die Nukleotide zufällig eingebaut werden (Heather and Chain 2016). Wird also zufällig ein ddNTP eingebaut, führt es zum Abbruch der Reaktion. Die unterschiedlichen langen DNA-Stränge werden anschließend aufgereinigt, um nicht-gebundene Nukleotide zu entfernen, und mittels Kapillarelektrophorese aufgetrennt. Durch die unterschiedliche Fluoreszenzmarkierung der vier ddNTP ist es möglich, die genaue Basenabfolge der DNA-Sequenz zu erhalten (Slatko et al. 2018). Ein Schema der Durchführung einer Sanger-Sequenzierung ist in Abbildung 3 dargestellt.

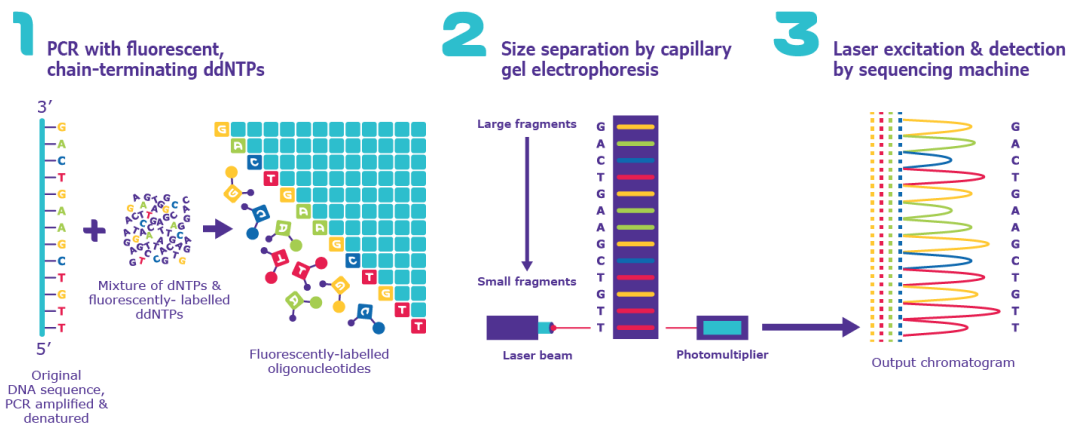


Abbildung 3: Die drei Schritte der Sanger-Sequenzierung

(Sanger-Sequenzierung- Schritte und Methode (n.d) https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/marketing/global/images/technical-documents/protocols/genomics/sequencing/sanger-sequencing_steps_process_diagram/sanger-sequencing_steps_process_diagram.png, 16.07.2023 13:16)

Die Sanger-Sequenzierung ist in ihrem Durchsatz limitiert, da die gängigen Sequencer zur Kapillarelektrophorese lediglich 96 Kapillare besitzen. Außerdem ist die Fragmentgröße auf 500 bis 800 bp beschränkt, sodass größere Bereiche mit mehreren überlappenden Teilstücken analysiert werden müssen. Des Weiteren ist die Abarbeitung insgesamt teurer und aufwändiger als die Sequenzierung mittels NGS, vor allem wenn mehrere Gene auf einmal sequenziert werden sollen.

2.2.1. SNP ID-Analyse mit Sanger-Sequenzierung

Die SNP der Patient*innen können auf verschiedene Weisen nachgewiesen werden. Bei der Analyse mittels Sanger-Sequenzierung werden die Primer für die SNP, welche im verwendeten TWIST Custom Panel ID2020 integriert sind, in die Reaktion eingesetzt. Die SNP mit den jeweiligen Primern sind in Tabelle 1 beschrieben.

Tabelle 1: SNP nach Pengelly et.al mit Primer-Sequenzen, Basenpaargröße und MAF

Chromosom	Gen	rs-Nummer	HGVS	Primer forward	Primer reverse	bp	MAF
2	<i>ABCB11</i>	rs497692	c.3084A	AGTCCCTTAAACACCACCCTCTCC	GGTTGTCGGTCCAGCAGTTGAA	290	0,544797
2	<i>COL4A4</i>	rs10203363	c.3594G	AACTTATGGCTGACCAGTGTGATAGG	CACGAGGACCTGGAGGAGAGATTC	296	0,472765
3	<i>SUMF1</i>	rs2819561	c.*61T	CTCTACACAGTCTTATTGTTACAGGTATCG	TTGGTCTCACAAAGGCGGTTCC	276	0,590071
5	<i>VCAN</i>	rs309557	c.5808T	CATAGTTCTAATGGCAAGTCTGTGTCA	TGGGCATCCTCCTCCTCCTCTA	206	0,504382
7	<i>ABCA13</i>	rs17548783	c.12111T	TTTGTCTCCTGGTGCTGAT	TCCTTACCTGCCTCGTGAGT	206	0,480745
8	<i>PDP1</i>	rs4735258	c.*36T	GTAGTTCAGTTCAATTCTCATGTTGTAGGG	TGAGCTAGGCAGGACCAGACTC	248	0,470757
9	<i>TDR7</i>	rs1381532	c.33A	TAGGCAAAGATGCTGGAAGG	ACCACTGCTGGCACACTTCT	188	0,51109
10	<i>HPSE2</i>	rs10883099	c.*17C	TAGTGATGGTGGACGACGGGAC	CTGGAGTGGAGGAGTGGAAAGCA	202	0,506933
12	<i>WNK1</i>	rs7300444	c.3960C	TAGACGTGCCCAAATGACAG	CAGTGGGCACTGTAACCTCA	209	0,420008
13	<i>FREM2</i>	rs9532292	c.7398A	GGTGTGACCAGCCCTATGAG	TGATCTTCAGGGGAAAATGG	218	0,373783
15	<i>SLC12A6</i>	rs4577050	c.3003C	TGCTTTTTGCCCTTTTATG	GCCCAGTATCATCCCTTTCA	212	0,653517
16	<i>AARS</i>	rs2070203	c.903C	TCTGCTAATGCTCCTCAATCCTCCA	ACCCACGCCCTGTGTTGTCA	245	0,470352
17	<i>COG1</i>	rs1037256	c.1782G	ATCACTGCCCAAGGACGTTT	CACTGTTGAGGGCATCCTGT	194	0,521692
18	<i>LAMA3</i>	rs9962023	c.2901T	TTCTGGGCTCTATTTTCACCA	CGGAGCTCTTCACATTCACA	206	0,658658
19	<i>DNMT1</i>	rs2228611	c.1389A	TGTGCTTTGTTTCCTGTCTAGTGTGTA	GCTATGTTGTCCAGGCTCGTCTC	287	0,518105
20	<i>FERMT1</i>	rs10373	c.114T	ACCATCCCAATGAAGAGCAG	GCTCTCCAGGGCATTACAAG	209	0,517533
21	<i>NDUFV3</i>	rs4148973	c.468T	TTCACAGATGAAGGGGTTCC	CTGAAACGTCAGCCTCATCA	196	0,558026
22	<i>SERPIND1</i>	rs4675	c.1446T	GCACCAAGGCACGATCACAGT	CCAAGGCACTTCAGACACCTAGAC	219	0,46092

Es werden drei Multiplex-Ansätze vorbereitet. Die ausgewählten SNP mit flankierenden Bereichen werden mittels Multiplex-PCR amplifiziert. Die dafür benötigten drei Ansätze sind wie folgt aufgeteilt:

Multiplex 1: SNP der Gene *ABCB11*, *COL4A4*, *SUMF1*, *VCAN*, *ABCA13*, *PDP1*

Multiplex 2: SNP der Gene *TDR7*, *HPSE2*, *WNK1*, *FREM2*, *SLC12A6*, *AARS*

Multiplex 3: SNP der Gene *COG1*, *LAMA3*, *DNMT1*, *FERMT1*, *NDUFV3*, *SERPIND1*.

Zusammenstellung der Primermixe:

	Multiplex 1 - 3
Primermenge je Primer [μL]	1,5
Tris-HCL pH 8 [μL]	32
Total [μL]	50

Der PCR-Ansatz wird mit dem QIAGEN Multiplex PCR *Plus* Kit durchgeführt.

Der PCR- Mastermix wird auf Eis vorbereitet:

Reagenz	Volumen pro Reaktion [μL]
2x Multiplex PCR Master Mix	12,5
dH ₂ O	9,0
Primermix	2,5
Total	24

Je nach Probenanzahl werden die Volumina des Mastermix angepasst.

24 μL des Mastermixes werden in 0,2 mL- PCR-Tubes vorgelegt. Als Template werden 1 μL DNA pro Probe eingesetzt.

PCR-Programm für die Multiplex-PCR:

Schritt	Temperatur [°C]	Zeit	Zyklenanzahl
1	95	5 min	
2	95	30 sek	28x
3	59	90 sek	
4	72	30 sek	
5	68	30 min	
6	20	Hold	

Nach der PCR erfolgt eine Aufreinigung, um störende Komponenten wie nicht eingebaute Primer oder Nukleotide zu entfernen. Die Aufreinigung wird mittels Agencourt AMPure XP (Beckman Coulter) am Biomek NX^P durchgeführt, die Elution erfolgt in 40 µL dH₂O.

Für die Sequenzierreaktion werden die Primer auf 1,25 µM mit Tris-HCL pH 8 verdünnt. Der Ansatz erfolgt mit dem BigDye™ Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems), welches alle benötigten Komponenten für die Reaktion enthält.

Der Master-Mix wird auf Eis angesetzt:

Reagenz	Volumen pro Reaktion [µL]	
	forward	reverse
BigDye™ Terminator v1.1 Ready Reaction Mix	0,3	0,3
BigDye™ Terminator v1.1 5X Sequencing Buffer	2,0	2,0
dH ₂ O	4,2	4,2
Primer forward (1,25 µM)	2	-
Primer reverse (1,25 µM)	-	2
aufgereinigtes PCR-Produkt	1	1

Je nach Probenanzahl werden die Volumina des Mastermix angepasst.

PCR-Programm für die Sequenzierreaktion:

Schritt	Temperatur [°C]	Zeit	Zyklenanzahl
1	95	60 sek	
2	95	10 sek	25x
3	55	10 sek	
4	60	2 min 30 sek	
5	10	Hold	

Vor der Kapillarelektrophorese findet erneut eine Aufreinigung statt, um nicht eingebaute ddNTP und dNTP zu entfernen. Aufgereinigt wird mit dem CleanSEQ Dye-Terminator Removal Kit (Beckman Coulter) am Biomek NXP, eluiert wird in 75 µL 0,1 mM EDTA pH 8,0.

Die Kapillarelektrophorese erfolgt mit dem 3730xl DNA Analyzer der Firma Applied Biosystems. Die Auswertung findet mit der Software SEQUENCE Pilot Version 5.1.0. (JSI Medical systems) statt.

2.3. NGS - Next-Generation-Sequencing

Im Laufe der Zeit fand eine Weiterentwicklung der DNA-Sequenzierungstechnologien statt, die unter dem Begriff „Next Generation Sequencing“, kurz NGS, zusammengefasst werden. Alle Technologien haben gemeinsam, dass klonal amplifizierte Moleküle oder Einzelmoleküle als Vorlage genutzt werden, welche massiv parallel sequenziert werden. Durch die parallele Sequenzierung mehrerer Proben auf einmal ist der Durchsatz im Vergleich zur Sanger-Sequenzierung deutlich höher. Kosten können so eingespart werden und die Dauer der Analyse wird reduziert. Des Weiteren können mehrere Gene oder das ganze Exom oder Genom gleichzeitig sequenziert werden. Aber auch mit der Technologie des NGS können nur „short reads“ von bis zu 800 bp generiert werden, sodass die entstandenen Daten über größere DNA-Abschnitte zusammengesetzt werden müssen (Hu et al. 2021).

2.3.1. Library Preparation

Ausgangsmaterial für die Analyse ist genomische DNA, die aus EDTA-Vollblut extrahiert wurde. In der Library Preparation werden die DNA-Moleküle für die anschließende Sequenzierung vorbereitet. Diese Preparation ist abhängig von der Sequenzierplattform. Da es sich, wie bereits erwähnt, beim NGS um short-read-Technologien handelt, muss die Patient*innen-DNA in möglichst einheitlich große Fragmente geschnitten werden. Dies geschieht mit Hilfe einer Fragmentierung, die physikalisch, chemisch oder enzymatisch durchgeführt werden kann. Damit die Enden der Fragmente für die anschließende Adapterligation präpariert sind, werden diese repariert, indem ein Poly-A-Schwanz angehängt wird. Die Universaladapter, die eine einzelne überhängende Thyminbase zur Bindung besitzen, enthalten Sequenzen, die spezifisch für die Sequenzierplattform sind. Die Sequenz der Adapter ist komplementär zu kurzen Oligonukleotiden, welche auf der Flowcell befestigt sind, sodass sie eine Immobilisierung der Fragmente auf der Flowcell ermöglichen. In der Library Preparation werden zusätzlich kurze Indexsequenzen an jedes einzelne Fragment einer jeden Probe hinzugefügt. Diese Indexsequenzen sind für jede Probe individuell, was die Identifizierung jeder Probe ermöglicht. So können nach der Sequenzierung die erhaltenen Sequenzdaten nicht nur einer Probe eindeutig zugeordnet werden, sondern es wird auch das Poolen mehrere Proben auf einmal ermöglicht. Eine anschließende Größenselektion, die mithilfe von Beads oder elektrophoretisch stattfinden kann, reichert die DNA-Fragmente in der gewünschten Größe an und entfernt zusätzlich Verunreinigungen (Hu et al. 2021). In Abbildung 4 ist ein Schema der Durchführung einer Library Preparation mittels Ligation dargestellt.

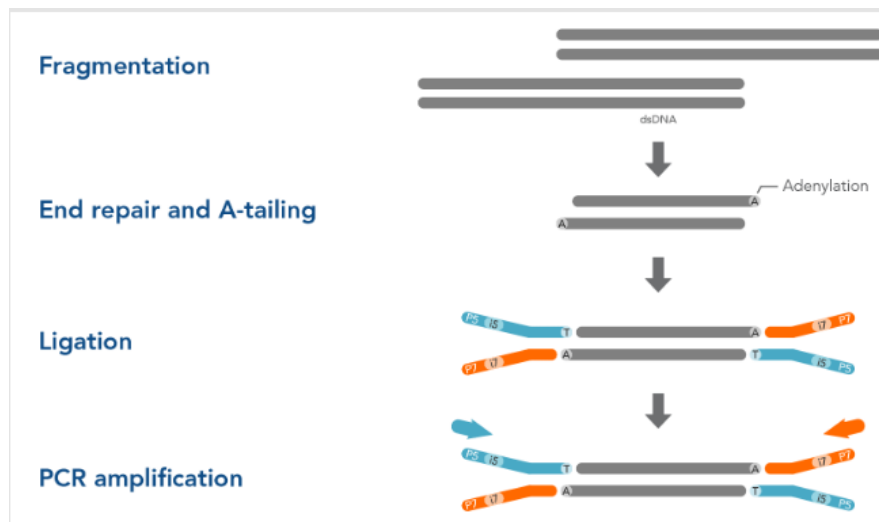


Abbildung 4: Überblick der Library Preparation mittels Ligation

(Ligation-based library preparation (n.d), <https://eu.idtdna.com/pages/technology/next-generation-sequencing/library-preparation/ligation-based-library-prep>, 15.07.2023 13:58)

2.3.2. Klonale Amplifikation

Da Short-read-Sequenzierer keine Einzelmoleküle sequenzieren können, ist eine klonale Amplifikation Voraussetzung für die Sequenzierung. Diese erfolgt bei den Geräten der Firma Illumina mittels Bridge-PCR. Die in der Library Preparation entstandenen Fragmente werden denaturiert, sodass sie einzelsträngig vorliegen. Anschließend werden sie über die Flowcell gespült, auf der die komplementären Oligonukleotide gebunden sind, an welche sie hybridisieren und so immobilisiert werden. Diese Forward- und Reverse- Oligonukleotide werden zu Beginn der Bridge-PCR an die Flowcell gebunden und dienen als Primer, mithilfe dessen von jedem individuellen Fragment eine identische Kopie erstellt wird. Die einst eingesetzten Moleküle werden anschließend entfernt und ausgehend vom in der Library Preparation erzeugten Molekül werden neue Moleküle generiert. Durch einen Zyklus aus Denaturierungs-, Annealing- und Elongationsschritten bei 60 °C können die 3'-Enden des kopierten Template-Moleküls an das komplementäre Oligonukleotid auf der Flowcell binden, indem es eine Brücke bildet. Am Ende entsteht für jedes gebundene Fragment eine Kopie auf der Flowcell, wodurch Cluster von bis zu 1000 Kopien eines Ausgangsmoleküls generiert werden

(Buermans and den Dunnen 2014). Abbildung 5 zeigt den Ablauf der Bridge-PCR im Schema.

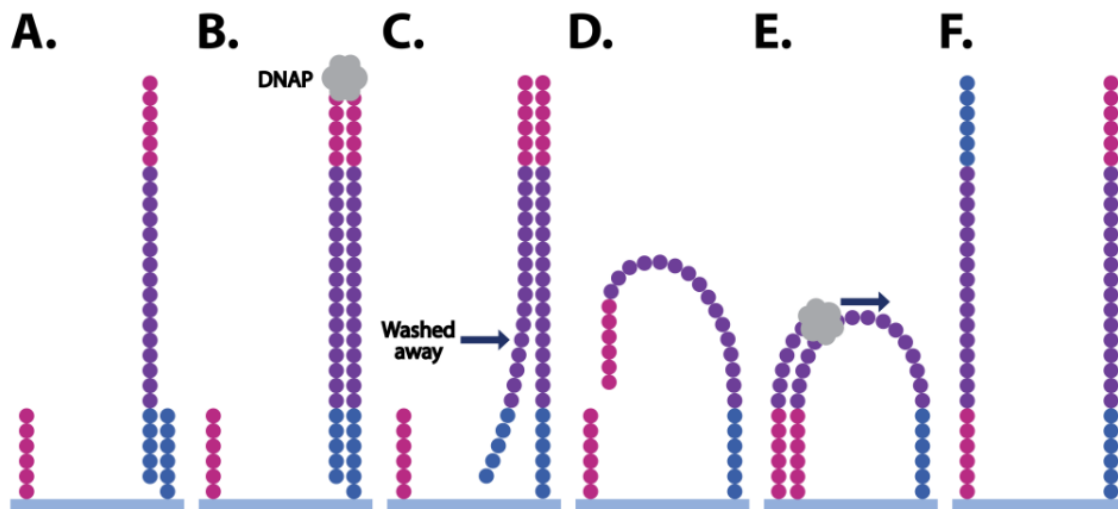


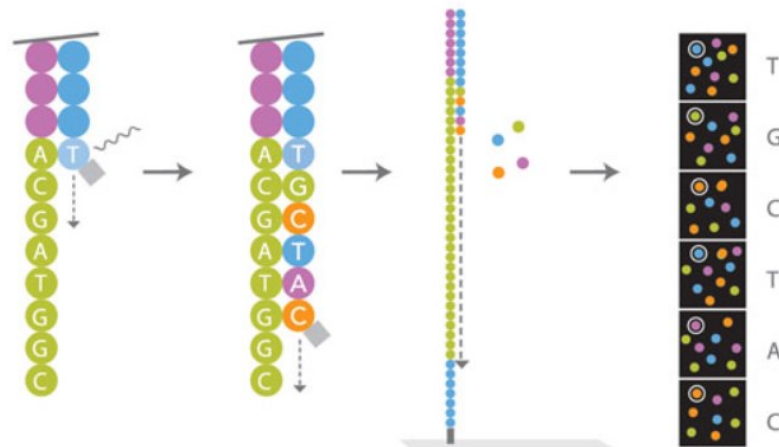
Abbildung 5: Ablauf der einzelnen Schritte der Bridge-PCR

- (A) ssDNA-Molekül hybridisiert an auf der Flowcell gebundene Oligonukleotidprimer
- (B) eine Kopie des Templates wird gebildet und (C) der Originalstrang wird entfernt. (D) – (F) Bildung einer Brückenstruktur und Clusterbildung durch isothermale Amplifikation des Ausgangsmoleküls (Nguyen, J. (2022). Bridge Amplification Sequencing. Apollo Institute. <https://apollo-institute.org/bridge-amplification-sequencing>, 14.07.2023 18:43)

2.3.3. SBS - Sequenzierung

Illuminas Technologie der Sequenzierung beruht auf dem Prinzip des „Sequencing by synthesis“ (SBS) mit reversiblen fluoreszenzmarkierten Terminatoren. Die DNA-Polymerase baut in jedem Zyklus ein fluoreszenzmarkiertes dNTP ein, an welches ein reversibler Terminator gebunden ist. Der Einbau des passenden Nukleotids kann optisch in real-time anhand des Fluoreszenzsignals gemessen werden (Zhong et al. 2021). Nach dem Messen werden der Fluoreszenzfarbstoff und der Terminator, der verhindert, dass innerhalb eines Zyklus ein weiteres Nukleotid eingebaut wird, abgespalten, sodass das nächste Nukleotid in die wachsende Kette eingebaut werden kann. Dieser Vorgang wird so lange wiederholt, bis das Fragment vollständig sequenziert ist (siehe auch Abbildung 6). Die Firma Illumina nutzt je nach Sequenzierplattform zwei (z.B. NextSeq) oder vier (z.B. MiSeq) Farben für die

Markierung der Nukleotide. Des Weiteren können single-read oder paired-end Sequenzierungen durchgeführt werden. Bei der paired-end Sequenzierung erfolgt die Erfassung der Basensequenz eines Fragments am forward und am reversen Strang, wodurch mehr Reads und Daten mit höherer Abdeckung erreicht werden (Hu et al. 2021). Die Sequenzierung bei single read findet einzelsträngig statt. Die Nachteile der SBS-Methodik sind, dass kürzere Reads als bei der Sanger-Methode entstehen (300 - 500 bp) und eine höhere Fehlerquote vorliegt (Slatko et al. 2018).



©2013 Illumina, Inc. All rights reserved.

Abbildung 6: Ablauf der Sequenzierung mittels SBS

Einbau eines fluoreszenzmarkierten Nucleotids mit Terminator, Detektion des Lichtsignals mit anschließender Entfernung des Terminatorsignals. Darstellung der Sequenzierung eines Fragments mit optischer Detektion des Lichtsignals. (Next generation sequencing. (n.d.). <https://www.medicinische-genetik.de/diagnostik/allgemeine-informationen/methoden/next-generation-sequencing>, 15.07.2023 19:34)

2.3.4. NGS-Applikationen

Die Applikationen des NGS sind sehr vielfältig. In der Routinediagnostik werden vor allem drei Hauptanalyseebenen durchgeführt: Gen-Panels, Whole-Exome-Sequencing (WES) und Whole-Genome-Sequencing (WGS).

Gen-Panels zeichnen sich dadurch aus, dass nur bestimmte, ausgewählte Gene analysiert werden. Sie enthalten dadurch nur eine Teilmenge von Genen, die mit einer bestimmten Krankheit in Verbindung stehen. Da die Zahl der Gene pro Panel

von einigen wenigen bis zu mehreren 100 reichen kann, schwankt die diagnostische Ausbeute.

Beim Whole-Exome-Sequencing werden alle ca. 22000 proteincodierende Gene des Menschen sequenziert. Neben den proteincodierenden Genen werden auch die regulatorischen Regionen und andere funktionelle Bereiche angereichert (Nisar et al. 2021). Dies entspricht nur etwa 1-2 % des Genoms, enthält allerdings rund 85 % aller bekannten potenziell pathogenen Varianten (Rehm et al. 2013). Die WES-Analyse findet Einsatz bei komplexen Phänotypen mit mehreren Differenzialdiagnosen oder bei Phänotypen, die nicht mit einem bereits bekannten Krankheitsgen assoziiert sind. Hierzu bieten sich Trio-Analysen aus Eltern und Kind an, wo beispielsweise *de novo* Varianten leichter gefunden werden können (Zhong et al. 2021). Die hohe Anzahl an zu analysierenden Genen erschwert die Auswertung und Interpretation der Daten, sodass auch beim Exom eine Panel-Auswertung erfolgen kann. Hierbei werden durch geeignete HPO-Terms nur die Gene herausgefiltert, die zum Phänotyp passen. HPO steht für Human Phenotype Ontology, und die HPO-Datenbank assoziiert Krankheiten mit dem Auftreten bestimmter Phänotypen. Durch die Auswahl an den Phänotyp angepassten HPO-Terms soll die Diagnoserate verbessert und die Auswertung erleichtert werden (Zhao et al. 2020). Werden so keine potenziell pathogene Varianten gefunden, kann die Auswertung auf weitere Gene ausgeweitet werden. Die Tiefe der Abdeckung und Sensitivität sind geringer als bei der Gen-Panel-Analyse und auch die diagnostische Ausbeute schwankt, liegt im Mittel bei ca. 25 % für mendelsche Erkrankungen (Rehm et al. 2013; Zhong et al. 2021).

Das WGS zeigt eine ähnliche, aber gleichmäßigere Abdeckung als das WES. Hierbei wird das gesamte Genom sequenziert, wodurch eine Anreicherung der gesuchten Bereiche wie bei Gen-Panels und WES entfällt (Nisar et al. 2021). Die diagnostische Ausbeute liegt etwas höher als beim WES (27 %) (Zhong et al. 2021). Mit der Ganz-Genom-Sequenzierung können zusätzlich Kopienzahlen analysiert werden (Nakagawa and Fujita 2018).

Die NGS-Analyse wird mit den Protokollen und dazugehörigen Reagenzien der Firma Twist Bioscience durchgeführt. Die Library Preparation, welche die Fragmentierung, das End-Repair und das dA-Tailing umfasst, folgt dem Protokoll „Enzymatic Fragmentation and Twist Universal Adapter System“, die Anreicherung der gewünschten Target-Bibliothek wird im „Twist Target Enrichment Standard Hybridization v1 Protocol“ abgearbeitet.

Das selbst zusammengestellte TWIST Custom Panel ID2020 umfasst 0,541 Mb, das TWIST Human Comprehensive Exome 36,8 Mb.

Die anschließende Vorbereitung zur Sequenzierung erfolgt mit den Protokollen der Firma Illumina, die Sequenzierung findet auf einem NextSeq 550 (Illumina) statt. Die Qualitätskontrolle wird mithilfe der Software von Illumina Sequencing Analysis Viewer v2.4.7 und der Plattform varvis® Version 1.22.3 (Limbus Medical Technologies GmbH) analysiert. Die Auswertung der Daten erfolgt ebenfalls mit der Plattform varvis® und SEQUENCE Pilot der Firma JSI medical systems Version 5.1.0.

2.3.5. SNP-ID Analyse mittels NGS

In Zusammenarbeit mit der Firma Twist Bioscience sollte eine Möglichkeit zur SNP-ID Analyse entwickelt werden, die mit dem verwendeten Protokoll für Custom-Panel- oder Exomanreicherung und -sequenzierung vereinbar ist, sodass Patient*innen-Proben und SNP-ID zusammen in einem NGS-Lauf analysiert werden. Ziel dabei war zusätzlich, EDTA-Vollblut als Ausgangsmaterial für die SNP-ID zu nutzen. Im Folgenden werden das SureDirect Blood PCR Kit (Agilent) und das TWIST Human Sample ID Kit verglichen.

- SureDirect Blood PCR Kit, Agilent

Um die 24 von Pengelly et. al beschriebenen SNP aus EDTA-Vollblut anzureichern, wurde das SureDirect Blood PCR Kit der Firma Agilent Technologies verwendet. Um eine optimale Sequenzierung und Abdeckung zu erreichen, wurden die Primer (100 pM) wie folgt vor dem Poolen mit Tris-HCl pH 8 verdünnt:

Tabelle 2: Primerverdünnung für SureDirect Blood PCR Kit

Gen	rs-Nummer	Verdünnung			
		Lauf 1	Lauf 2	Lauf 3	Lauf 4
<i>NPHS2</i>	rs1410592	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:5 (20µM)
<i>IL12RB2</i>	rs2229546	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>ABCB11</i>	rs497692	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>COL4A4</i>	rs10203363	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>SUMF1</i>	rs2819561	1:2 (50µM)	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)
<i>EVC</i>	rs4688963	1:2 (50µM)	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)
<i>VCAN</i>	rs309557	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)	1:5 (20µM)
<i>GRM1</i>	rs2942	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>ABCA13</i>	rs17548783	1:2 (50µM)	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)
<i>PDP1</i>	rs4735258	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)	1:5 (20µM)
<i>TDRD7</i>	rs1381532	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>HPSE2</i>	rs10883099	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>SOX6</i>	rs4617548	1:2 (50µM)	1:5 (20µM)	1:5 (20µM)	1:5 (20µM)
<i>WNK1</i>	rs7300444	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>FREM2</i>	rs9532292	1:2 (50µM)	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)
<i>L2HGDH</i>	rs2297995	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)
<i>SLC12A6</i>	rs4577050	1:2 (50µM)	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)
<i>AARS</i>	rs2070203	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:5 (20µM)
<i>COG1</i>	rs1037256	1:10 (10µM)	1:2 (50µM)	1:5 (20µM)	1:5 (20µM)
<i>LAMA3</i>	rs9962023	1:10 (10µM)	1:2 (50µM)	1:5 (20µM)	1:5 (20µM)
<i>DNMT1</i>	rs2228611	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)	1:5 (20µM)
<i>FERMT1</i>	rs10373	1:10 (10µM)	1:5 (20µM)	1:5 (20µM)	1:10 (10µM)
<i>NDUFV3</i>	rs4148973	1:10 (10µM)	1:2 (50µM)	1:5 (20µM)	1:5 (20µM)
<i>SERPIND1</i>	rs4675	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)	1:10 (10µM)

Die verdünnten Primer werden im Verhältnis 1:1 gepoolt und in die PCR-Reaktion eingesetzt.

Der PCR- Mastermix wird auf Eis vorbereitet:

Reagenz	Volumen pro Reaktion [µL]
dH ₂ O	5,5
SureDirect Blood PCR 2x Mastermix	12,5
Primer Mix	2
Total	20,0

Je nach Probenanzahl werden die Volumina angepasst.

20 µL des Mastermixes werden in 0,2 mL- PCR-Tubes vorgelegt. Als Template werden 5 µL EDTA-Vollblut pro Probe eingesetzt.

PCR-Programm für die Reaktion:

Schritt	Temperatur [°C]	Zeit	Zyklenanzahl
1	90	5 min	
2	95	30 sek	30x
3	60	30 sek	
4	72	1 min	
5	72	5 min	
6	4	Hold	

Nach der PCR werden die PCR-Tubes bei 13200 rpm für 20 min zentrifugiert, anschließend werden 10 µL jedes Überstandes in eine Hardshell-Platte überführt.

Die 10 µL dienen als Input für die Durchführung des Protokolls Enzymatic Fragmentation and Twist Universal Adapter System (Twist Bioscience).

Folgende Änderungen zum Protokoll wurden durchgeführt:

- Fragmentierungszeit: 0 min, da die Fragmente durch die Primerwahl bereits die gewünschte Größe haben
- Für die SNP-ID müssen immer andere Indices als die Indices der Proben des Custom Panel oder Exom verwendet werden, um bei der Datenauswertung die Unterscheidung zwischen Proben- und SNP-ID-Daten zu gewährleisten

Nach der Library Preparation wird eine Qualitätskontrolle mittels Qubit™ 1X dsDNA HS Assay Kit (Invitrogen) durchgeführt.

Je 2 µL, bzw. in späteren Testläufen 80 ng jeder Probe werden anschließend gepoolt und 1:1000 mit dH₂O verdünnt. 3 µL der Verdünnung werden in den denaturierten Library-PhiX-Pool (Custom Panel und Exom) hinzugegeben und auf den NextSeq 550 geladen. Nach dem Lauf wird eine Qualitätskontrolle des NGS-Laufes durchgeführt.

Die Auswertung der SNP ID Daten erfolgt wie auch die Auswertung des Custom Panel/Exom durch varvis® (Limbus Medical Technologies GmbH).

- Twist Human Sample ID Kit

In Zusammenarbeit mit Twist Bioscience wurde die α- und β-Version des mittlerweile kommerziell erhältlichen Twist Human Sample ID Kit getestet. Besonders an diesem Kit ist, dass weitere SNP hinzugefügt und/oder SNP geändert wurden, unter anderem vier SNP auf den Gonosomen. Insgesamt können nun 31 SNP nachgewiesen werden. Allerdings sind nicht alle SNP im TWIST Custom Panel ID2020 oder Exom integriert, sodass sich die Auswertung im TWIST Custom Panel ID2020 auf 18 SNP und im Exom auf 20, bzw. 18 SNP (je nach Geschlecht) begrenzt. Die SNP werden in folgender Tabelle dargestellt:

Tabelle 3: SNP im TWIST Custom Panel ID2020 und Exom

Chromosom	rs-Nummer	Custom Panel ID2020	Exom
1	rs6666954		
2	rs4411641		
2	rs497692	✓	✓

2	rs10203363	✓	✓
3	rs2819561	✓	✓
3	rs11130795		
4	rs6841061		
6	rs9388856		
6	rs2942	✓	✓
7	rs17548783	✓	✓
8	rs4735258	✓	✓
8	rs1393978		
9	rs12682834		
9	rs1381532	✓	✓
10	rs10883099	✓	✓
12	rs7300444	✓	✓
12	rs10748087		
13	rs9532292	✓	✓
15	rs4577050	✓	✓
16	rs2070203	✓	✓
17	rs1037256	✓	✓
18	rs9962023	✓	✓
19	rs2228611	✓	✓
20	rs10373	✓	✓
20	rs6512586		
21	rs4148973	✓	✓
22	rs4675	✓	✓
Y	rs2032653		✓
Y	rs34581739		✓
Y	AMELY		
X	AMELX		

Der Unterschied zur Durchführung mit dem SureDirect Blood PCR Kit ist, dass die Multiplex-PCR mit vorgefertigtem Mastermix und Primermix mit 2 µL EDTA-Vollblut oder 10 ng DNA gestartet wird. Nach der Multiplex-PCR fällt die Zentrifugationszeit auf 3 min bei 2250 rpm und es werden lediglich 1 µL Überstand pro Probe als

Eingangsmaterial in das Twist Bioscience Library Preparation Kit eingesetzt. Wichtig ist auch hier wieder, dass unterschiedliche Indices verwendet werden.

Nach der Library Preparation werden sowohl die SNP-ID-Library als auch die Custom Panel- oder Exom-Library auf 4,3 nM mit RSB-Puffer verdünnt. Die Qualitätskontrolle wird mit Qubit™ 1X dsDNA HS Assay Kit (Invitrogen) durchgeführt.

Um die Flowcell optimal zu beladen und eine ausreichende Abdeckung aller Proben zu gewährleisten, wird der verdünnte SNP-ID-Pool je nach Probenanzahl mit dH₂O verdünnt. Der verdünnte SNP-ID-Pool wird im Verhältnis 1:12 mit dem Custom Panel- oder Exom-Pool (4,3 nM) gemischt.

Die Vorbereitung für die Sequenzierung erfolgt anhand des Protokolls der Firma Illumina. Im Gegensatz zum SureDirect Blood PCR Kit wird der TWIST SNP-ID-Library-Pool vor der Vorbereitung für die Sequenzierung eingespikt und somit auch denaturiert.

Nach dem Lauf wird eine Qualitätskontrolle durchgeführt.

Die Auswertung der SNP ID Daten erfolgt erneut durch varvis® (Limbus Medical Technologies GmbH).

3. Ergebnisse

Die Durchführung der Tests wurden im Rahmen der molekulargenetischen Routinediagnostik am MVZ Humane Genetik München durchgeführt. Da hierbei eine große Menge an Patient*innen analysiert wurden, werden im folgenden Ergebnisteil nur repräsentative Beispiele gezeigt. Die genomischen Positionen im NGS folgen dem Referenzgenom hg37.

3.1. Sanger-Sequenzierung

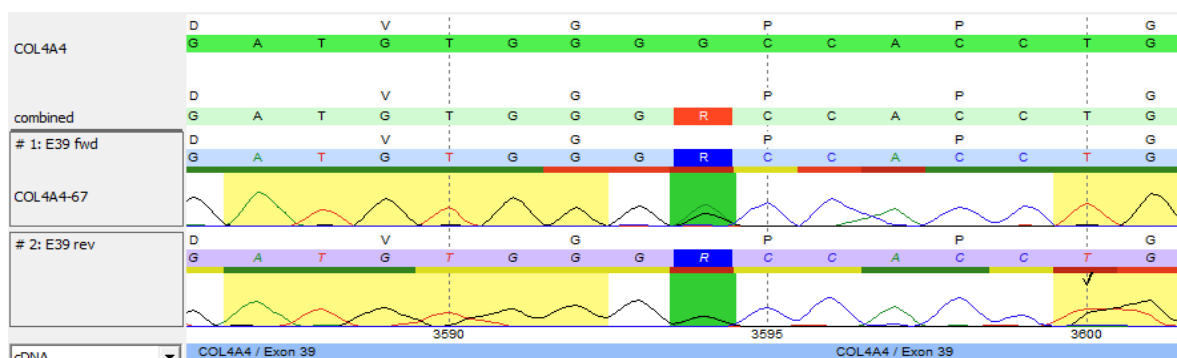
Die bei der Sanger-Sequenzierung entstandenen Ergebnisse der SNP-ID werden mit denen der NGS-Analyse verglichen. Dies kann aufgrund der unterschiedlichen Analysetechniken nicht in derselben Auswertepattform geschehen, sondern wird

manuell verglichen. In der SeqPilot Software für die Auswertung der Sanger-Sequenzierung ist die relevante Position dunkelgrün hinterlegt.

Hier werden repräsentative Ergebnisse zweier Proben präsentiert, die vollständigen Ergebnisse sind im Anhang zu finden:

Probe A

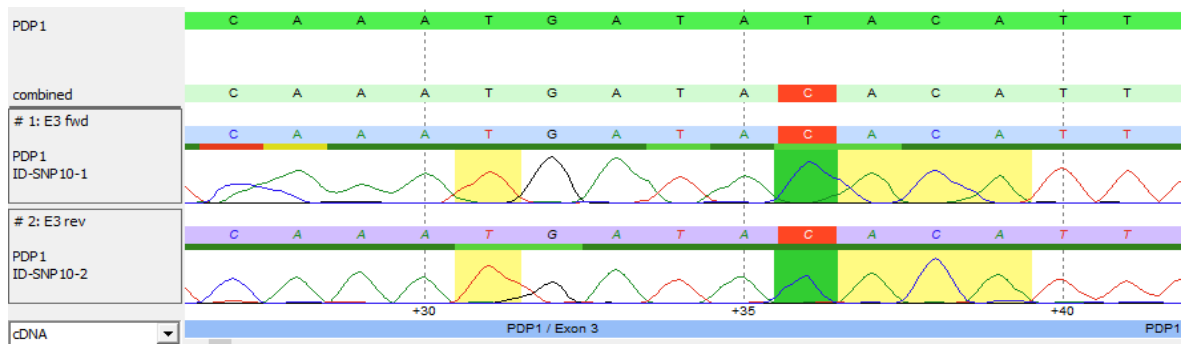
SNP rs10203363 (COL4A4 Exon 39 c.3594)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
227896976	rs1023363	COL4A4	Exon 39 c.3594	C	T	0,3930130	heterozygot

Bei diesem SNP wird in der Sanger-Sequenzierung der Minus-Strang sequenziert, sodass die Referenzbase komplementär ist (Guanin, im NGS Cytosin). Die Sanger-Sequenz zeigt in beiden Strängen ein eindeutig heterozygoten Ergebnis. Im NGS wird mit einer AltAF von 0,3930130 (= ~39 %) ebenfalls ein heterozygoten Ergebnis erzielt. Die AltAF gibt an, wie häufig das alternative Allel detektiert wurde.

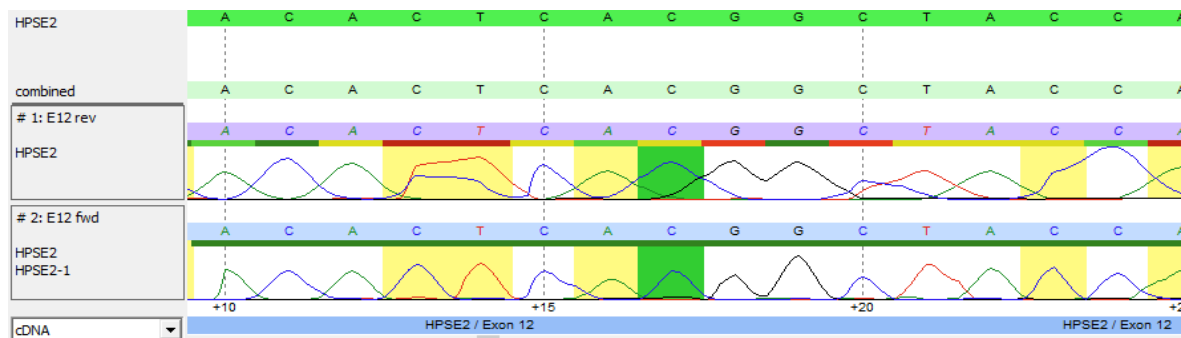
SNP rs4735258 (*PDP1* Ex3/3'UTR c.*36)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
94935937	rs4735258	<i>PDP1</i>	Ex3/3'UTR c.*36	T	C	1	homozygot

Beide Methoden zeigen das übereinstimmende Ergebnis einer homozygoten Transition (T>C). Im NGS wird dies mit einer AltAF von 1 deutlich.

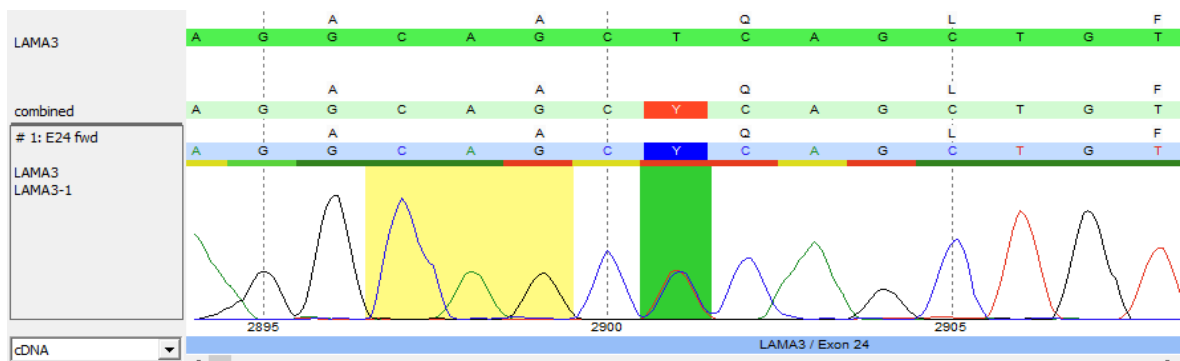
SNP rs10883099 (*HPSE2* Ex13/3'UTR c.*17)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
100219314	rs10883099	<i>HPSE2</i>	Ex13/3'UTR c.*17	G	G	0	Ref

Bei diesem SNP erfolgt die Sanger-Sequenzierung anhand des Minusstranges, zeigt aber ein deutliches, homozygotes Ergebnis für die Referenzbase G. Dieses Ergebnis wird auch im NGS erzielt. Die unterschiedlichen Exonbezeichnungen kommen durch die unterschiedlichen Transkripte, die für beide Methoden verwendet werden, zustande.

SNP rs9962023 (*LAMA3* Exon 24 c.2901)



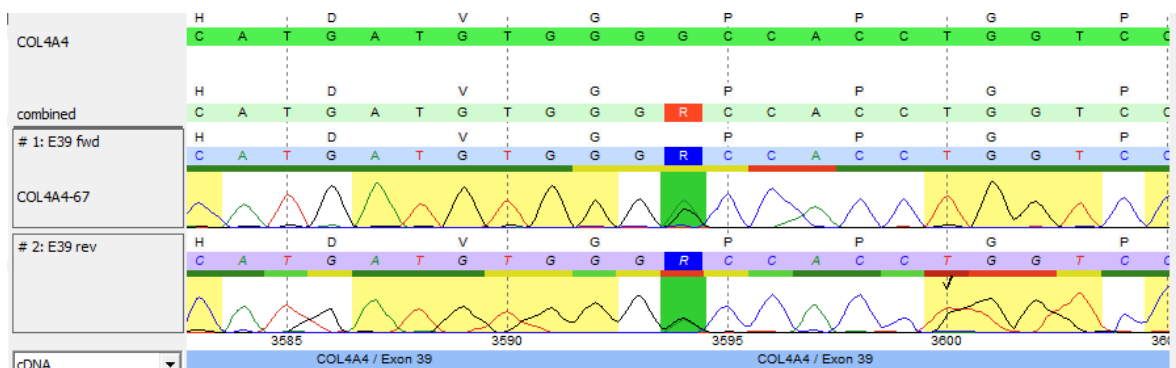
Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
21413869	rs9962023	<i>LAMA3</i>	Exon 24 c.2901	T	C	0,4651162	heterozygot

Der SNP im Gen *LAMA3* ist nur in der Forward-Leserichtung darstellbar. Trotzdem ist das heterozygote Ergebnis (T>C) eindeutig und wird auch im NGS mit einer AltAF von 0,4651162 erreicht.

Auch bei Probe B werden teilweise in der Sanger-Sequenz die Minusstränge sequenziert, was zu komplementären Referenzbasen führt. Dies wird bei der Beschreibung der jeweiligen SNP nicht erneut aufgeführt.

Probe B

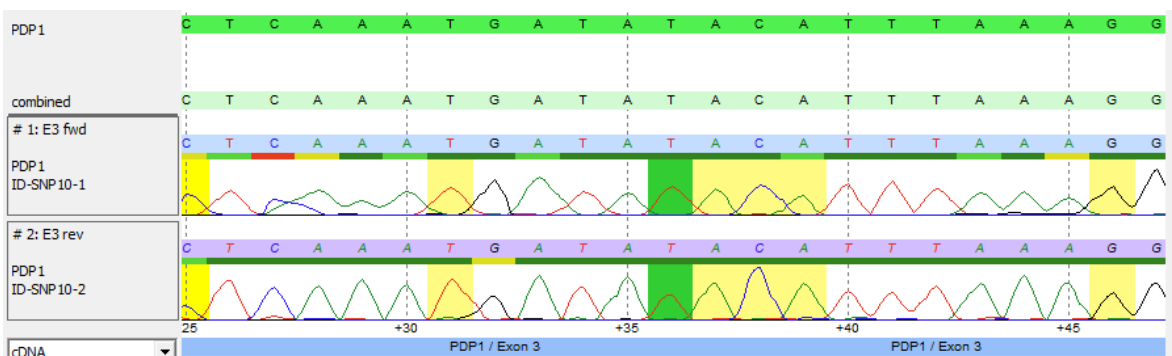
SNP rs10203363 (*COL4A4* Exon 39 c.3594)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
227896976	rs10203363	COL4A4	Exon 39 c.3594	C	T	0,5136612	heterozygot

Ein heterozygoten Ergebnis zeigen beide Methoden für den SNP rs10203363 im COL4A4-Gen (G>C, bzw., C>T). Im NGS wird zu ca. 52 % das alternative Allel detektiert.

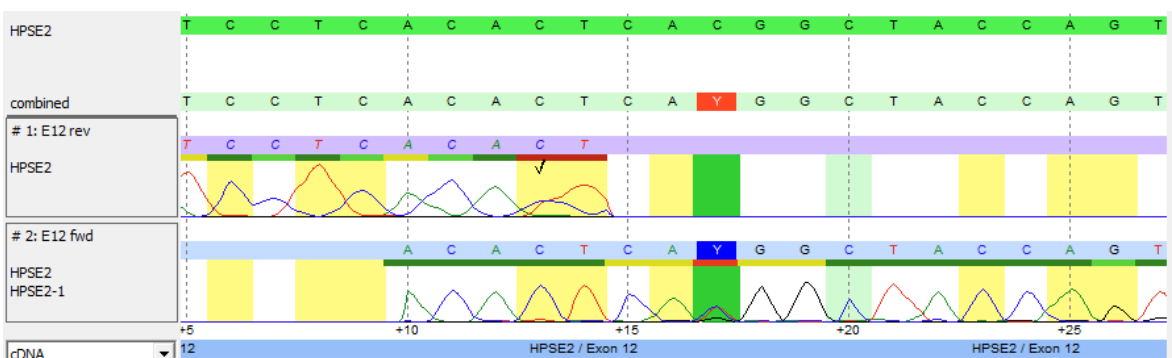
SNP rs4735258 (*PDP1* Ex3/3'UTR c.*36)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
94935937	rs4735258	PDP1	Ex3/3'UTR c.*36	T	T	0	Ref

Das Referenzallel (Thymin) wird in homozygoter Konstellation sowohl in der Sanger-Sequenz als auch im NGS für den SNP rs4735258 detektiert.

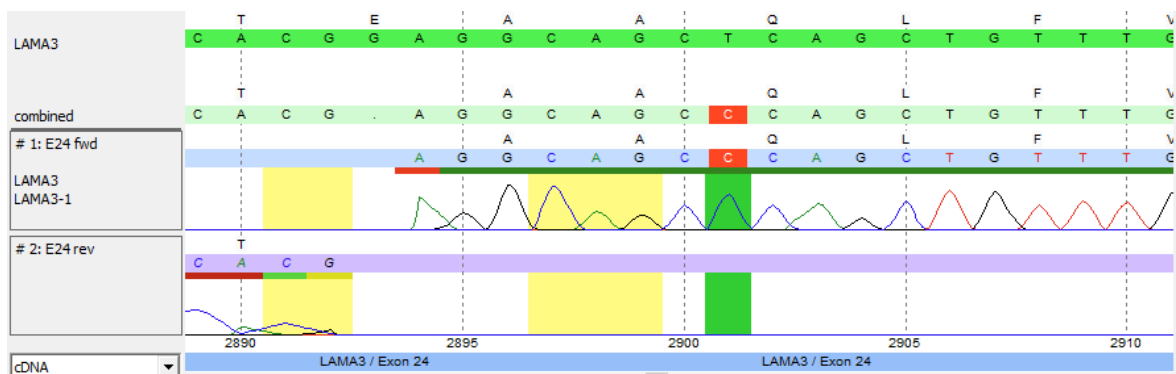
SNP rs10883099 (*HPSE2* Ex13/3'UTR c.*17)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
100219314	rs10883099	<i>HPSE2</i>	Ex13/3'UTR c.*17	G	A	0,5	heterozygot

Der SNP rs10883099 wird nur in der Forward-Leserichtung abgedeckt. Hier ist die Sequenz nicht eindeutig sauber, trotzdem kann von einem heterozygoten Ereignis (C>T) ausgegangen werden. In der NGS-Analyse wird ebenfalls ein heterozygoten Ergebnis mit einer AltAF von 0,5 erreicht.

SNP rs9962023 (*LAMA3* Exon 24 c.2901)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
21413869	rs9962023	<i>LAMA3</i>	Exon 24 c.2901	T	C	1	homozygot

Ebenfalls zu kurz ist die Leseweite in der Reverse-Leserichtung in der Sanger-Sequenzierung des SNP rs9962023. In der Forward-Leserichtung wird die alternative Base Cytosin eindeutig detektiert, was für ein homozygoten Ergebnis spricht, welches auch dem Ergebnis der NGS-Analyse entspricht.

Sowohl bei Probe A als auch bei Probe B stimmen die Allele der SNP aus Sanger-Sequenzierung und NGS überein. Demnach kann eine Probenverwechslung oder Kontamination mit Fremd-DNA nahezu ausgeschlossen werden. Falls eine Probenverwechslung und/oder Kontamination vorliegen würde, müssten sich die Allele aus NGS und Sanger in mindestens einem SNP unterscheiden.

Der SNP rs2819561 im *SUMF1*-Gen kann im NGS nicht detektiert werden, was bei allen Patient*innen so beobachtet wurde. Aus diesem Grund wird dieser SNP aus der Bewertung genommen und steht für den Vergleich der SNP nicht zur Verfügung. Es bleiben also lediglich 17 SNP zur Identifikation einer Verwechslung oder Kontamination.

Anhand der zwei Beispiele ist bereits gut ersichtlich, dass die ausgewählten SNP sich zur Diskriminierung von Proben eignen, da die beiden Proben sich in neun der 18, bzw. 17 SNP unterscheiden.

3.2. SureDirect Blood PCR Kit

Da der SNP-ID-Library Pool in den Proben-PhiX-Pool eingespiket wird, ist es möglich, eine automatische Gegenüberstellung der SNP mit der Plattform varvis® zu erhalten. Hauptaugenmerk bei der Testung lag neben der Übereinstimmung der Allele der SNP pro Patient*in, dem korrekten Geschlechtsabgleich und Mix-up check auf der Tiefe der Calls bei unterschiedlichen Verdünnungen.

Im Folgenden wird pro Testlauf auf eine repräsentative Probe gezeigt. Sowohl der Geschlechtsabgleich als auch der Mix-up check erfüllen die erforderlichen Kriterien.

Testlauf 1

Tabelle 4: Beispiel Probe 1 SureDirect Blood PCR Kit (Lauf 1)

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
1	rs2229546	C	A	A	Yes	567	33	1,00	1,00
1	rs1410592	G	A	A	Yes	530	38	0,52	0,41
1	rs497692	T	C	C	Yes	441	90	0,45	0,54
1	rs1023363	C	T	T	Yes	509	36	1,00	1,00
1	rs2819561	A	G	G	Yes	542	33476	0,49	0,48
1	rs4688963	T	T	T	Yes	541	18033	0,00	0,00
1	rs309557	T	C	C	Yes	521	11	1,00	1,00
1	rs2942	G	A	A	Yes	405	79	0,44	0,40
1	rs17548783	T	T	T	Yes	548	3668	0,00	0,00
1	rs4735258	T	C	C	Yes	567	20	0,49	0,50
1	rs1381532	A	G	G	Yes	584	52	0,48	0,46
1	rs10883099	G	A	A	Yes	525	98	1,00	1,00
1	rs4617548	A	A	A	Yes	387	2209	0,00	0,00
1	rs7300444	C	T	T	Yes	603	80	0,49	0,46

1	rs9532292	A	G	G	Yes	469	15170	0,47	0,49
1	rs2297995	G	A	A	Yes	425	33	0,44	0,45
1	rs4577050	G	A	A	Yes	565	12963	0,46	0,49
1	rs2070203	G	G	G	Yes	617	22	0,00	0,00
1	rs1037256	G	A	A	Yes	527	9	1,00	1,00
1	rs9962023	T	C	C	Yes	497	7	1,00	1,00
1	rs2228611	T	C	C	Yes	509	19	0,48	0,62
1	rs10373	A	G	G	Yes	542	13	0,49	0,38
1	rs4148973	T	G			517	0	0,50	0,00
1	rs4675	T	C	C	Yes	581	59	0,48	0,62

In der Tabelle werden die genomische Referenzen der 24 SNP angegeben, ebenso die Base, die im NGS (sample call) und die Base, die im SNP ID (SNP ID call) ermittelt wurde. Diese werden im Match verglichen. Stimmen die Basen überein, wird Match ‚Yes‘ angegeben, sollte es eine Unstimmigkeit geben, wird ‚No‘ angezeigt. Bleibt die Spalte leer, kann entweder in der SNP ID oder im Custom Panel/Exom keine Base detektiert werden. Die Sample AltAF und SNP ID AltAF geben an, zu wie viel Prozent die alternative Base eingebaut wurde. Eine AltAF von 0 bedeutet, dass kein alternatives Allel detektiert wird, eine AltAF von ca. 0,5 spricht für Heterozygotie und 1 für Homozygotie für das alternative Allel. Ob ein Match zutrifft, hängt neben der detektierten Base auch von der übereinstimmenden Zygotie ab.

Zusätzlich in der Tabelle ersichtlich ist die Tiefe (depth) von Probe und SNP ID. Diese gibt die durchschnittliche Anzahl an sequenzierten Basen an, die die Referenzbase abdecken. Diese sollte bei der Probe über 150 liegen, um ein aussagekräftiges Ergebnis zu erhalten. Bei der Probe kann man sehen, dass die Sample depth zwischen 387 und 617 liegt, was für eine gute, ausgeglichene Abdeckung der SNP spricht. Die Tiefe beim SNP ID liegt zwischen 0 und 33476, was zeigt, dass die Tiefen der unterschiedlichen SNP stark schwanken. Der SNP auf Chromosom 21 kann nicht detektiert werden, hier wird kein Match angezeigt. Alle anderen SNP erhalten das Match ‚Yes‘, womit davon ausgegangen werden kann, dass diese Probe weder kontaminiert noch vertauscht wurde. Allerdings erreichen manche SNP nicht die benötigte Tiefe von 20. Zur endgültigen Bewertung müssen alle Proben des Laufes in die Bewertung einfließen, was hier aufgrund der Menge an Daten nicht möglich ist. In diesem Lauf wurden 24 Proben analysiert.

Die Mittelwerte der SNP sind in folgender Abbildung gezeigt:

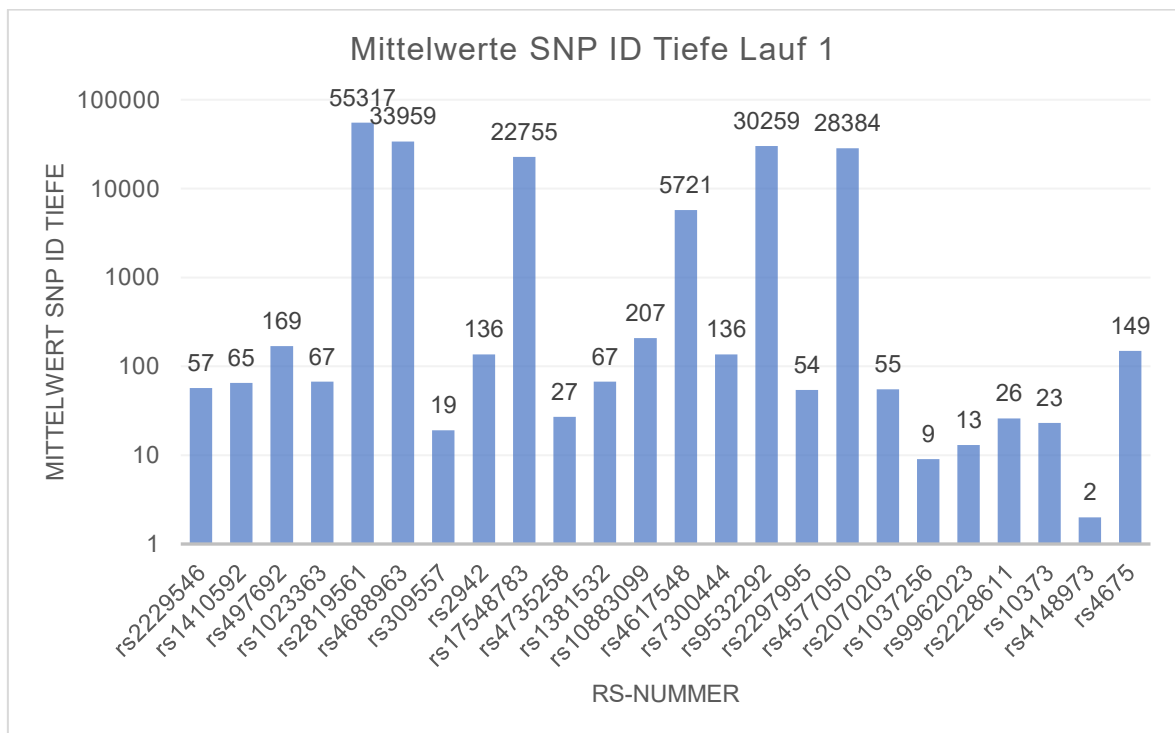


Abbildung 7: Mittelwerte der SNP ID Tiefe Lauf 1, logarithmische Darstellung

Im Diagramm Abbildung 7 werden die 24 SNP gegen die Mittelwerte der Tiefen aller im Lauf analysierten Proben (24) gezeigt. Zur besseren Darstellung wird eine logarithmische Skalierung der Mittelwerte gewählt. In der Abbildung ist zu sehen, dass die SNP, die bei Probe 1 eine sehr hohe oder eine sehr niedrige Tiefe erreichen, auch bei den anderen Proben des Laufes hohe Werte erreichen.

Für eine ausgeglichene Tiefe beim SNP ID, bzw. damit alle SNP für ein sicheres Ergebnis mindestens mit einer Tiefe von 20 abdeckt werden, wurden die Primer für den nächsten Lauf neu verdünnt (siehe Tabelle 2). Die neue Verdünnung wurde anhand der Mittelwerte der Tiefe aller SNP ID Proben pro SNP des Laufes gewählt. Angepasst wurden die Primer, die sehr hohe oder niedrige (< 20) Tiefen anzeigen.

Testlauf 2

Tabelle 5: Beispiel Probe 2 SureDirect Blood PCR Kit (Lauf 2)

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
2	rs2229546	C	A	A	Yes	323	37	0,50	0,43
2	rs1410592	G	A	A	Yes	357	41	0,48	0,63
2	rs497692	T	C	C	Yes	242	81	1,00	1,00
2	rs1023363	C	T	T	Yes	352	52	0,44	0,45
2	rs2819561	A	G	G	Yes	297	7	0,49	0,71
2	rs4688963	T	C	C	Yes	282	3	1,00	1,00
2	rs309557	T	C	C	Yes	319	280	0,50	0,45
2	rs2942	G	A	A	Yes	266	90	0,44	0,51
2	rs17548783	T	C	C	Yes	355	5	1,00	1,00
2	rs4735258	T	C	C	Yes	383	681	0,51	0,49
2	rs1381532	A	G	G	Yes	335	67	0,48	0,44
2	rs10883099	G	A	A	Yes	317	144	1,00	1,00
2	rs4617548	A	G	G	Yes	262	33	1,00	1,00
2	rs7300444	C	T	T	Yes	363	105	0,46	0,50
2	rs9532292	A	G	G	Yes	246	4	0,47	0,75
2	rs2297995	G	G	G	Yes	243	40	0,00	0,00
2	rs4577050	G	A	A	Yes	401	4	0,50	0,43
2	rs2070203	G	G	G	Yes	364	44	0,00	0,00
2	rs1037256	G	G	G	Yes	371	129150	0,00	0,00
2	rs9962023	T	C	C	Yes	331	150235	0,50	0,49
2	rs2228611	T	C	C	Yes	354	630	0,50	0,48
2	rs10373	A	A	A	Yes	299	1629	0,00	0,00
2	rs4148973	T	G	G	Yes	329	2962	0,51	0,46
2	rs4675	T	C	C	Yes	361	118	0,47	0,48

Die Übersicht der SNP des zweiten Testlaufes entspricht der Übersicht aus Lauf 1, jedoch handelt es sich um eine andere Probe, sodass die Basen abweichen können. Bei Probe 2 erhalten alle SNP das Match ‚Yes‘, was bedeutet, dass die detektierte Base im NGS der Base im SNP ID entspricht und die Zygote identisch ist. Auch hier kann davon ausgegangen werden, dass die Zuordnung der Probe zum* zur Patient*in korrekt ist, muss aber im Kontext aller Proben gesehen werden.

Die Tiefen der SNP der Proben (sample depth) sind gleichmäßig, sodass davon ausgegangen werden kann, dass die Auswertung der Probe erfolgreich sein wird. Trotz Anpassung der Primerverdünnungen ist ersichtlich, dass nicht alle SNP die erforderliche Tiefe von > 20 erreicht haben (SNP rs4688963, rs4688963, rs17548783, rs9532292, rs4577050) und die Tiefe mancher SNP auch noch

deutlich zu hoch ist (vor allem SNP rs1037256 und rs9962023). Auch nach diesem Lauf werden anhand der Mittelwerte der SNP ID Tiefe aus allen Proben des Laufes (40) (siehe Abbildung 8) die Primerkonzentrationen für den nächsten Testlauf erneut angepasst.

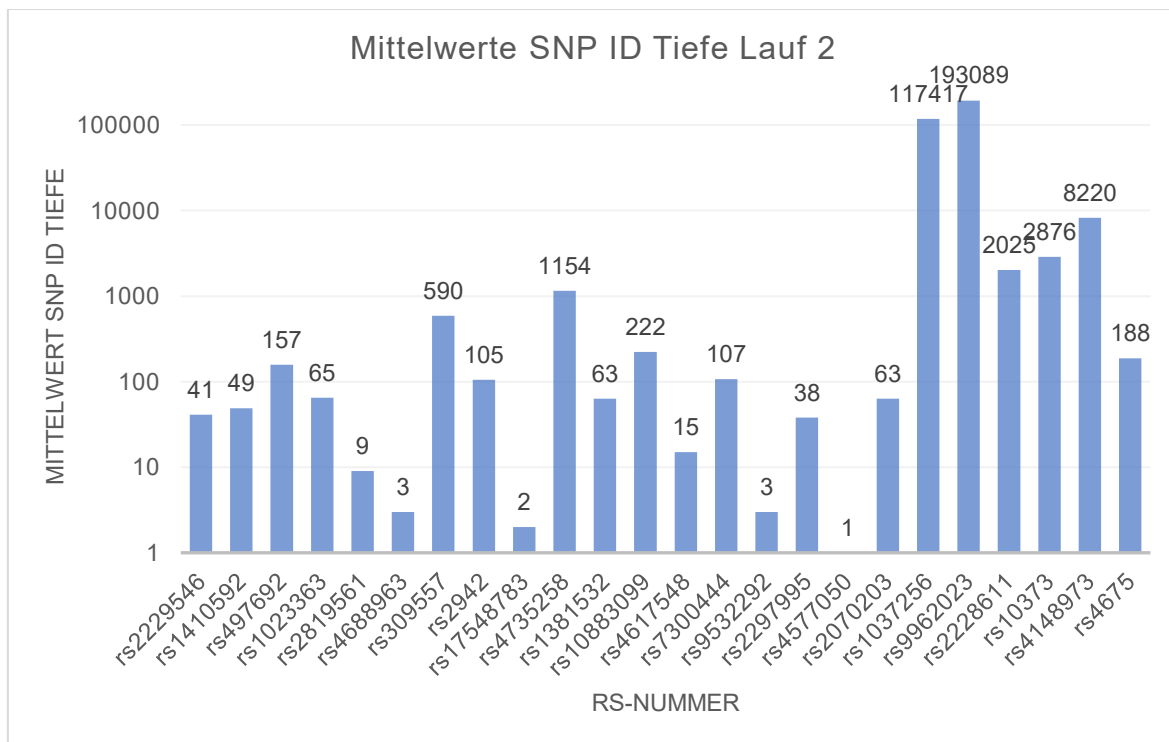


Abbildung 8: Mittelwerte der SNP ID Tiefe Lauf 2, logarithmische Darstellung

In der Abbildung werden wieder die SNP gegen die Mittelwerte der SNP ID Tiefen aller Proben des Laufes dargestellt. Die Mittelwerte der SNP ID Tiefen bestätigen, dass manche SNP nach wie vor nicht ausreichend abgedeckt sind (beispielsweise SNP rs17548783 und rs4577050) oder noch deutlich zu hoch sind, wie der SNP rs1037256 und rs9962023. Für den nächsten Testlauf werden die Verdünnungen der Primer deshalb weiter angepasst (siehe Tabelle 2 Lauf 3).

Testlauf 3

Tabelle 6: Beispiel Probe 3 SureDirect Blood PCR Kit (Lauf 3)

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
3	rs2229546	C	A	A	Yes	303	175	0,51	0,56
3	rs1410592	G	A	A	Yes	236	198	1,00	1,00

3	rs497692	T	C	C	Yes	193	512	1,00	0,99
3	rs1023363	C	T	T	Yes	257	171	0,46	0,49
3	rs2819561	A	G	G	Yes	247	874	0,49	0,49
3	rs4688963	T	T	T	Yes	217	647	0,00	0,00
3	rs309557	T	C	C	Yes	239	10003	0,55	0,50
3	rs2942	G	A	A	Yes	203	461	0,43	0,49
3	rs17548783	T	T	T	Yes	245	190	0,00	0,00
3	rs4735258	T	T	T	Yes	260	4391	0,00	0,00
3	rs1381532	A	G	G	Yes	223	260	0,54	0,46
3	rs10883099	G	A	A	Yes	239	490	1,00	1,00
3	rs4617548	A	G	G	Yes	192	105	0,48	0,41
3	rs7300444	C	T	T	Yes	253	811	0,48	0,45
3	rs9532292	A	G	G	Yes	182	797	0,45	0,48
3	rs2297995	G	A	A	Yes	206	189	0,47	0,47
3	rs4577050	G	A	A	Yes	256	2115	0,47	0,50
3	rs2070203	G	A			258	0	1,00	0,00
3	rs1037256	G	G	G	Yes	240	2645	0,00	0,00
3	rs9962023	T	C	C	Yes	233	57409	1,00	1,00
3	rs2228611	T	T	T	Yes	261	13102	0,00	0,00
3	rs10373	A	G	G	Yes	227	29486	0,49	0,49
3	rs4148973	T	G	G	Yes	246	397	0,54	0,45
3	rs4675	T	C	C	Yes	259	556	1,00	1,00

Die Ergebnisse von Probe 3 im dritten Testlauf zeigen, dass auch hier wieder alle SNP Match ‚Yes‘ anzeigen. Kein Match konnte bei dem SNP rs2070203 erzielt werden. Da dies bei allen Proben des Laufes der Fall war, wurde dieser SNP aus der Bewertung genommen. Auch ohne den SNP konnten alle Proben erfolgreich voneinander diskriminiert werden, sodass eine Probenverwechslung oder Kontamination ausgeschlossen werden kann. Trotz der erneuten Anpassungen der Primerverdünnungen schwanken die SNP ID Tiefen weiterhin, auch wenn alle SNP die erforderliche Tiefe von > 20 erreichen. Die SNP rs309557, rs9962023, rs2228611 und rs10373 bleiben weiterhin deutlich zu hoch.

Die Tiefen der SNP der Proben sind etwas geringer im Vergleich zu denen der vorherigen Läufe, trotzdem konnten die Proben gut ausgewertet werden. Bei diesem Lauf wurden 40 Proben sequenziert.

Die Mittelwerte der SNP ID Tiefen aller Proben des Laufes sind in folgender Abbildung dargestellt:

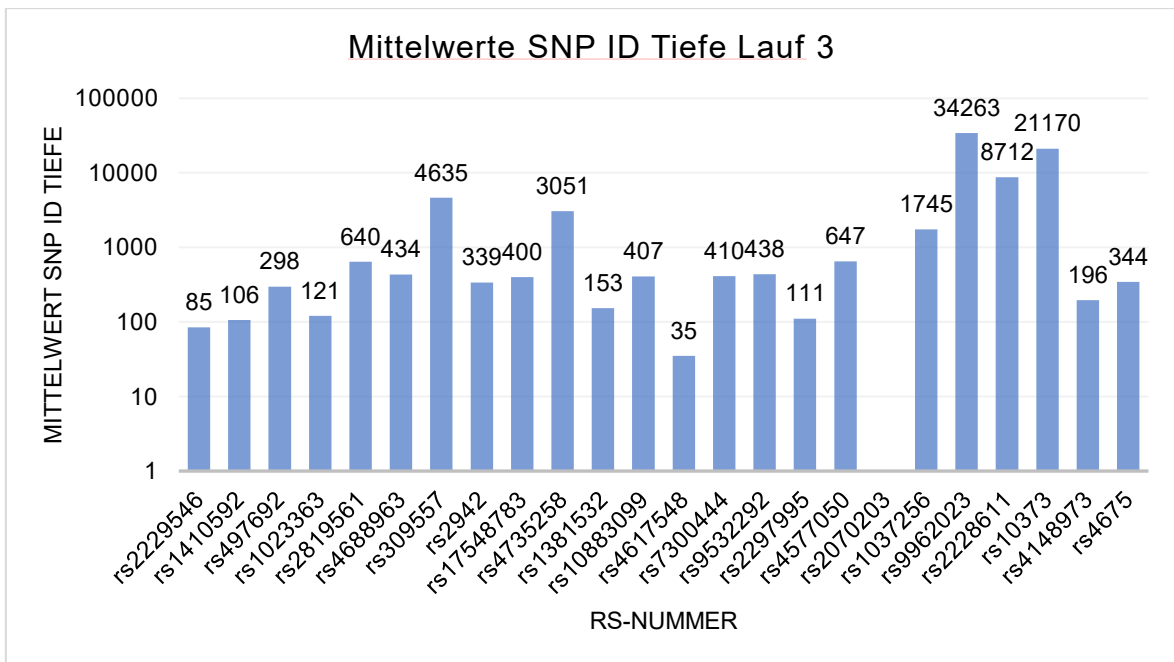


Abbildung 9: Mittelwerte SNP ID Tiefe Lauf 3, logarithmische Darstellung

Abbildung 9 zeigt die Mittelwerte der SNP ID Tiefen aller Proben gegen die 24 SNP.

Zu sehen ist, dass der SNP rs2070203 bei keiner Probe detektiert werden konnte, sodass dieser nicht bewertet wurde. In der Abbildung ist ebenfalls erkennbar, dass sich die SNP ID Tiefen etwas angleichen, was dafür spricht, dass die Anpassungen der Primerverdünnungen wirksam waren.

Für den vierten Testlauf werden nur noch wenige Primerverdünnungen angepasst (siehe Tabelle 2).

Testlauf 4

Tabelle 7: Beispiel Probe 4 SureDirect Blood PCR Kit (Lauf 4)

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
4	rs2229546	C	A	A	Yes	318	318	0,47	0,52
4	rs1410592	G	A	A	No	317	5	0,47	0,86
4	rs497692	C	T	T	Yes	275	13	0,45	0,60
4	rs1023363	T	C	C	Yes	228	15	1,00	1,00
4	rs2819561	A	A	A	Yes	299	50	0,00	0,00
4	rs4688963	T	C	C	Yes	258	9	0,53	0,65
4	rs309557	T	C	C	Yes	288	33	0,45	0,45
4	rs2942	G	A	A	Yes	215	11	0,47	0,48
4	rs17548783	T	C	C	Yes	288	28	1,00	1,00

4	rs4735258	T	C	C	Yes	345	201	0,49	0,42
4	rs1381532	A	G	G	Yes	291	7	0,51	0,67
4	rs10883099	G	A	A	Yes	268	27	1,00	1,00
4	rs4617548	A	G	A	No	196	2	0,53	0,00
4	rs7300444	C	C	C	Yes	365	12	0,00	0,00
4	rs9532292	A	A	A	Yes	252	5	0,00	0,00
4	rs2297995	G	A	A	Yes	248	7	1,00	1,00
4	rs4577050	G	G	G	Yes	312	2	0,00	0,00
4	rs2070203	G	A	A	Yes	290	548	0,54	0,51
4	rs1037256	G	A	A	Yes	308	20	0,55	0,46
4	rs9962023	T	C	C	Yes	270	32	1,00	1,00
4	rs2228611	T	C	C	Yes	308	279	1,00	0,99
4	rs10373	A	G	A	No	285	1	1,00	0,00
4	rs4148973	T	G	G	Yes	260	2	1,00	1,00
4	rs4675	T	T	T	Yes	326	19	0,00	0,00

Die Probe von Lauf 4 zeigt eine recht ausgeglichene Sample depth, allerdings wird bei drei der 24 SNP ein Match ‚No‘ angezeigt (rs1410592, rs4617548, rs10373). Bei SNP rs1410592 wird zwar im Sample und im SNP ID call das gleiche Allel detektiert, allerdings in unterschiedlicher Zygotie (Sample AltAF 0,47; SNP ID AltAF 0,86). Diese Diskrepanz kommt sehr wahrscheinlich durch die geringe Tiefe im SNP ID zustande, da die Base nur 5-fach abgedeckt ist. Bei den SNP rs4617548 und rs10373 kommt es durch die sehr geringe Tiefe im SNP ID zu einer SNP ID AltAF von 0, wodurch das Referenzallel (A) in homozygoter Form angegeben wird. Da mehrere Proben Match ‚No‘ anzeigen, vor allem bei SNP rs1410592, rs17548783, rs4617548 und rs10373, kann nicht mit endgültiger Sicherheit davon ausgegangen werden, dass keine Probenverwechslung vorliegt. In Lauf 4 wurden ebenfalls 40 Proben sequenziert.

Die Mittelwerte der Proben in Testlauf 4 werden in folgender Abbildung gezeigt.

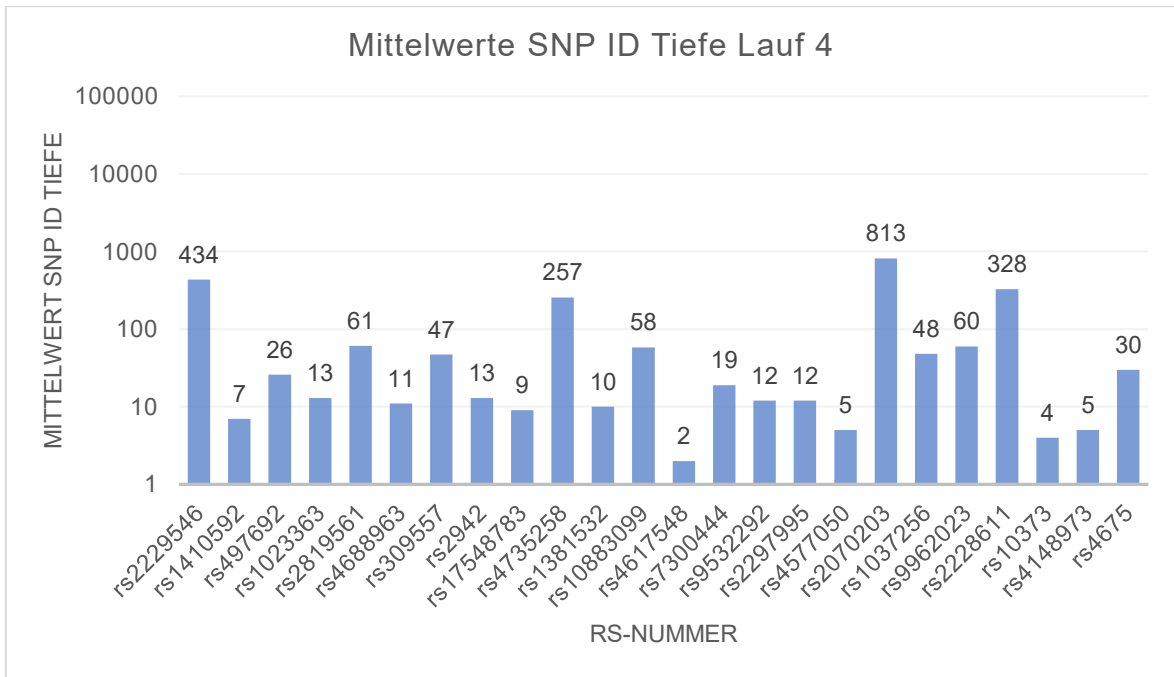
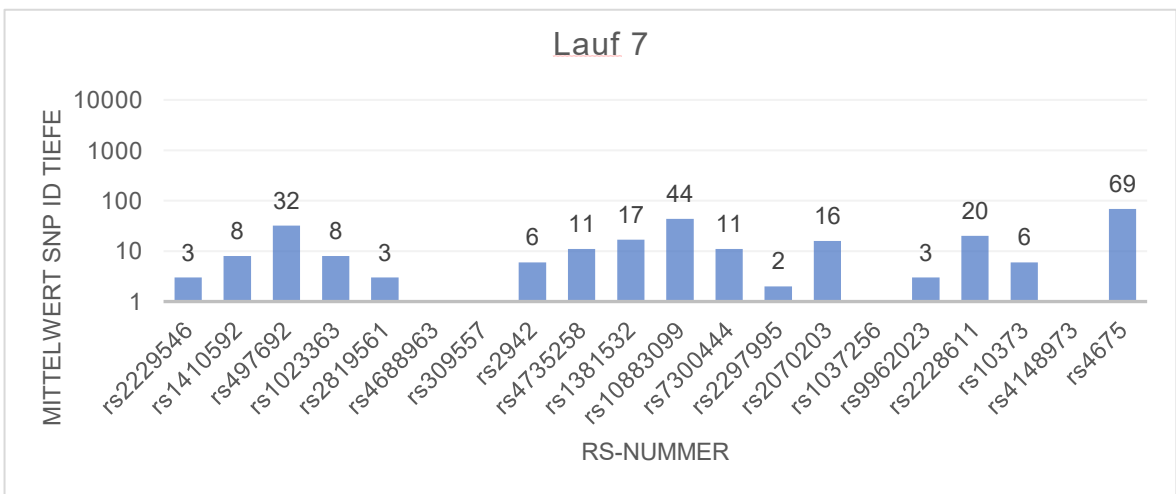
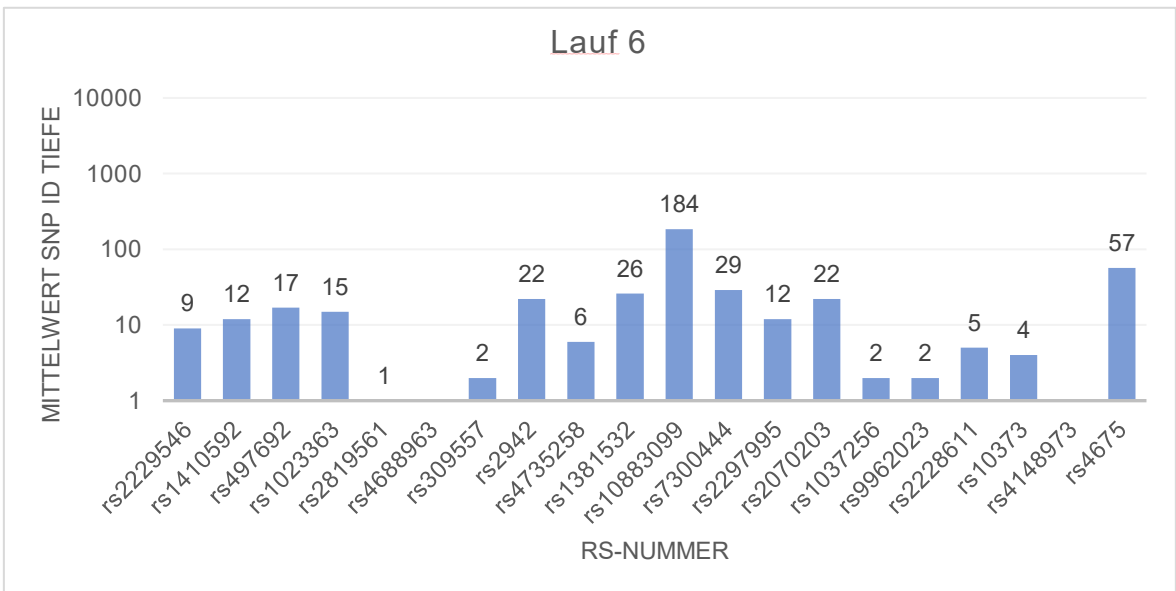
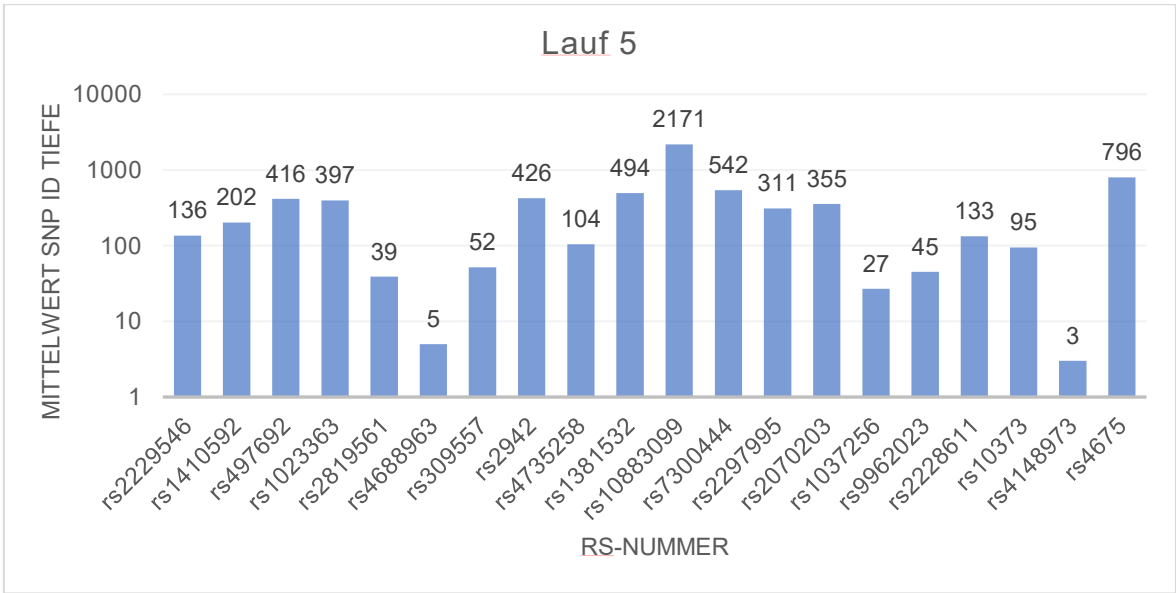


Abbildung 10: Mittelwerte SNP ID Tiefe Lauf 4, logarithmische Darstellung

Auch nach der Anpassung der Primerverdünnungen schwanken die Mittelwerte der SNP Tiefen sehr, was in Abbildung 10 gezeigt ist. Hier werden die Mittelwerte der SNP ID Tiefen aller Proben des Laufes gezeigt. Zu sehen ist, dass die Mittelwerte insgesamt deutlich geringer sind als in den vorherigen Läufen.

Ebenso können zwischen den einzelnen SNP-ID-Proben deutliche Unterschiede festgestellt werden, woraufhin für die nächsten Läufe 80 ng pro Probe eingesetzt wurden. Zusätzlich dazu wurden vier kritische SNP aus der Bewertung herausgenommen (rs17548783, rs4617548, rs9532292, rs4577050). Die Verdünnungen der Primer wurden, sofern nicht anders angegeben, aus Testlauf 4 übernommen.

Im Folgenden werden repräsentative Beispiele aus ausgewählten Läufen gezeigt.



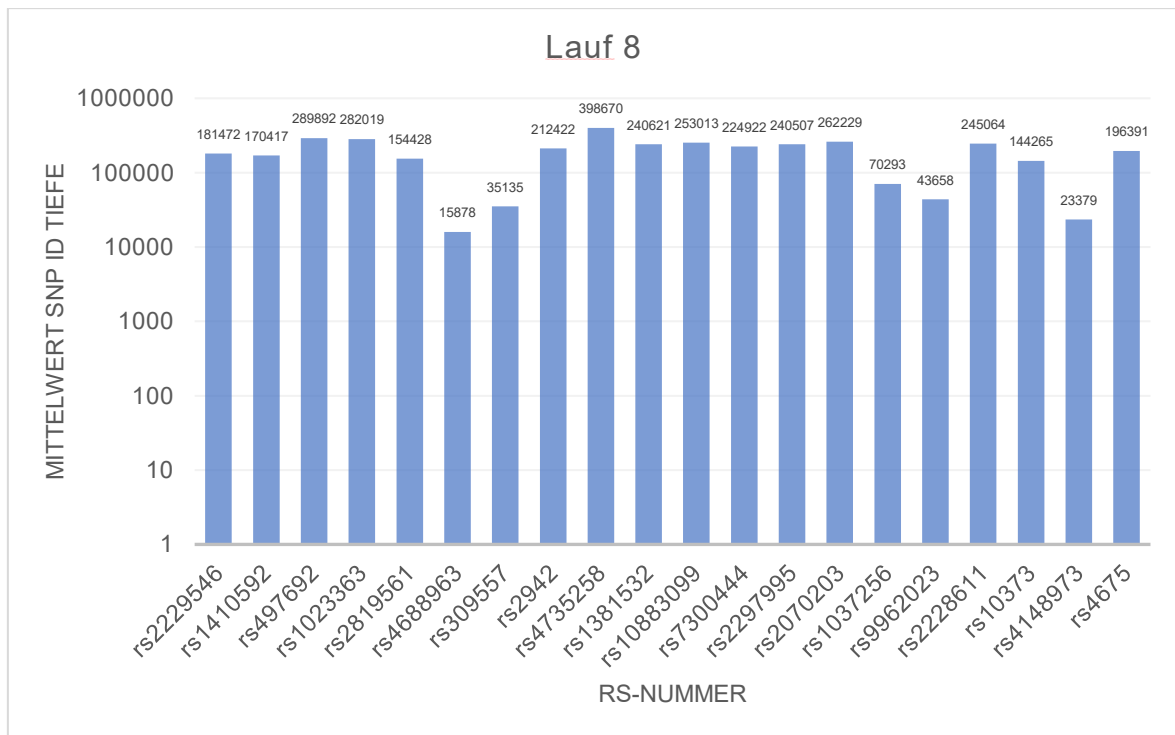


Abbildung 11: Mittelwerte SNP ID Tiefe Lauf 5 – 8, logarithmische Darstellung

Zu sehen ist, dass im Lauf 5 (8 Proben – Exomlauf) die Mittelwerte der SNP bis auf einzelne Ausnahmen (rs4688963, rs4148973) die Mindest-Tiefe von 20 erreichen. Bei diesen zwei SNP, deren Mittelwert unter 20 ist, können aufgrund der zu geringen Tiefe nicht bei allen Proben Basen detektiert werden, diese weisen das Match ‚No‘ auf. Auch bei weiteren SNP, die im Mittel zwar die Tiefe von 20 erreichen, z.B. rs1037256, kommt es wegen der geringen Tiefe zu falschen oder fehlenden Detektionen. Dies führt dazu, dass eine mögliche Probenverwechslung oder Kontamination nicht eindeutig bewertet werden kann. Im Zuge der nächsten Läufe (Beispiel Lauf 6 (40 Proben) und 7 (32 Proben)) wurden bei gleichbleibenden Bedingungen im Mittel immer geringere Tiefen detektiert oder manche SNP gar nicht mehr abgedeckt, wodurch keine einzige Probe in allen 20 SNP Match ‚Yes‘ aufweisen konnte (Daten nicht gezeigt). Auch hier kann eine Probenkontamination oder Verwechslung nicht sicher ausgeschlossen werden.

Um einer möglichen Degradierung der Primer entgegenzuwirken, wurden in Lauf 8 (24 Proben) die Primer unverdünnt gepoolt und in die Multiplex-PCR eingesetzt. Alle Proben erhielten für jeden SNP das Match ‚Yes‘ mit einer sehr hohen Abdeckung. Bei diesem Lauf ist jedoch zu beachten, dass die Proben die jeweiligen SNP nicht

mehr mit der erforderlichen Tiefe von 150 abdecken (Daten nicht gezeigt) und insgesamt die Qualität der Daten im Vergleich zu anderen Läufen schlechter ist. Weiter ist zu sehen, dass bei allen Läufen die Mittelwerte der SNP ID Tiefen das gleiche Muster haben und gleichmäßig im Verlauf der Läufe sinken (ausgenommen Lauf 8).

3.3. Twist Human Sample ID Kit

Bei der Austestung des Human Sample ID Kits lag das Hauptaugenmerk sowohl beim α - als auch beim β - Kit neben der korrekten Probenidentifikation auf der optimalen Verdünnung des SNP-ID-Pools zum Einspielen in den Library-PhiX-Pool. Dies ist notwendig, da im Gegensatz zum SureDirect Blood PCR Kit die Primerverdünnungen nicht manuell angepasst werden können, und im Vergleich zum Originalprotokoll Anpassungen an die im Labor gegebenen Bedingungen erfolgen mussten. Zur Auswertung wurde das Sample Identity Tool der varvis® Plattform verwendet, das neben dem Vergleich der SNP (identity check) auch den Geschlechtsabgleich (Sex determination) und den Mix-up check beinhaltet (Erklärung siehe 1.2).

In diesem Kit sind die angepassten SNP (siehe Tabelle 3) enthalten. Es werden nur die Ergebnisse der im TWIST Custom Panel ID2020 oder Exom enthaltenden SNP gezeigt. Die Tabellen der Beispiele sind ebenso aufgebaut wie im Abschnitt 3.2.

Im Folgenden werden ebenfalls repräsentative Beispiele pro Testlauf (α - Kit) gezeigt. Bei allen Proben wird davon ausgegangen, dass sowohl der Geschlechtsabgleich als auch der Mix-up check die Anforderungen erfüllt haben.

Tabelle 8: Beispiel Probe 9 TWIST Human Sample ID Kit

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
9	rs497692	T	C	C	Yes	257	175	1,00	1,00
9	rs10203363	C	T	T	Yes	299	104	0,46	0,45
9	rs2819561	A	G	A	No	288	126	0,46	0,00
9	rs2942	G	A	A	Yes	258	38	0,40	0,57
9	rs17548783	T	C	C	Yes	317	108	1,00	1,00
9	rs4735258	T	C	C	Yes	351	29	1,00	1,00
9	rs1381532	A	G	G	Yes	267	89	1,00	1,00
9	rs10883099	G	A	A	Yes	327	232	1,00	1,00

9	rs7300444	C	C	C	Yes	318	72	0,00	0,00
9	rs9532292	A	G	G	Yes	246	143	0,49	0,41
9	rs4577050	G	A	A	Yes	321	62	0,49	0,47
9	rs2070203	G	A	A	Yes	385	319	0,48	0,50
9	rs1037256	G	A	A	Yes	328	35	0,47	0,51
9	rs9962023	T	C	C	Yes	262	35	1,00	1,00
9	rs2228611	T	T	T	Yes	334	69	0,00	0,00
9	rs10373	A	G	G	Yes	311	27	1,00	1,00
9	rs4148973	T	G	G	Yes	373	130	1,00	0,99
9	rs4675	T	T	T	Yes	371	137	0,00	0,00

Probe 9 zeigt in allen integrierten SNP das Match ‚Yes‘, außer im SNP auf Chromosom 3. Hier wird ein diskrepantes Ergebnis angezeigt, da im Sample call ein Guanin und im SNP ID call ein Adenin detektiert werden. Im Laufe der Testung des Kits konnte deutlich gemacht werden, dass dieser SNP beim Großteil aller Proben ausfällt, bzw. diskrepante Ergebnisse trotz guter Abdeckung entstehen. Nach Rücksprache mit der Firma TWIST Bioscience wurde der SNP aus der Wertung genommen (siehe auch Diskussion). Es bleiben also lediglich 17 SNP im TWIST Custom Panel ID2020 zur Auswertung. Da im Exom SNP auf den Geschlechtschromosomen integriert sind, können bei weiblichen Proben 17 SNP und bei männlichen 19 SNP bewertet werden. Da alle Parameter zur Sample Identity die Kriterien erfüllen und auch die Qualitätsparameter des Laufes sich im erforderlichen Rahmen befinden, kann davon ausgegangen werden, dass es nicht zu einer Probenverwechslung oder Kontamination gekommen ist.

Die Sample depth bei Probe 9 erreicht in allen SNP die Abdeckung von mindestens 150, sodass einer Auswertung nichts im Wege steht. Auch sind die Abdeckungen der unterschiedlichen SNP recht ausgeglichen. Die SNP ID Tiefe erreicht in allen SNP die erforderliche Abdeckung von 20, allerdings schwanken die Tiefen der einzelnen SNP. Die Mittelwerte der SNP ID Tiefen aller Proben des Laufes sind in Abbildung 12 gezeigt.

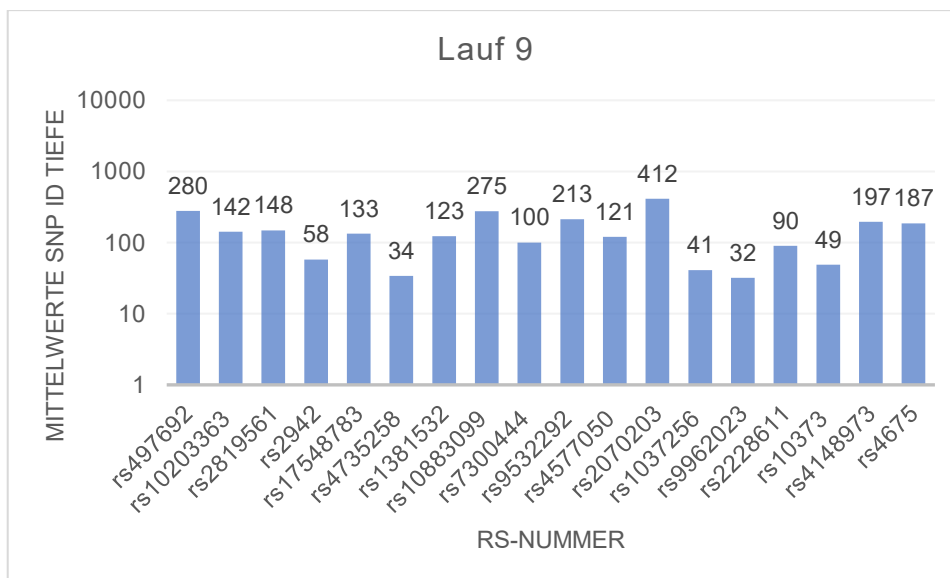


Abbildung 12: Mittelwerte der SNP ID Tiefe Lauf 9, logarithmische Darstellung

Zu sehen ist, dass alle SNP im Mittel die erforderliche Tiefe von > 20 erreichen, allerdings manche SNP (z.B. rs9962023 und rs4735258) nur knapp über der Grenze sind, während andere SNP sehr gut abgedeckt sind (z.B. rs2070203).

Bei Probe 9 handelt es sich um eine Probe in einem TWIST Custom Panel ID2020 Lauf von insgesamt 32 Proben. Die SNP ID-Library wurde mit einer Verdünnung von 1:80 in die Sample Library eingespielt. Da einige SNP nur sehr knapp die erforderliche Tiefe erreichen, wird für den nächsten Lauf mit 32 Proben eine Verdünnung von 1:50 gewählt, um zu verhindern, dass einzelne SNP ausfallen. Auch bei dieser geringeren Verdünnung sollten die Proben die erforderliche Abdeckung erreichen, da die in diesem Lauf gezeigte Abdeckung sehr gut ist.

Bei Probe 10 handelt es sich um eine männliche Probe aus einem Exomlauf. Der Exomlauf mit insgesamt acht Proben wurde 1:80 verdünnt.

Tabelle 9: Beispiel Probe 10 TWIST Human Sample ID Kit

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
10	rs497692	T	C	C	Yes	119	2191	0,52	0,49
10	rs10203363	C	T	T	Yes	127	1190	0,45	0,50
10	rs2819561	A		A		0	1542	0,00	0,00
10	rs2942	G	A	A	Yes	132	658	0,43	0,49
10	rs17548783	T	C	C	Yes	173	1023	0,46	0,50
10	rs4735258	T	C	C	Yes	50	600	0,50	0,47

10	rs1381532	A	G	G	Yes	86	1302	0,98	1,00
10	rs10883099	G	A	A	Yes	84	2870	0,46	0,52
10	rs7300444	C	T	T	Yes	113	671	0,47	0,47
10	rs9532292	A	G	G	Yes	134	1080	0,46	0,50
10	rs4577050	G	G	G	Yes	86	938	0,00	0,00
10	rs2070203	G	A	A	Yes	159	3722	0,46	0,51
10	rs1037256	G	G	G	Yes	193	591	0,00	0,00
10	rs9962023	T	C	C	Yes	121	462	0,56	0,49
10	rs2228611	T	C	C	Yes	59	714	0,61	0,47
10	rs10373	A	G	G	Yes	81	373	0,45	0,52
10	rs4148973	T	G	G	Yes	178	2101	1,00	1,00
10	rs4675	T	C	C	Yes	144	1677	0,50	0,49
10	rs2032653	G	G	G	Yes	85	218	0,00	0,00
10	rs34581739	G	G	G	Yes	99	284	0,00	0,00

Zu sehen ist, dass im Exom zwei SNP (rs2032653 und rs34581739) auf dem Y-Chromosom abdeckt werden. Alle integrierten SNP zeigen das Match ‚Yes‘, außer rs2819561, hier bleibt das Feld „Match“ leer. Der SNP wird bei der Bewertung nicht berücksichtigt.

Alle anderen Parameter wie Geschlechtsabgleich und Mix-up Check erfüllen die Kriterien, wodurch eine Probenverwechslung oder Kontamination ausgeschlossen werden kann. Die Sample depth ist deutlich geringer als im vorherigen Beispiel, dafür ist die SNP ID depth deutlich höher (siehe Diagramm Abbildung 13).

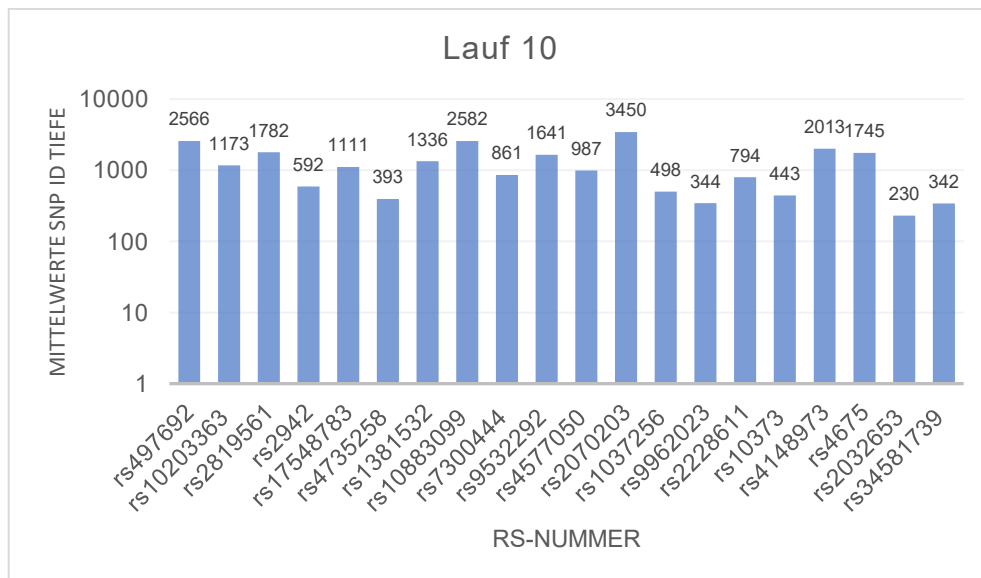


Abbildung 13: Mittelwerte der SNP ID Tiefe Lauf 10, logarithmische Darstellung

Die Mittelwerte aller Proben im Lauf der SNP ID Tiefen der einzelnen SNP sind deutlich höher als im Lauf 9 und erfüllen mehr als deutlich die Kriterien (> 20). Auch

die SNP, die in Lauf 9 weniger gut abgedeckt werden, sind hier sehr gut abgedeckt. Dies lässt darauf schließen, dass die Verdünnung von 1:80 zu gering für einen Lauf mit nur acht Proben ist und die SNP ID durchaus die Tiefe der Proben beeinflusst. Für die nächsten Exomläufe wird die SNP ID-Library 1:200 verdünnt und zum Sample Library-Pool gegeben.

Nach allen durchgeführten Testläufen (α - und β - Kit) ergeben sich folgende Verdünnungen der SNP ID-Library je nach Probenanzahl:

Tabelle 10: Verdünnung der SNP ID-Library je nach Probenanzahl

Probenanzahl	8	24	32	40	48	56	64	72
Verdünnung	1:200	1:80	1:50	1:30	1:30	1:30	1:10	1:5

3.3.1. Beispiel einer Probenverwechslung/Kontamination

Für die Überprüfung einer Probenverwechslung oder Kontamination müssen nach dem NGS-Lauf die Qualitätsparameter des Laufes und die Ergebnisse zur Sample Identity analysiert werden. Im Folgenden wird ein Beispiel einer Probenkontamination gezeigt.

- Mix-up check:

Similarity score

Probe 11 → Probe 12 0,068

Probe 12 → Probe 11 0,068

Der Similarity Score zwischen den beiden Proben von 0,068 ist geringer als ein Score für nicht-verwandte Proben sein sollte (Erklärung siehe 1.2). Somit besteht bei nicht bekannter Verwandtschaft der Verdacht auf eine Kontamination oder Probenverwechslung.

- Überprüfung des Geschlechtsabgleiches:

berechnetes Geschlecht *einggegebenes Geschlecht*

Probe 11: männlich männlich

Probe 12: weiblich weiblich

Der Geschlechtsabgleich zeigt eine Übereinstimmung zwischen berechnetem und eingegebenem Geschlecht.

- Identity check:

Probe 11 zeigt mehrere Allele mit Match ‚No‘ an:

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
11	rs17548783	T	C	C	No	157	62	0,69	1,00
11	rs4735258	T	C	T	No	167	30	0,15	0,00
11	rs4675	T	C	T	No	191	103	0,11	0,00

Zu sehen ist, dass bei SNP rs17548783 zwar die gleiche Base (Cytosin) detektiert wird, allerdings liegt beim NGS eine Heterozygotie vor und beim SNP ID Homozygotie. Bei SNP rs4735258 und rs4675 wird jeweils beim SNP ID ein Thymin, beim Sample call wird allerdings ein Cytosin sequenziert. Die Tiefen sind sowohl bei der Probe als auch beim SNP ID ausreichend, sodass die unterschiedliche Detektion nicht aufgrund einer zu geringen Abdeckung zustande kommen kann.

Der Verdacht einer Kontamination oder Probenverwechslung bleibt weiterhin bestehen.

In weiteren Untersuchungen zur Findung einer Ursache konnte gezeigt werden, dass die Proben auf der Platte nebeneinander liegen.

Zusammenfassung:

- zu ähnliche Proben (Mix-up check) bei nicht bekannter Verwandtschaft,
- Identity check zeigt Allele mit Match ‚No‘ an (bei ausreichender Abdeckung)

- Berechnetes und eingegebenes Geschlecht stimmen überein
- Möglichkeit einer Kontamination durch räumliche Nähe

Durch die Betrachtung aller Kriterien zeigt sich, dass es sich bei diesem Fall um eine Probenkontamination handelt.

Weiteres Vorgehen: da es sich um eine Probenkontamination und nicht um eine Probenverwechslung handelt, werden beide Proben erneut analysiert. Eine erneute DNA-Extraktion ist nicht zwingend notwendig und aufgrund der unterschiedlichen Geschlechter müssen beide Proben nicht in unterschiedlichen Runs laufen.

Zusammenfassend sind in Tabelle 11 die Kriterien zur Bewertung der SNP ID Ergebnisse im Zusammenhang mit den Ergebnissen des NGS dargestellt:

Tabelle 11: Übersicht Kriterien Probenverwechslung/Kontamination

	Probenverwechslung ausgeschlossen	Probenverwechslung/Kontamination
Sex determination	Eingegebenes + berechnetes Geschlecht stimmen überein	<ul style="list-style-type: none"> • Geschlechter stimmen überein • Geschlechter stimmen nicht überein
Identity check	<ul style="list-style-type: none"> • Alle Proben Match ‚Yes‘ (ausgenommen rs2819561) mit Abdeckung > 20x • Filtern ‚Empty‘: 14 SNP im Custom Panel, 12, bzw. 14 SNP bei Exom pro Patient*in (je nach Geschlecht) • Pro Patient*in 17 SNP 	Filter ‚No‘ zeigt unterschiedliche Ergebnisse bei Sample call + SNP ID call an (bei ausreichender Abdeckung)
Mix-up Check	Similarity Score > 0.080	Similarity Score < 0.080 + keine Verwandtschaft bekannt

3.3.2. Beispiel eines unklaren Ergebnisses

Bei der Überprüfung der Sample Identity nach einem NGS-Lauf werden folgende Ergebnisse erzielt:

- Geschlechtsabgleich bei allen Proben korrekt
- Mix-up check bei allen Proben > 0,080

- Identity check zeigt bei Probe 13 im SNP rs4735258 Match ‚No‘ bei ausreichender Abdeckung an:

Probe	rs-Nummer	Genomic ref	Sample call	SNP ID call	Match	Sample depth	SNP ID depth	Sample altAF	SNP ID altAF
13	rs4735258	T	C	T	No	135	108	0,53	0

Im NGS wird ein Cytosin (heterozygot) detektiert, wohingegen im SNP ID ein Thymin homozygot detektiert wird

Da bereits ein Match ‚No‘ bei ausreichender Abdeckung eine Probenverwechslung oder Kontamination bedeuten könnte, werden hier weitere Untersuchungen angeschlossen.

Mit dem IGV-Viewer werden die einzelnen Reads des SNP im NGS und SNP ID visualisiert:

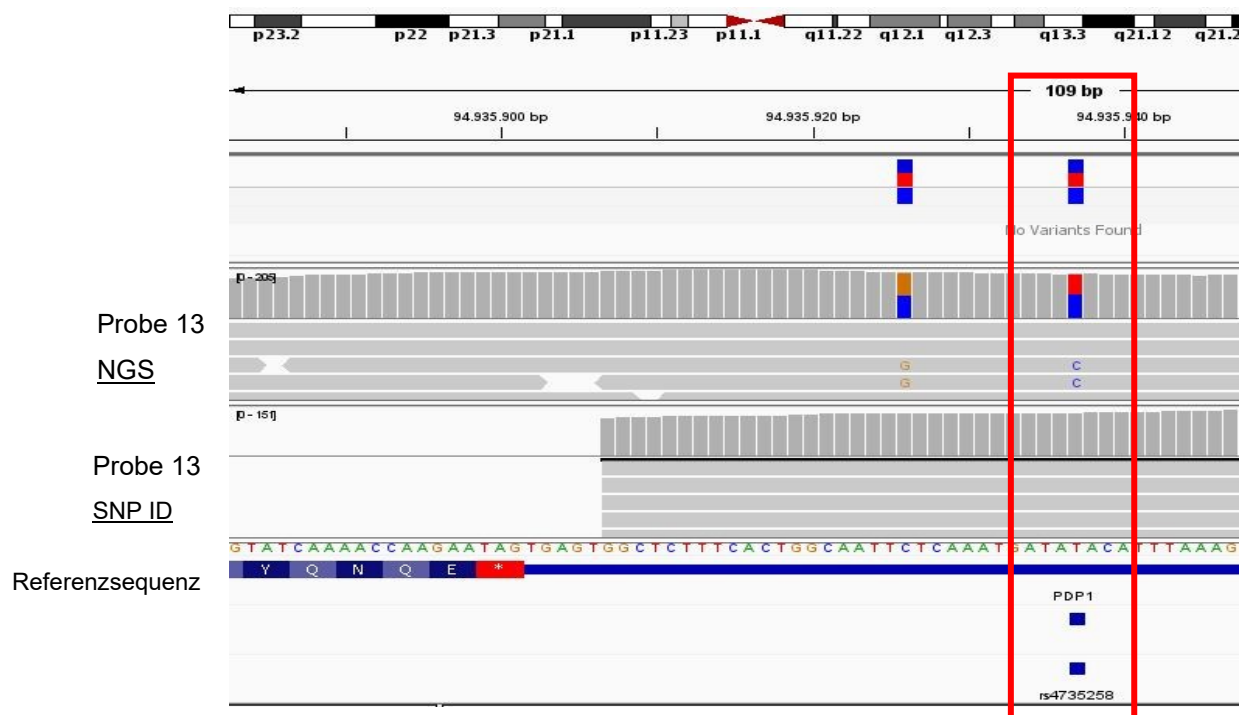
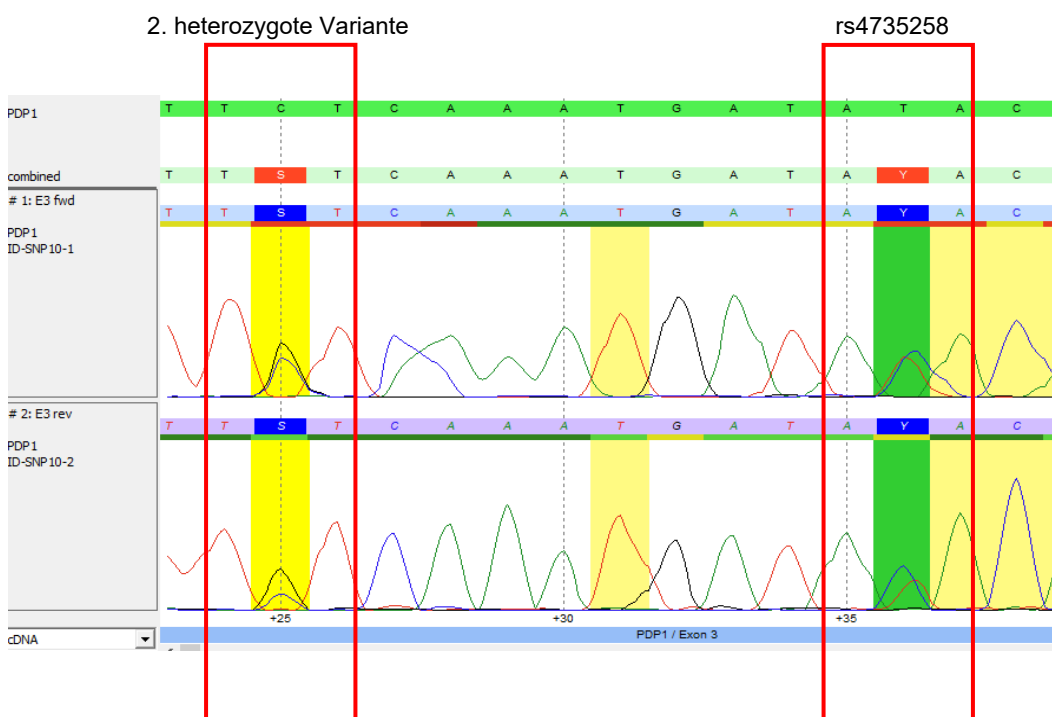


Abbildung 14: Rohdaten IGV-Viewer Probe 13

In Abbildung 14 ist das Chromosom 8, auf dem der SNP rs4735258 liegt, dargestellt. Angegeben ist auch die genomische Lokalisation des Bereichs auf dem Chromosom sowie die Reads und Abdeckung im Bereich des SNP im NGS und im SNP ID. Zu sehen ist, dass der SNP im SNP ID gut abgedeckt wird, allerdings

hauptsächlich durch forward Reads. Es wird ein Thymin homozygot detektiert. Im NGS wird der SNP sowohl von forward als auch reverse Reads ausreichend abgedeckt, eine Heterozygotie für CT wird detektiert. In unmittelbarer Nähe (11 bp entfernt) wird im NGS eine weitere heterozygote Variante nachgewiesen, welche im SNP ID nicht angezeigt wird. Zur Überprüfung wird eine Sanger-Sequenzierung des SNP aus einer erneuten DNA-Extraktion durchgeführt.

- Ergebnis Sanger-Sequenzierung SNP rs4735258:



Zu sehen ist, dass der SNP in heterozygoter Form vorliegt. Auch die zweite heterozygote Variante kann hier detektiert werden. Das Ergebnis des SNP entspricht also dem des NGS, somit liegt keine Probenverwechslung bei diesem Lauf vor.

4. Diskussion

Überall, wo Menschen manuell arbeiten, passieren Fehler. Gerade bei Hochdurchsatzmethoden wie dem NGS, bei dem viele Proben auf einmal bearbeitet

und analysiert werden, steigt die Zahl an Fehlern und führt damit zu einem erhöhten Risiko für Probenverwechslung oder Kontamination der Proben. Um den Patient*innen eine sichere Analyse und damit ein sicheres Ergebnis zu gewährleisten, aber auch um die Labore abzusichern, ist es wichtig, ein Verfahren zu finden, welches mögliche Verwechslungen erkennt. Dieser Test sollte so stattfinden, dass Verwechslungen oder Kontaminationen erkannt und behoben werden, bevor ein Befund an Einsender*in oder Patient*in gelangt. Zusätzlich wichtig ist, dass der Test QM-konform ist und den aktuell gültigen Richtlinien folgt.

Bei der Austestung der verschiedenen Methoden im Labor war es ebenfalls wichtig, dass die Methode möglichst einfach und robust und in der Routinediagnostik neben den weiteren Methoden und Analysen gut durchführbar ist. Auch sollte der Kostenpunkt beachtet werden.

Alle bisher durchgeführten Methoden zur Probenidentifikation, wie beispielweise die AZF-Locusanalyse oder die Testung der STR-Marker, mussten zusätzlich zur Routinediagnostik durchgeführt werden. Auch die Auswertung ist aufwendiger, da die STR-Marker mit einer anderen Software ausgewertet werden als die NGS-Daten. Darüber hinaus ist die AZF-Locusanalyse limitiert, weil hier nur das Geschlecht überprüft werden kann.

Durch die von Pengelly et.al. beschriebenen SNP wurde eine neue Möglichkeit geschaffen, Proben voneinander zu unterscheiden. Mit dem Abgleich per Sanger-Sequenzierung können diese gut nachgewiesen werden, allerdings ist die Analyse sehr aufwendig, da jeder SNP einzeln bearbeitet werden muss. Bei einem Probenaufkommen von ca. 200-300 NGS-Proben pro Monat wie es im MVZ Humane Genetik in München der Fall ist, wäre dieser Weg zu ineffizient und damit nicht umsetzbar. Eine Reduktion der Anzahl an SNP wäre gleichermaßen keine Option, da die Sensitivität dadurch sinken würde. Mit der Sanger-Sequenzierung kann aktuell kein Geschlechtsabgleich durchgeführt werden, da keine SNP auf dem X- oder Y-Chromosom analysiert werden. Erschwert wird die Auswertung dadurch, dass kein automatischer Vergleich der Ergebnisse stattfinden kann, da diese Analysen mit verschiedenen Software ausgewertet werden. Somit muss für die Auswertung wie auch für die Durchführung der Methode zusätzliche Zeit aufgewendet werden.

Bei der Sanger-Sequenzierung fällt auf, dass einige SNP nur einzelsträngig abgedeckt werden. Dies ist sehr wahrscheinlich die Folge eines nicht ausreichend optimierten Primerdesigns. Für eine optimale Abdeckung der SNP sollten die Primer neu designt werden, damit die SNP sowohl in forward- als auch in reverse-Leserichtung abdeckt werden.

Da die Probenidentifikation im MVZ Humane Genetik München nicht routinemäßig mittels Sanger-Sequenzierung durchgeführt wird, wird auf ein Neudesign der Primer verzichtet. Die Primer werden zwar zur Nach-Sequenzierung von ausgefallenen Proben oder unklaren Ergebnissen (siehe weitere Diskussion) verwendet, hierfür sind die vorhandenen Primer und deren Abdeckungen der SNP aber ausreichend.

Ein weiterer Nachteil der Sanger-Sequenzierung zur Probenidentifikation ist, dass diese nur aus genomischer DNA erfolgen kann. Dadurch werden Fehler, die vor oder bei der DNA-Extraktion passieren, nicht erfasst. Fehler bei der Extraktion werden zwar bereits durch geeignete Verfahren (siehe Einleitung) minimiert, können allerdings trotzdem nicht mit letzter Sicherheit ausgeschlossen werden. Alternativ könnte eine erneute Extraktion der DNA aus EDTA-Blut erfolgen, um die SNP aus einer zweiten DNA durchzuführen, allerdings würde dies erneute Arbeit, Aufwand und Kosten bedeuten. Zusätzlich müssten weitere Lagerungsmöglichkeiten für die doppelten DNA-Röhrchen geschaffen werden.

Ziel dieser Arbeit ist, eine Methode zu implementieren, die die Analyse der SNP aus EDTA-Blut erlaubt und diese mit NGS-Daten vergleicht. Die Analyse aus Blut schließt Fehler, die nach dem Probeneingang und vor der Durchführung der NGS-Analyse stattfinden, aus. Zusätzlich soll der Aufwand der SNP-Analyse, bezogen auf Durchführbarkeit und Auswertung, gering sein. Die Analyse aus Blut wird sowohl beim SureDirect Blood PCR Kit als auch beim TWIST Human Sample ID Kit ermöglicht. Bei beiden Analysen ist der zusätzliche Arbeitsaufwand geringer und einfacher, da die Protokolle nahezu ident zu denen des bereits etablierten NGS Workflow sind. Ein großer Vorteil beider Analysen liegt in der Auswertung. Durch die Möglichkeit des Einspiens der SNP ID-Library in die Proben-Library kann eine Gegenüberstellung der SNP jeder Probe aus Blut und DNA durch die Plattform varvis® erfolgen, was den Arbeitsaufwand deutlich senkt. Das Einspien der SNP-ID-Library ist, auch wenn dies nur in sehr geringen Volumina geschieht, allerdings

abhängig von der Auslastung der Flowcell. Beim aktuellen TWIST Custom Panel ID2020 wäre die MidOutput-Flowcell des NextSeq 550 bei 192 Proben ausgelastet, wenn alle Proben in mindestens 98 % 20 x abgedeckt werden sollen. Da maximal 72 Proben auf einmal sequenziert werden, ist das Einspiken möglich. Mit der HighOutput-Flowcell können maximal acht Exom-Proben sequenziert werden, um die Mindestcoverage zu erreichen.

Die Überlastung der Flowcell, bzw. das Risiko, dass Proben nicht mindestens 20x abgedeckt werden, muss trotzdem verhindert oder minimiert werden. Dies wird beim SureDirect Blood PCR Kit durch die optimale Primer-Verdünnung und beim TWIST Human Sample ID Kit durch die Verdünnung des fertigen SNP ID-Library-Pools in die Proben-Library gewährleistet.

4.1. SureDirect Blood PCR Kit

Im Rahmen der Austestung der Methode mit dem SureDirect Blood PCR Kit wurden zu Beginn gute Ergebnisse erzielt und die Proben konnten größtenteils gut voneinander diskriminiert werden, wodurch eine Probenverwechslung in den meisten Fällen ausgeschlossen werden konnte. Die Anpassungen der Primerverdünnungen anhand der erzielten Ergebnisse ist ein Vorteil der Methode, da dadurch eine Überladung der Flowcell verhindert werden kann. Ein weiterer Vorteil ist, dass die Primer manuell angepasst werden können. Dies bedeutet, dass nicht nur die Konzentration verändert werden kann, sondern auch Primer hinzugefügt oder verändert werden können, um weitere SNP abzudecken oder die Lage der Primer zu verändern.

Im Laufe der Austestung hat sich gezeigt, dass die Tiefen der SNP pro Probe sehr schwanken. So werden die SNP mancher Proben gut abgedeckt, wohingegen andere weniger gut abgedeckt werden und somit die Auswertung erschwert wird oder nicht möglich ist. Um dies zu verhindern, wurde jede Probe im weiteren Testverlauf auf 80 ng mit RSB-Puffer verdünnt. Diese Anpassung zeigte jedoch nur geringe Verbesserungen in den Ergebnissen, was sehr wahrscheinlich an der Stabilität der Primer liegt. In den weiteren Läufen wurden die SNP teilweise sehr wenig bis gar nicht abgedeckt, sodass die für eine eindeutige Probenidentifikation notwendige Anzahl an auswertbaren SNP unterschritten wurde. Durch zahlreiche

Tests wurden die verschiedenen Komponenten des Kits getestet und auch an verschiedenen Zeitpunkten der Analyse Qualitätskontrollen eingefügt. Um das SureDirect Blood PCR Kit zu testen, wurde das PCR-Produkt nach der Multiplex-PCR auf ein 2 % Agarose-Gel aufgetragen, was zeigen sollte, dass bei der PCR überhaupt ein Produkt in der richtigen Basenpaargröße amplifiziert wird. Zusätzlich wurde nach der Multiplex-PCR eine Qubit-Messung durchgeführt. Nach der Library Preparation wurde neben der Qubit-Messung zur Normalisierung auf 80 ng eine TapeStation-Messung angeschlossen. Auch hier sollte kontrolliert werden, ob die amplifizierten Fragmente die richtige Größe besitzen. Alle Tests zeigten die zu erwarteten Ergebnisse, sodass rückgeschlossen werden konnte, dass die Multiplex-PCR und Library Preparation erfolgreich ablaufen. Als letzten Schritt wurden die Primer unverdünnt gepoolt und in die Multiplex-PCR eingesetzt. Die Ergebnisse der Qualitätskontrollen nach der Multiplex-PCR und nach der Library Preparation zeigten die zu erwarteten Ergebnisse. Die Ergebnisse nach der Sequenzierung (Lauf 8) zeigen, dass mit den unverdünnten Primern sehr hohe Abdeckungen erreicht werden. Dies lässt darauf schließen, dass die Primer zwar funktionieren, jedoch im verdünnten Zustand zu instabil sind. Es wäre möglich, die Primer vor jedem Lauf neu zu verdünnen, um der vermuteten schnellen Degradation entgegenzuwirken und eine ausreichende Abdeckung zur Probenidentifikation zu erlangen, allerdings würde dies einen zusätzlichen hohen Aufwand bedeuten. Ein weiteres Problem wäre, dass durch die sehr hohe Abdeckung der SNP die Abdeckung der Patient*innen-Proben beeinflusst wird und diese nicht mehr mit der gewünschten Coverage abgedeckt werden könnten. Die Probenidentifikation würde so mit einem Qualitätsverlust der Patienten-Proben einhergehen, was nicht wünschenswert ist. Es wäre weiters möglich gewesen, die gepoolte SNP-ID-Library höher als 1:1000 zu verdünnen, um dies zu umgehen. Denkbar wäre auch gewesen, abhängig von der Probenmenge im Lauf, unterschiedliche Verdünnungen zu testen, um ausgeglichene Ergebnisse zu erzielen, ohne die Probenqualität zu beeinflussen. Diese Tests blieben aufgrund der Instabilität der Primer und der Möglichkeit, das α - und β - Kit des TWIST Human Sample ID Kit zu testen, aus.

4.2. TWIST Human Sample ID Kit

Wie auch die Methode mit dem SureDirect Blood PCR Kit lässt sich die Multiplex-PCR beim TWIST Human Sample ID Kit aus EDTA-Blut durchführen, was einen Vorteil darstellt. Weiters ist ebenso keine zusätzliche Ausstattung notwendig, da die Analyse mit nahezu gleichen Reagenzien und Geräten abläuft wie der NGS Workflow. Zudem folgt es bis auf die Multiplex-PCR dem normalen NGS-Protokoll, was einen einheitlichen Workflow, leichtere Einarbeitung und Integration in die Routinediagnostik verspricht. Ein weiterer großer Vorteil liegt ebenfalls in der einfachen Auswertung, da die gleiche Art Barcodes für die SNP-ID-Analyse und die NGS-Analyse verwendet werden. Auch hier erfolgt die Auswertung mit der Plattform varvis®. Insgesamt ist die Analyse sehr robust und es werden zusätzliche Sequenzier- oder Analysekosten durch das Einspielen in die Proben-Library gespart. Die Kosten pro Probe belaufen sich auf ca. 30 €, was im Vergleich zu anderen, ebenfalls bereits kommerziell erhältlichen Kits zur Probenidentifikation zwar etwas teurer ist, aber den Vorteil gegenüber diesen Methoden hat, dass die Analyse aus EDTA-Blut durchführbar ist. Nicht immer liegt allerdings EDTA-Blut als Primärmaterial der Proben vor. Häufig werden vor allem bei pränatalen Trio-Analysen Gewebe- oder Fruchtwasserproben zur Analyse eingesandt. Auch kann es vorkommen, dass die Analyse erst nach der vorgeschriebenen Mindestaufbewahrungszeit von acht Wochen gestartet wird und somit das EDTA-Blut in der Zwischenzeit verworfen wurde. In solchen Fällen werden dann 10 ng der extrahierten DNA in die Multiplex-PCR eingesetzt. Auch aus DNA zeigt die Analyse mit dem TWIST Human Sample ID Kit gute Ergebnisse, im Gegensatz zur Analyse mit dem SureDirect Blood PCR Kit.

Ein großer Nachteil des Kits ist, dass die Primer-Sequenzen der SNP nicht bekannt sind. Somit liegt kein Wissen über die Sequenz und Position der Primer vor und es kann keine Aussage darüber getroffen werden, warum weiterhin manche SNP höhere Abdeckungen erreichen als andere. Dies erschwert das Einstellen der Verdünnungen, da manche SNP sehr viel schwächer abgedeckt werden als andere. Möglich ist, dass manche Primer oder SNP in schwierig zu sequenzierenden Regionen liegen, wodurch geringere Abdeckungen erreicht werden. Zusätzlich ist

keine manuelle, individuelle Anpassung der SNP möglich, da ein fertiger Primermix verwendet wird.

Eine Einschränkung in der Auswertung stellt aktuell die fehlende Integration mancher SNP in das verwendete Custom Panel und Exom dar. Durch den Ausfall des SNP auf Chromosom 3 (rs2819561) stehen lediglich 17, bzw. 19 SNP bei männlichen Exomproben zur Verfügung. Nach Pengelly et. al. sind dies zu wenig SNP, um zwei Individuen sicher voneinander zu unterscheiden, da die Wahrscheinlichkeit, dass zwei Individuen die gleichen Allelkonstellationen haben, mit steigender Anzahl an SNP sinkt. Durch den zusätzlichen Mix-up check und die Geschlechtsbestimmung wird die geringere Anzahl an SNP in Kauf genommen, zumal die Umstellung auf eine neue Version des Exom-Kits, welches alle SNP enthält, die im TWIST Human Sample ID Kit integriert sind, bereits im Gange ist.

Der SNP rs2819561 zeigte von Anfang an diskrepante Ergebnisse, obwohl die Abdeckungen gut sind. Bei genauerem Blick auf die Rohdaten in der IGV Software ist bei allen Proben ersichtlich, dass das Amplikon im SNP ID sehr unsauber sequenziert wird und auch weitere Varianten im Amplikon sehr viel Hintergrund zeigen. Im Vergleich dazu ist ein solcher Hintergrund bei anderen Amplikons nicht zu erkennen. Sehr wahrscheinlich wird der SNP gar nicht gecallt, weil dieser aufgrund der schlechten Qualität durch die Plattform varvis® aussortiert wird. Nicht vorhandene Varianten werden in der Auswertung als Referenz, in diesem Fall Adenin, angegeben. Hiermit erklärt sich auch, wieso der SNP bei manchen Proben übereinstimmende Ergebnisse bei NGS und SNP ID zeigt, nämlich dann, wenn auch im NGS ein Adenin detektiert wird. Durch die nicht vorhandenen Angaben zur Primersequenz und -position kann dieses Problem nicht vollständig gelöst werden. Bei Anpassungen des Kits soll der Fehler aber von TWIST Bioscience überarbeitet werden.

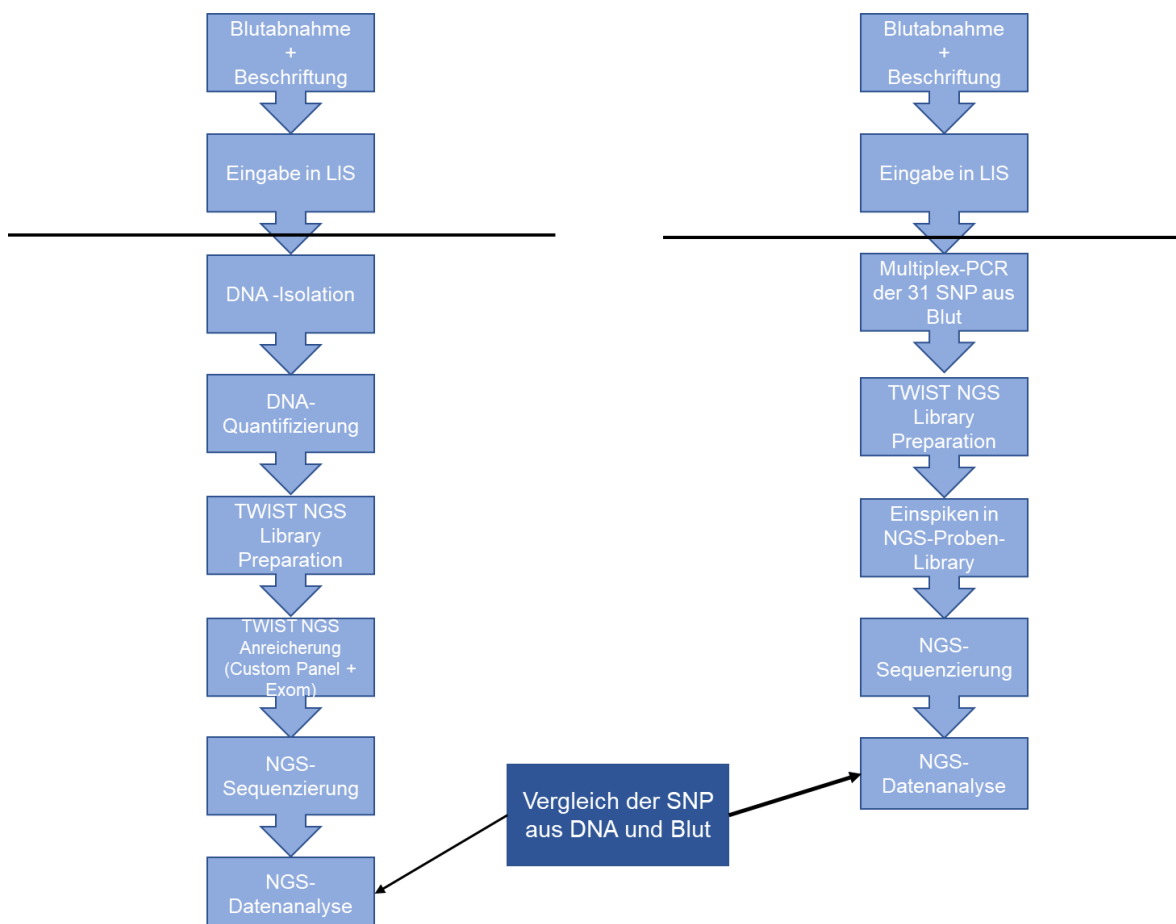
Beim TWIST Human Sample ID Kit werden alle Proben des SNP ID auf 4,3 nM mit RSB verdünnt und gepoolt. Diese Änderung im Vergleich zum Originalprotokoll wird aufgrund der gegebenen Bedingungen im Labor durchgeführt. Die anschließende Verdünnung ist abhängig von der Laufgröße und verhindert das Überladen der Flowcell und somit einen Qualitätsverlust der Proben durch Verringerung der Abdeckung. Die Verdünnung muss hierbei so gewählt werden, dass alle SNP

ausreichend abgedeckt werden und eine Balance von Proben und SNP ID Tiefe besteht. Da die Flowcells meistens nicht voll ausgelastet sind, wird eine etwas geringere Verdünnung gewählt, um zu verhindern, dass SNP ausfallen oder aufgrund einer zu geringen Abdeckung mittels Sanger-Sequenzierung wiederholt werden müssen.

Trotz der Verdünnung und der Einstellung auf 4,3 nM kommt es in der Routinediagnostik vor, dass SNP-Proben ausfallen oder aufgrund der zu geringen Abdeckung Match ‚No‘ anzeigen. Häufig ist dies auf das Ausgangsmaterial zurückzuführen, da die Analyse aus Blut schwieriger ist. Eine erneute Sequenzierung mittels NGS ist in der Routinediagnostik nicht durchführbar, weshalb diese ausgefallenen SNP nachträglich mittels Sanger-Sequenzierung analysiert werden müssen, was einen erhöhten Aufwand bedeutet. Auch bei unklaren Ergebnissen (siehe 3.3.2) wird eine Sanger-Analyse angeschlossen. Bei diesem Beispiel wurden die Kriterien zur Probenidentifikation (Mix-up check und Geschlechtsbestimmung) erfüllt, jedoch zeigte ein SNP bei ausreichender Abdeckung Match ‚No‘ an. In den Rohdaten konnte eine weitere heterozygote Variante nachgewiesen werden, die in den BED-files des SNP ID nicht angezeigt wurde. Durch diese zweite Variante kommt es sehr wahrscheinlich zu einem Alleldropout und dem scheinbar homozygoten Ergebnis im Identity check. Aufgrund der fehlenden Informationen zu den Sonden oder Primern im TWIST Human Sample ID Kit kann nicht sicher ausgeschlossen werden, dass die Sonde dort bindet oder die Bindung durch die zweite Variante beeinträchtigt wird. In der Sanger-Sequenzierung aus einer zweiten DNA-Extraktion, die den SNP und angrenzende Bereiche umfasst, wurde sowohl die zweite Variante als auch der SNP in Heterozygotie nachgewiesen. Das diskrepante Ergebnis kann somit vermutlich durch einen Alleldropout erklärt werden. Dieses Beispiel zeigt, dass es durch andere Varianten oder SNP in der Nähe zur Beeinflussung der Detektion kommen kann, wodurch eine genauere Analyse bei ausgefallenen SNP oder diskrepanten Ergebnissen notwendig ist, auch wenn dies erhöhten Aufwand bedeutet.

5. Fazit

In folgender Abbildung wird der bereits in der Einleitung gezeigte Workflow der Proben im Labor gezeigt. Dem gegenüber wird der Workflow der Probenidentifikation mittels SNP ID dargestellt.



Unterhalb der schwarzen Linien werden die Schritte, die im Labor passieren, dargestellt. Ab Eingang der Probe ins Labor kann mittels der SNP-ID-Analyse eine Probenverwechslung oder Kontamination festgestellt werden. Mögliche Fehler, die vorher passieren, beispielsweise ein falsches Etikettieren, können nicht erkannt werden, außer wenn das Geschlecht in der Analyse nicht dem angegebenen Geschlecht entspricht. In diesen Fällen muss immer Rücksprache mit den Einsender*innen gehalten werden, um auch beispielsweise Gonosomenstörungen auszuschließen.

Die Analyse der SNP zur Probenidentifikation allein reicht allerdings nicht aus, um eine Probe von anderen zu diskriminieren. Wie in Tabelle 11 gezeigt, müssen zusätzlich auch die weiteren Parameter wie Geschlechtsabgleich und Mix-up check in die Bewertung einfließen, damit eine Kontamination oder Verwechslung ausgeschlossen werden kann.

Fehler können während der Durchführung von Hochdurchsatzsequenzierungen, aber auch bei allen anderen Analysen, jederzeit auftreten. Mit dem TWIST Human Sample ID Kit wurde aber eine gute Lösung gefunden, um die Fehler zu identifizieren, bevor ein Schaden für die Patient*innen entsteht, auch wenn noch einige Details angepasst werden müssen.

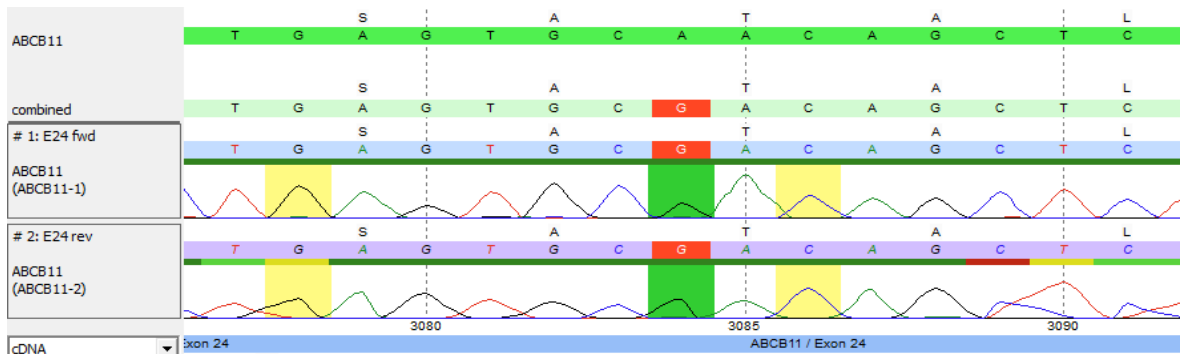
Dem erhöhten Aufwand und den zusätzlichen Kosten durch die hinzukommende Analyse mit dem TWIST Human Sample ID Kit steht ein sehr hoher Qualitätsstandard gegenüber, der den Patient*innen und dem Personal hohe Sicherheit bietet. Trotz vieler Versuche ist es nicht gelungen, den SureDirect Blood PCR Kit so zu optimieren, dass robuste und verlässliche Ergebnisse entstanden sind.

6. Anhang

Die Ergebnisse sind analog zu den Ergebnissen im Ergebnisteil dargestellt.

Probe A

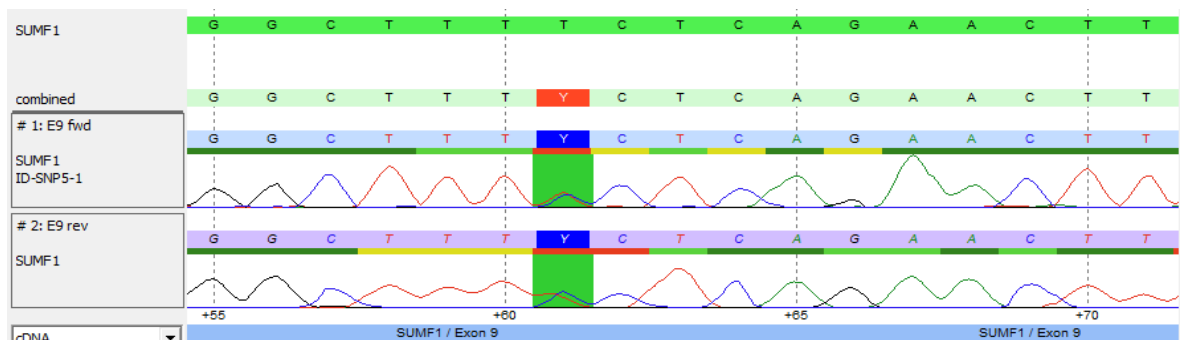
SNP rs497692 (*ABCB11* Exon 24 c.3084)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
169789016	rs497692	<i>ABCB11</i>	Exon 24 c.3084	T	C	0,9940828	homozygot

In der Sanger-Sequenz wird bei diesem SNP der Minus-Strang sequenziert, sodass die komplementäre Base als Referenz gilt. In diesem Fall ist dies ein Adenin, im NGS ein Thymin. Der AltAF von 0,9940828 spricht für ein homozygotes Ergebnis, da in 99 % das alternative Allel detektiert wird. Beide Methoden zeigen demnach ein homozygotes Ergebnis (G, bzw. C) und stimmen überein. Zudem zeigt die Sanger-Sequenz eine Abdeckung sowohl im Forward als auch im Reverse-Strang an der Position des SNP.

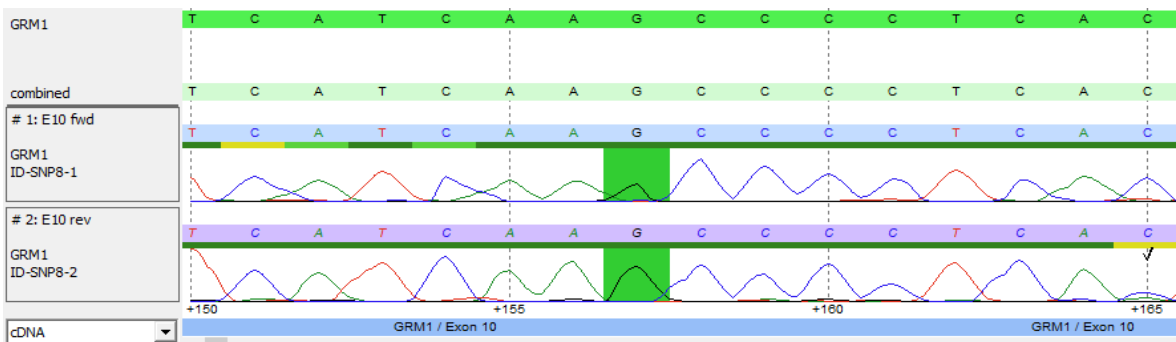
SNP rs2819561 (*SUMF1* Ex9/3'UTR c.*61)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
4403767	rs2819561	<i>SUMF1</i>	Ex9/3'UTR c.*61	A	-	no call	-

Die Sanger-Sequenzierung des Minus-Strangs zeigt ein eindeutig heterozygotes Ergebnis, sowohl in Forward- als auch in Reverse-Leserichtung. Im NGS konnte keine Base detektiert werden, sodass ein Vergleich der beiden Methoden für diesen SNP nicht möglich ist.

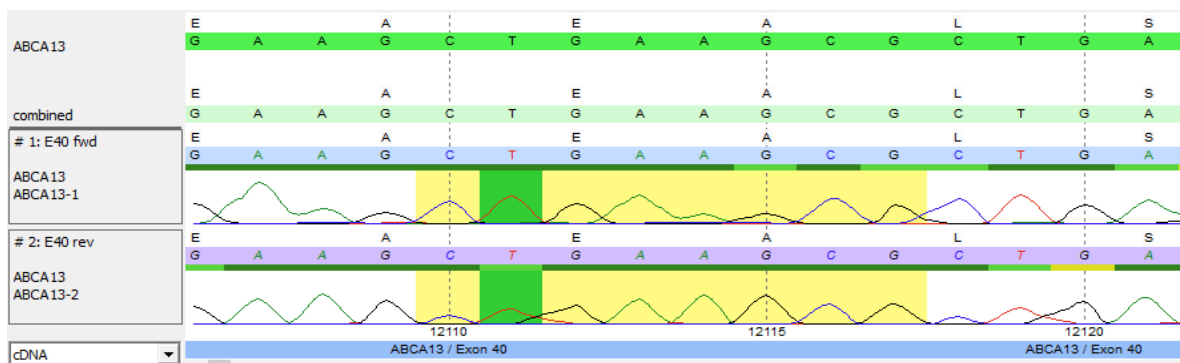
SNP rs2942 (*GRM1* Ex10/3'UTR c.*157)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
146755140	rs2942	<i>GRM1</i>	Ex10/3'UTR c.*157	G	G	0	Ref

Der SNP rs2942 zeigt in der Sanger-Sequenz ein Guanin, was der Referenz-Base entspricht. Auch im NGS wurde dieses Ergebnis mit einer AltAF von 0 erzielt. Die Ergebnisse stimmen demnach überein.

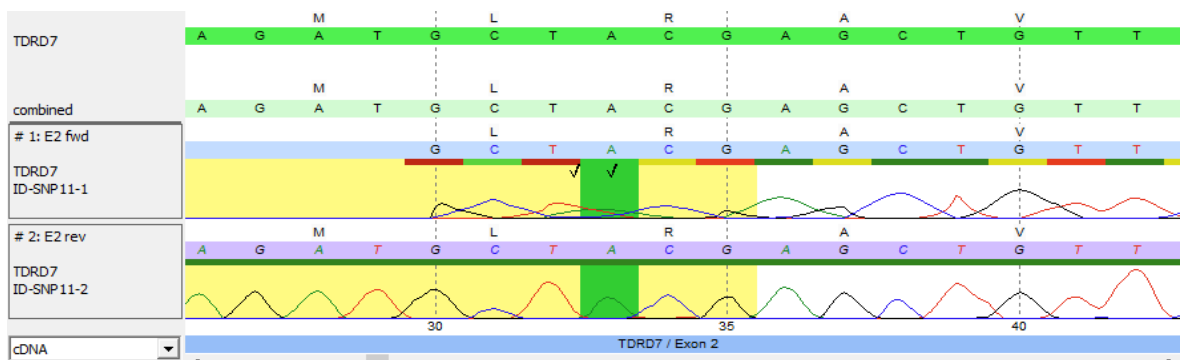
SNP rs17548783 (*ABCA13* Exon 40 c.12111)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
48450157	rs17548783	<i>ABCA13</i>	Exon 40 c.12111	T	T	0	Ref

Auch dieser SNP wird in der Sanger-Sequenz in beide Leserichtungen gut abgedeckt und zeigt in beiden Methoden ein homozygoten Ergebnis für die Referenz.

SNP rs1381532 (*TDRD7* Exon 2 c.33)

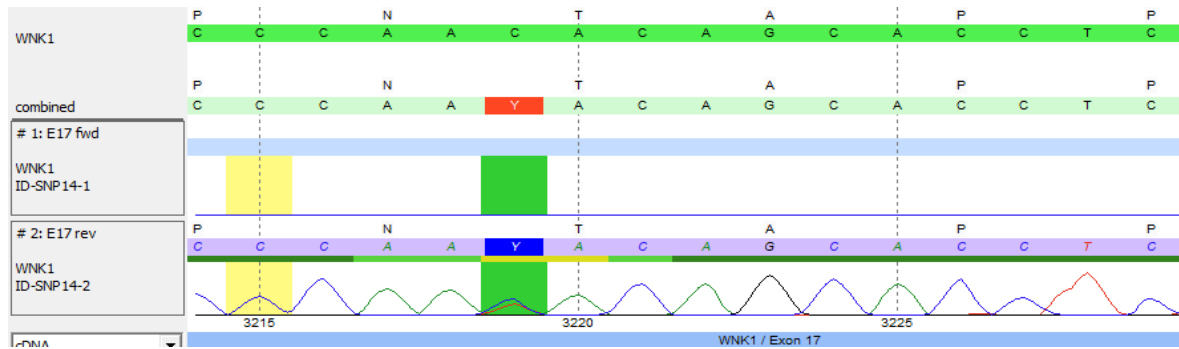


Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
100190780	rs1381532	<i>TDRD7</i>	Exon 2 c.33	A	A	0	Ref

Der SNP in Exon 2 des *TDRD7*-Gens wird in der Sanger-Sequenz eindeutig nur von der Sequenz des reverse-Primers abgedeckt. Der forward Primer fängt erst kurz vor dem SNP an, sodass das Sequenz-Profil noch unruhig ist. Trotzdem wird ein

Adenin, was der Referenzbase entspricht, klar detektiert. Auch im NGS wird die Referenzbase nachgewiesen.

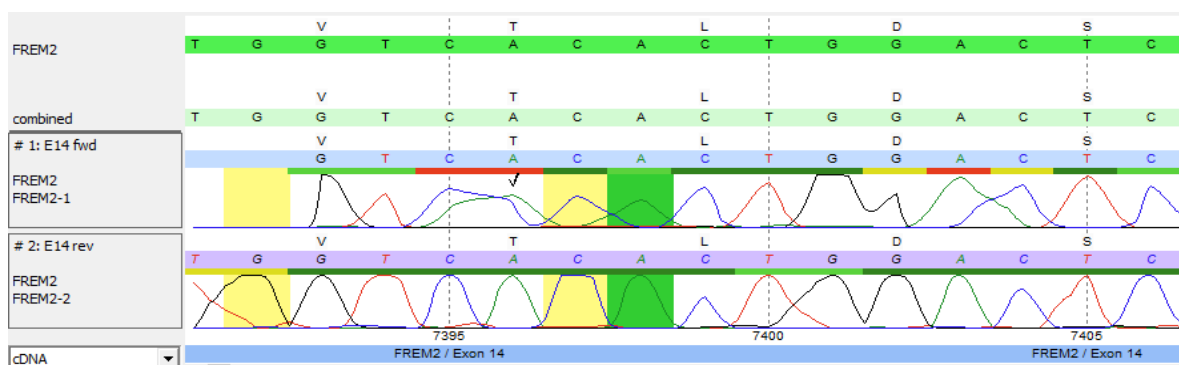
SNP rs7300444 (*WNK1* Exon 17 c.3219)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
993930	rs7300444	<i>WNK1</i>	Exon 17 c.3219	C	T	0,5047619	heterozygot

Der SNP rs7300444 ist in der Sanger-Sequenz nur in der Reverse-Leserichtung auswertbar, in der Forward-Leserichtung nicht. Trotzdem ist ein klares heterozygoten Ergebnis zu erkennen, welches auch in der NGS-Analyse bestätigt wird.

SNP rs9532292 (*FREM2* Exon 14 c.7398)

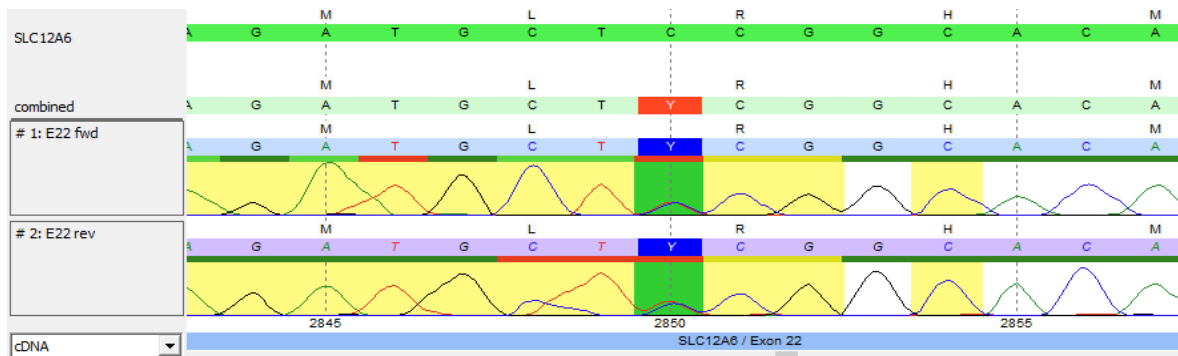


Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie

39433606	rs9532292	FREM2	Exon 14 c.7398	A	A	0	Ref
----------	-----------	-------	-------------------	---	---	---	-----

Eine klare Übereinstimmung wird für den SNP rs9532292 sowohl mit der Sanger-Sequenzierung als auch mit dem NGS erzielt.

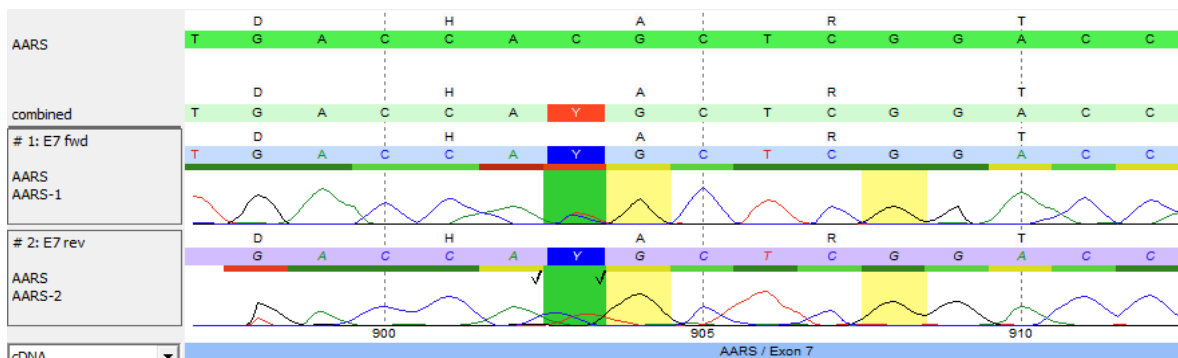
SNP rs4577050 (SLC12A6 Exon 25 c.2850)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
34528948	rs4577050	SLC12A6	Exon 25 c.2850	G	A	0,5654205	heterozygot

Die unterschiedlichen Exonbezeichnungen kommen bei diesem SNP durch die unterschiedlichen Transkripte zustande. Die Sanger-Sequenz zeigt sowohl im forward als auch im reversen Strang ein heterozygoten Ergebnis, was im NGS mit einer AltAF von 0,5654205 ebenfalls detektiert wird.

SNP rs2070203 (AARS Exon 7 c.903)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
70303580	rs2070203	AARS	Exon 7 c.903	G	A	0,4893920	heterozygot

Der SNP rs2070203 wird sowohl in der Sanger-Sequenz als auch im NGS in einer heterozygoten Konstellation detektiert. Auch hier erfolgt die Sanger-Sequenzierung am Minusstrang.

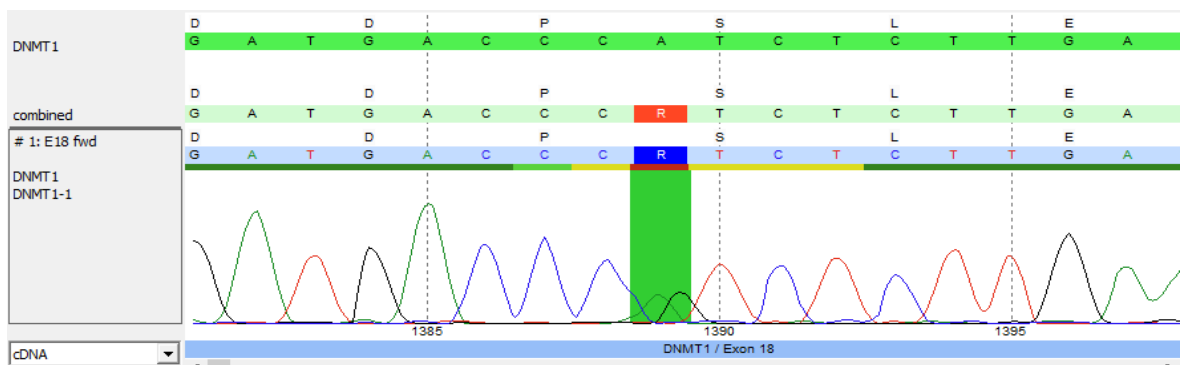
SNP rs1037256 (COG1 Exon 7 c.1782)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
71197748	rs1037256	COG1	Exon 7 c.1782	G	A	1	homozygot

Der SNP rs1037256 ist in der Sanger-Sequenz lediglich in der Forward-Leserichtung auswertbar. Trotzdem ist das homozygote Ergebnis für die Base Adenin eindeutig und wird auch im NGS (AltAF 1) erzielt.

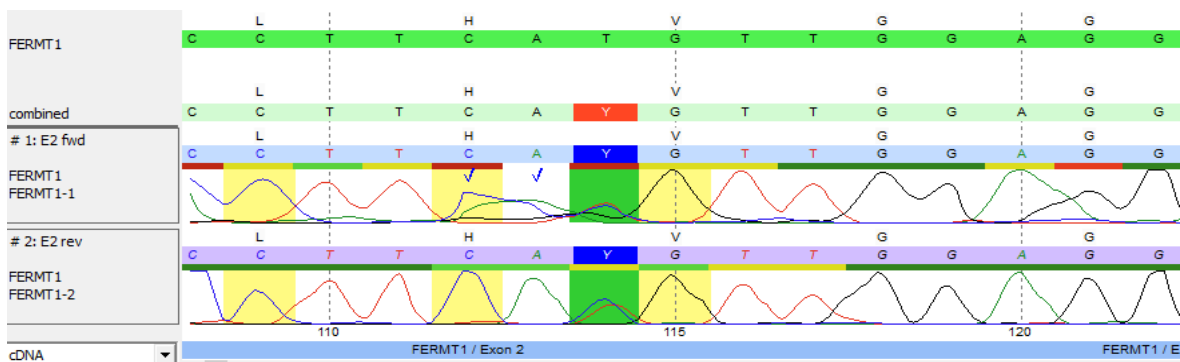
SNP rs2228611 (*DNMT1* Exon 18 c.1389)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
10267077	rs2228611	<i>DNMT1</i>	Exon 18 c.1389	T	C	0,5347593	heterozygot

Bei diesem SNP wird abermals der Minus-Strang sequenziert, sodass es im NGS und in der Sanger-Analyse zu komplementären Referenzbasen kommt. Zusätzlich ist der SNP in der Sanger-Sequenz nur in der Forward-Leserichtung auswertbar. Das heterozygote Ereignis deckt sich aber mit dem Ergebnis des NGS (AltAF 0,5347593).

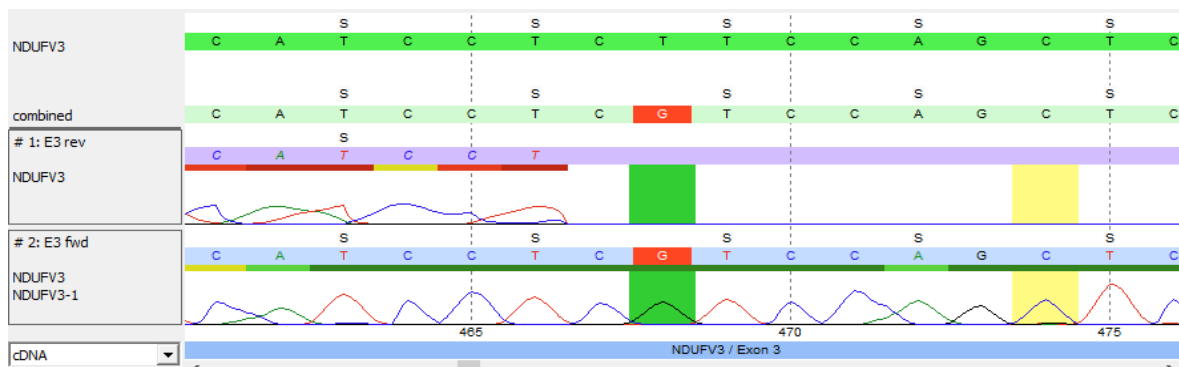
SNP rs10373 (*FERMT1* Exon 2 c.114)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
6100088	rs10373	<i>FERMT1</i>	Exon 2 c.114	A	G	0,5113636	heterozygot

Die Sanger-Sequenzierung dieses SNP erfolgt ebenfalls am Minusstrang. Der reverse Strang detektiert ein deutlich heterozygotes Ereignis, wohingegen der forward Strang ein etwas undeutlicheres heterozygotes Ergebnis zeigt, was aber eher auf eine ungenaue Sequenzierung zurückzuführen ist. Generell stimmt das Ergebnis der Sanger-Sequenzierung mit dem Ergebnis aus dem NGS überein.

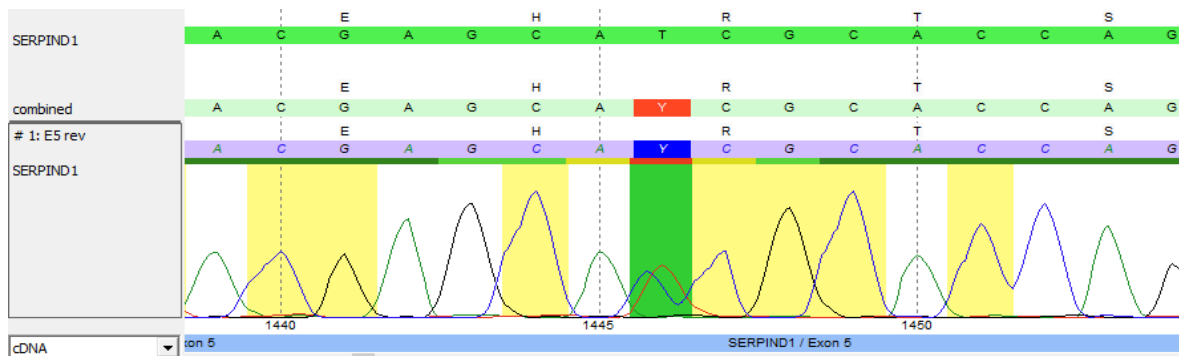
SNP rs4148973 (*NDUFV3* Exon 3 c.468)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
44323590	rs4148973	<i>NDUFV3</i>	Exon 3 c.468	T	G	1	homozygot

Das homozygote Ergebnis des SNP rs4148973 wird in der Sanger-Sequenzierung lediglich in der Forward-Leserichtung nachgewiesen, da die Reverse-Leserichtung kurz vor dem SNP endet. Im NGS wird mit 100 % das alternative Allel (Guanin) detektiert, was ebenso ein homozygotes Ergebnis bedeutet.

SNP rs4675 (*SERPIND1* Exon 5 c.1446)

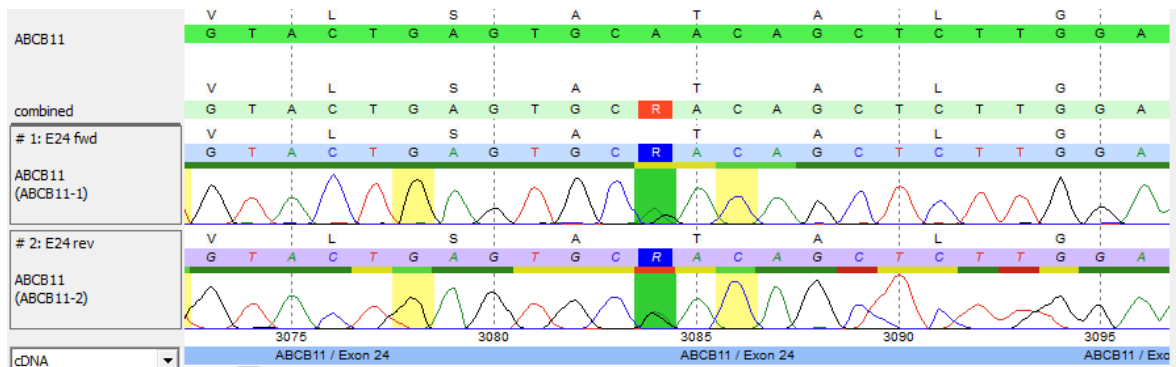


Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
21141300	rs4675	<i>SERPIND1</i>	Exon 5 c.1446	T	C	0,4917127	heterozygot

Ein heterozygotes Ergebnis wird in der Sanger-Sequenzierung für den SNP rs4675 nur mit in der Reverse-Leserichtung nachgewiesen. Die erzielten Ergebnisse aus Sanger und NGS stimmen überein.

Probe B

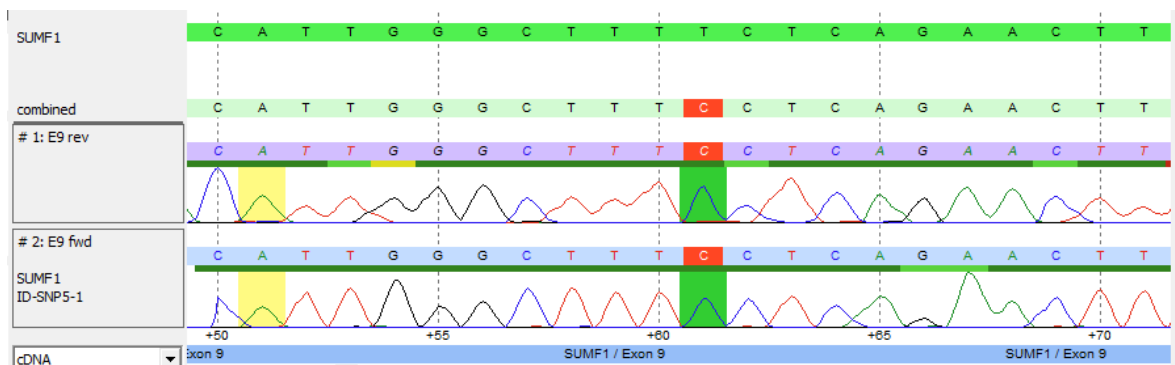
SNP rs497692 (*ABCB11* Exon 24 c.3084)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
169789016	rs497692	<i>ABCB11</i>	Exon 24 c.3084	T	C	0,56804734	heterozygot

Beide Methoden zeigen heterozygote Ergebnisse. Der SNP wird in der Sanger-Sequenz in beide Leserichtungen gut abgedeckt und auch das Ergebnis im NGS ist mit einer AltAF von 0,56804734 deutlich.

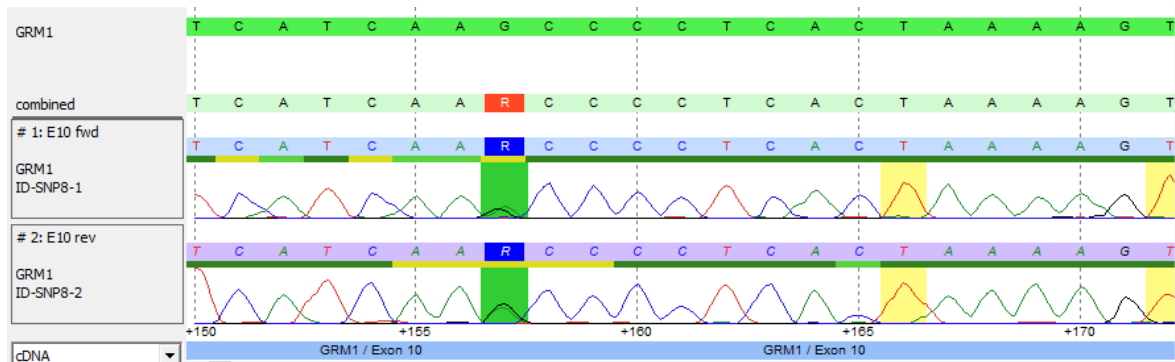
SNP rs2819561 (*SUMF1* Ex9/3'UTR c.*61)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
4403767	rs2819561	<i>SUMF1</i>	Ex9/3'UTR c.*61	A	-	no call	-

Der SNP rs2819561 zeigt in der Sanger-Sequenzierung ein für das Allel C homozygotes Ereignis und wird durch beide Leserichtungen deutlich. Wie bei Probe A wird auch bei Probe B im NGS kein Ergebnis erzielt, sodass ein Vergleich der beiden Methoden nicht möglich ist.

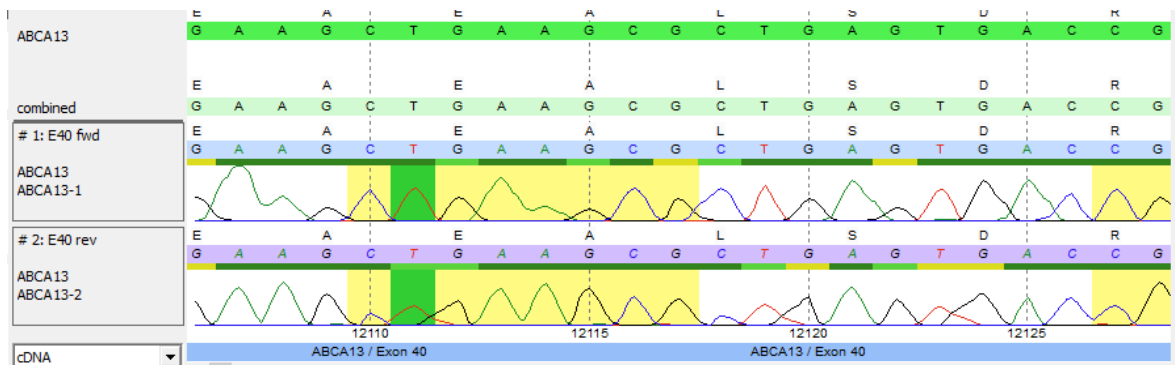
SNP rs2942 (*GRM1* Ex10/3'UTR c.*157)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
146755140	rs2942	<i>GRM1</i>	Ex10/3'UTR c.*157	G	A	0,4391891	heterozygot

Ein übereinstimmendes heterozygotes Ergebnis wird für den SNP rs2942 sowohl mit der Sanger-Sequenzierung als auch mit dem NGS (AltAF 0,4391891) erzielt.

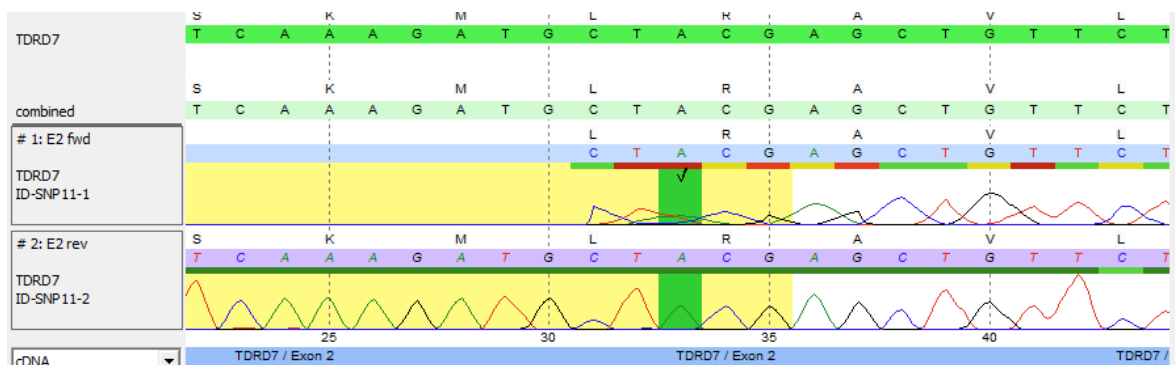
SNP rs17548783 (*ABCA13* Exon 40 c.12111)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
48450157	rs17548783	<i>ABCA13</i>	Exon 40 c.12111	T	T	0	Ref

Die Sanger-Sequenz für den SNP rs17548783 zeigt auf beiden Strängen ein Thymin, was dem Referenzallel entspricht. Auch in der NGS-Analyse wird dieses Ergebnis erzielt.

SNP rs1381532 (*TDRD7* Exon 2 c.33)

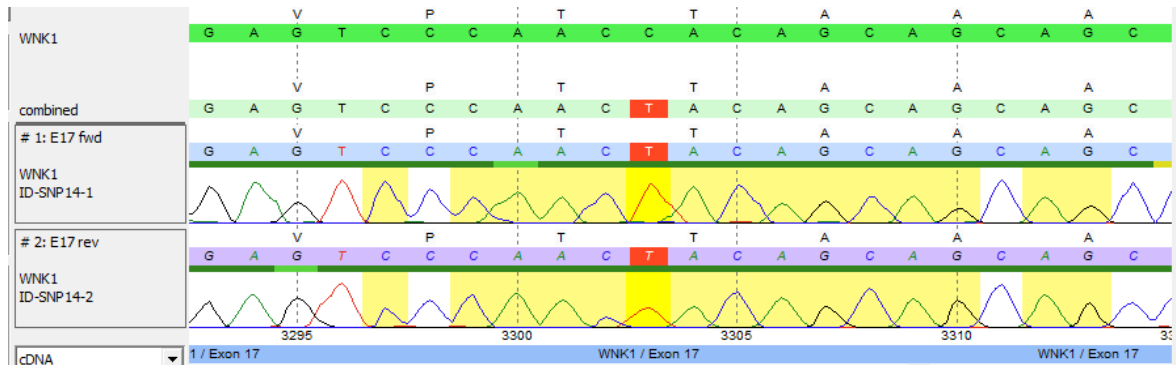


Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
100190780	rs1381532	<i>TDRD7</i>	Exon 2 c.33	A	A	0	Ref

Der SNP im Exon 2 des *TDRD7*-Gens wird nur in der Reverse-Leserichtung in der Sanger-Sequenz gut abgedeckt. Da die Forward-Leserichtung erst kurz vor dem SNP beginnt, wird diese Position etwas unsauber sequenziert. Das homozygote

Ergebnis für das Referenzallel (Adenin) ist jedoch eindeutig und wird auch in der NGS-Analyse erzielt.

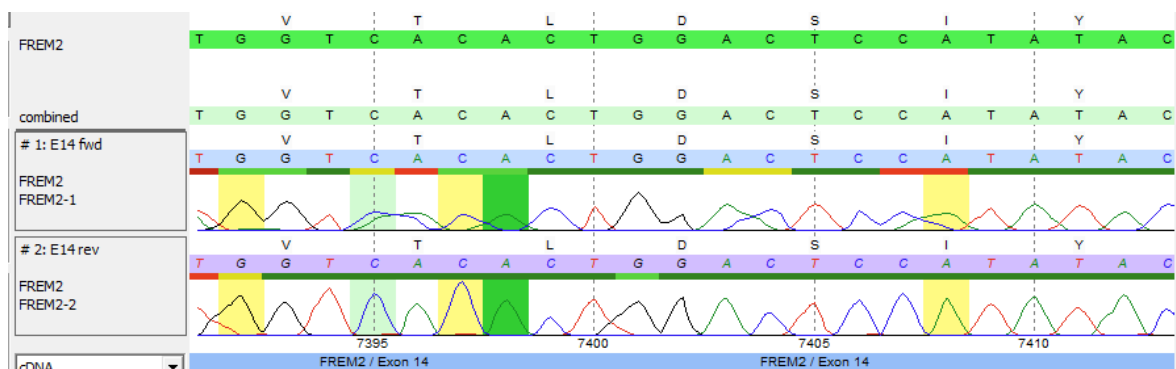
SNP rs7300444 (*WNK1* Exon 17 c.3219)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
993930	rs7300444	<i>WNK1</i>	Exon 17 c.3219	C	C	1	homozygot

Ein eindeutig homozygotes Ereignis für das alternative Allel wird mit beiden Methoden für den SNP rs7300444 erreicht.

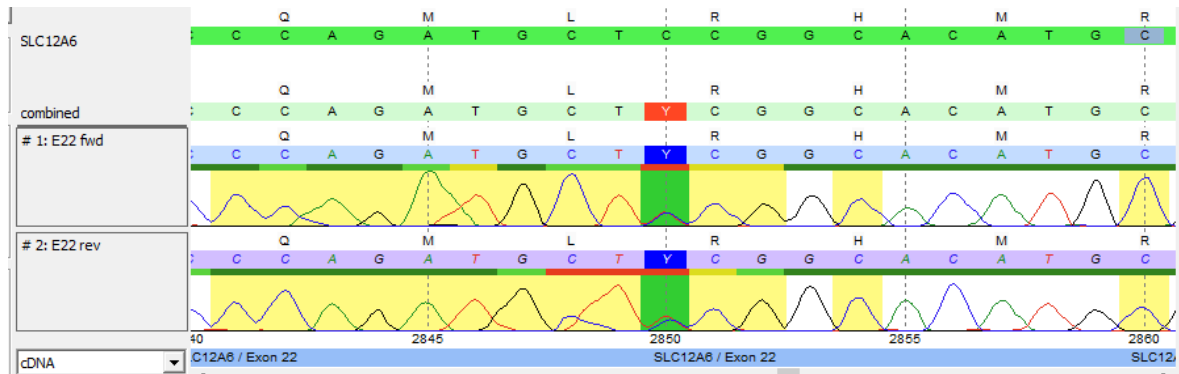
SNP rs9532292 (*FREM2* Exon 14 c.7398)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
39433606	rs9532292	<i>FREM2</i>	Exon 14 c.7398	A	A	0	Ref

Ein homozygotes Ergebnis für die Referenzbase Adenin wird sowohl in der Sanger-Sequenzierung als auch in der NGS-Analyse erzielt.

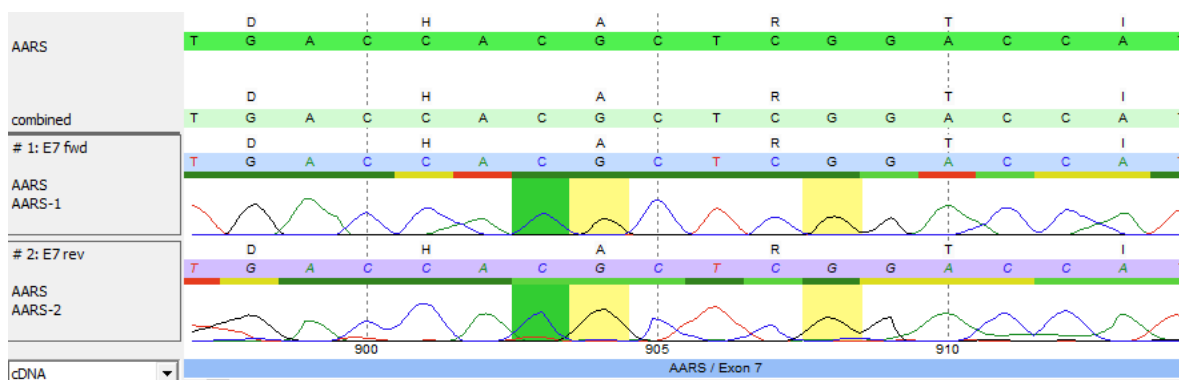
SNP rs4577050 (*SLC12A6* Exon 25 c.2850)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
34528948	rs4577050	<i>SLC12A6</i>	Exon 25 c.2850	G	A	0,5290697	heterozygot

Die Sanger-Sequenzierung zeigt für den SNP rs4577050 in beiden Strängen ein klares heterozygotes Ergebnis (C>T). Im NGS wird mit ca. 53 % das alternative Allel detektiert, was ebenfalls für Heterozygotie spricht.

SNP rs2070203 (*AARS* Exon 7 c.903)

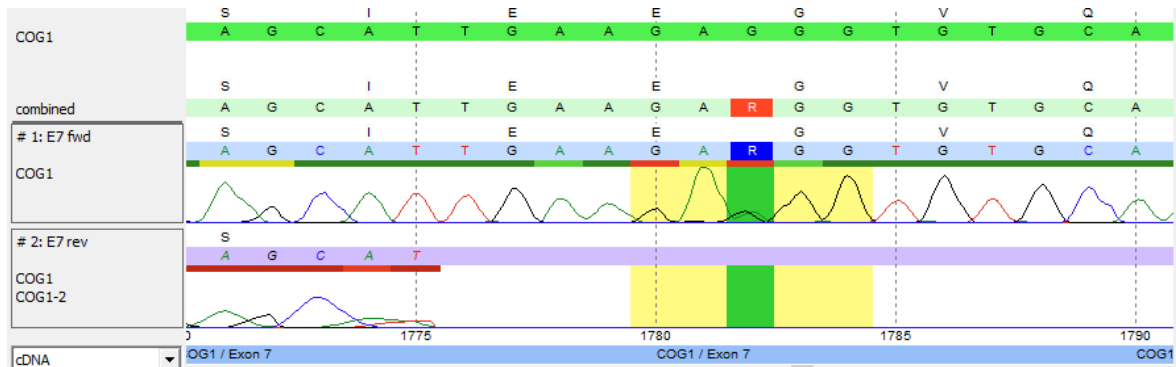


Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie

70303580	rs2070203	AARS	Exon 7 c.903	G	G	0	Ref
----------	-----------	------	-----------------	---	---	---	-----

Der SNP rs2070203 wird in der Sanger-Sequenz von beiden Strängen gut abgedeckt und zeigt ein homozygotes Ergebnis für das Referenzallel Cytosin. Das Ergebnis wird auch im NGS mit einer AltAF von 0 erreicht.

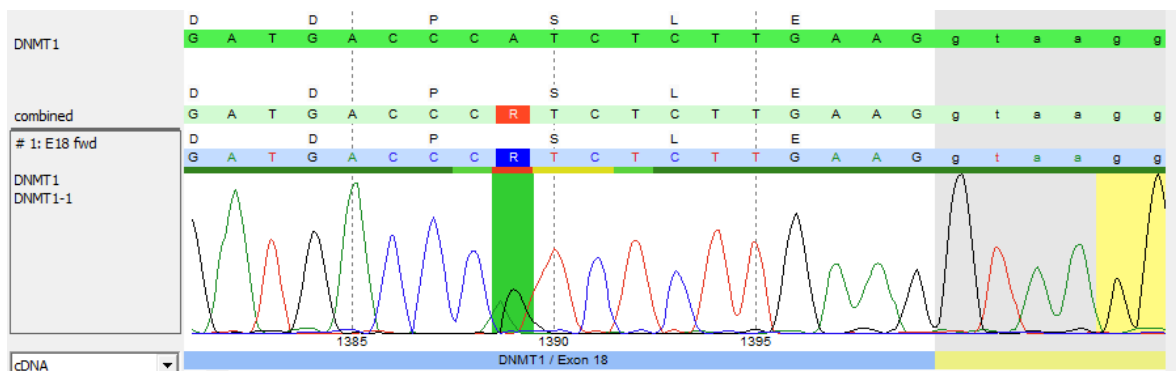
SNP rs1037256 (*COG1* Exon 7 c.1782)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
71197748	rs1037256	<i>COG1</i>	Exon 7 c.1782	G	A	0,4591194	heterozygot

Lediglich in der Forward-Leserichtung wird der SNP im *COG1*-Gen abgedeckt, da die Sequenz in der Reverse-Leserichtung am Ende unsauber wird und in der Leselänge nicht ausreicht. Trotzdem ist das heterozygote Ergebnis eindeutig, und wird auch im NGS (AltAF 0,4591194) erzielt.

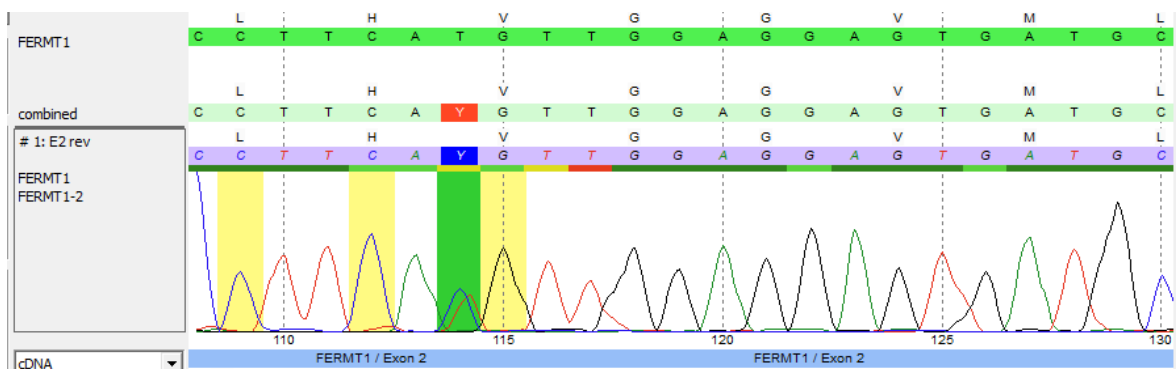
SNP rs2228611 (*DNMT1* Exon 18 c.1389)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
10267077	rs2228611	<i>DNMT1</i>	Exon 18 c.1389	T	C	0,4387755	heterozygot

Für den SNP im *DNMT1*-Gen ist ebenso wie bei Probe A die Auswertung nur anhand der Forward-Leserichtung möglich. Dieser zeigt Heterozygotie (A>G) für den SNP. Dieses Ergebnis wird mit einer AltAF von 0,4387755 im NGS bestätigt.

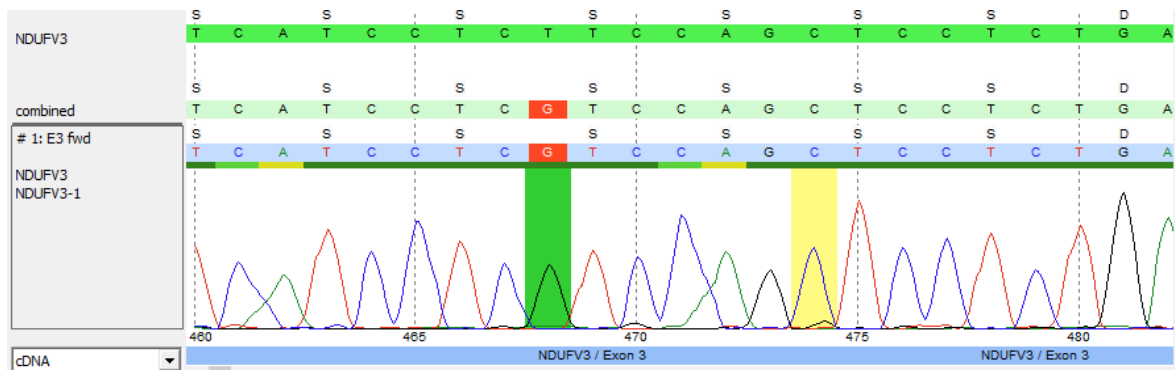
SNP rs10373 (*FERMT1* Exon 2 c.114)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
6100088	rs10373	<i>FERMT1</i>	Exon 2 c.114	A	G	0,4604316	heterozygot

Nur in der Reverse-Leserichtung in der Sanger-Sequenzierung für den SNP rs10373 auswertbar. Ein heterozygoten Ereignis (T>C) wird im reversen Strang erzielt, welches dem Ergebnis der NGS-Analyse entspricht.

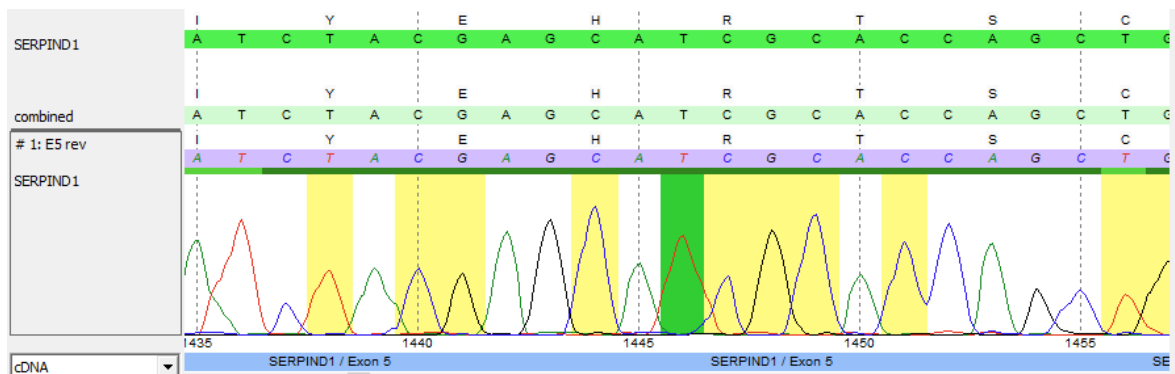
SNP rs4148973 (*NDUFV3* Exon 3 c.468)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
44323590	rs4148973	<i>NDUFV3</i>	Exon 3 c.468	T	G	1	homozygot

Homozygotie für das alternative Allel zeigt sowohl die Sanger-Sequenzierung als auch die NGS-Analyse für den SNP rs4148973. In der Sanger-Sequenzierung liegt zur Auswertung nur die Forward-Leserichtung vor, da in der Reverse-Leserichtung die Leseweite zu kurz ist und nicht über den SNP reicht.

SNP rs4675 (*SERPIND1* Exon 5 c.1446)



Genomic position	rs-Nummer	Gen	Mutation	Genomic ref	Sample call	Sample AltAF	Zygotie
21141300	rs4675	<i>SERPIND1</i>	Exon 5 c.1446	T	T	0	Ref

Ein homozygotes Ergebnis für die Referenzbase Thymin liegt in beiden Methoden für den SNP rs4675 vor, wenn auch lediglich einzelsträngig in der Sanger-Sequenzierung in Reverse-Leserichtung.

7. Literatur

- Bauer, P. et al. 2018. Deutsche Gesellschaft für Humangenetik S1-Leitlinie: Molekulargenetische Diagnostik mit Hochdurchsatz-Verfahren der Keimbahn, beispielsweise mit Next-Generation Sequencing. *medgen* 30, pp. 278–292.
- Bolton-Maggs, P.H.B., Wood, E.M. and Wiersum-Osselton, J.C. 2015. Wrong blood in tube - potential for serious outcomes: can it be prevented? *British Journal of Haematology* 168(1), pp. 3–13. doi: 10.1111/bjh.13137.
- Buermans, H.P.J. and den Dunnen, J.T. 2014. Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842(10), pp. 1932–1941. doi: 10.1016/j.bbadis.2014.06.015.
- Butler, J.M. 2007. Short tandem repeat typing technologies used in human identity testing. *BioTechniques* 43(4), p. Sii-Sv. doi: 10.2144/000112582.
- Consugar, M.B. et al. 2015. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genetics in Medicine* 17(4), pp. 253–261. doi: 10.1038/gim.2014.172.
- Du, Y. et al. 2017. A SNP panel and online tool for checking genotype concordance through comparing QR codes. *PLOS ONE* 12(9), p. e0182438. doi: 10.1371/journal.pone.0182438.
- Heather, J.M. and Chain, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* 107(1), pp. 1–8. doi: 10.1016/j.ygeno.2015.11.003.
- Hu, T., Chitnis, N., Monos, D. and Dinh, A. 2021. Next-generation sequencing technologies: An overview. *Human Immunology* 82(11), pp. 801–811. doi: 10.1016/j.humimm.2021.02.012.
- Huijsmans, C.J.J., Heilmann, F.G.C., van der Zanden, A.G.M., Schneeberger, P.M. and Hermans, M.H.A. 2007. Single nucleotide polymorphism profiling assay to exclude serum sample mix-up. *Vox Sanguinis* 92(2), pp. 148–153. doi: 10.1111/j.1423-0410.2006.00871.x.

- Kim, J.-J., Han, B.-G., Lee, H.-I., Yoo, H.-W. and Lee, J.-K. 2010. Development of SNP-based human identification system. *International Journal of Legal Medicine* 124(2), pp. 125–131. doi: 10.1007/s00414-009-0389-9.
- von Meyer, A. et al. 2017. Standard-Arbeitsanleitung zur peripher venösen Blutentnahme für die labormedizinische Diagnostik. *LaboratoriumsMedizin* 41(6), pp. 333–340. doi: 10.1515/labmed-2017-0127.
- Nakagawa, H. and Fujita, M. 2018. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Science* 109(3), pp. 513–522. doi: 10.1111/cas.13505.
- Nisar, H. et al. 2021. Whole-genome sequencing as a first-tier diagnostic framework for rare genetic diseases. *Experimental Biology and Medicine* 246(24), pp. 2610–2617. doi: 10.1177/15353702211040046.
- Pengelly, R.J., Gibson, J., Andreoletti, G., Collins, A., Mattocks, C.J. and Ennis, S. 2013. A SNP profiling panel for sample tracking in whole-exome sequencing studies. *Genome Medicine* 5(9), p. 89. doi: 10.1186/gm492.
- Rehm, H.L. et al. 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine* 15(9), pp. 733–747. doi: 10.1038/gim.2013.92.
- Slatko, B.E., Gardner, A.F. and Ausubel, F.M. 2018. Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology* 122(1). doi: 10.1002/cpmb.59.
- Tytgat, O., Fauvart, M., Stakenborg, T., Deforce, D. and Van Nieuwerburgh, F. 2021. STRide probes: Single-labeled short tandem repeat identification probes. *Biosensors and Bioelectronics* 180, p. 113135. doi: 10.1016/j.bios.2021.113135.
- Vogt, P.H., Falcao, C.L., Hanstein, R. and Zimmer, J. 2008. The AZF proteins. *International Journal of Andrology* 31(4), pp. 383–394. doi: 10.1111/j.1365-2605.2008.00890.x.
- Zhao, M. et al. 2020. Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genomics and Bioinformatics* 2(2). doi: 10.1093/nargab/lqaa032.

Zhong, Y., Xu, F., Wu, J., Schubert, J. and Li, M.M. 2021. Application of Next Generation Sequencing in Laboratory Medicine. *Annals of Laboratory Medicine* 41(1), pp. 25–43. doi: 10.3343/alm.2021.41.1.25.