

**„Forensisch relevante SNP-Genotypisierung  
mittels elektronischer Microarray-Technologie“**

Dissertation  
zur Erlangung des Grades  
„Doktor der Naturwissenschaften“

am Fachbereich Biologie  
der Johannes Gutenberg-Universität Mainz

Márta Kinga Balogh  
geb. am 27.07.1971 in Klausenburg

Mainz, 2008

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 20.07.2009

*Meinen Eltern gewidmet*

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Grundlagen der forensischen Molekulargenetik	1
1.2	Grundlagen der Einzelnukleotid-Polymorphismen	3
1.3	SNPforID-Projektübersicht und dessen Zielsetzung	6
1.3.1	SNP-Selektion	7
1.3.2	Etablierung von Multiplexstrategien	7
1.3.3	Methoden zur SNP-Genotypisierung	8
1.3.3.1	SNaPshot™-Minisequenzierung	9
1.3.3.2	Microarray-Genotypisierung	9
1.3.3.2.1	Microarray an Glasobjektträgern	10
1.3.3.2.2	Elektronische Microarray-Technologie	11
1.3.3.3	MALDI-TOF-Massenspektrometrie	11
1.3.3.4	Pyrosequencing-Technologie	12
1.3.3.5	SNPplex™/Genplex	13
1.4	Fragestellung und Zielsetzung der Arbeit	15
<b>2</b>	<b>Material und Methoden</b>	<b>16</b>
2.1	Untersuchungsmaterial	16
2.1.1	Einzelnukleotid-Polymorphismen	16
2.1.2	Probenauswahl zur Y-chromosomalen Validierungsstudie	19
2.1.3	Probenauswahl zur forensischen Validierungsstudie	19
2.1.4	Probenauswahl zur Microarray-Typisierung	20
2.2	Methoden	21
2.2.1	Probenquantifizierung	21
2.2.2	Whole Genome Amplification	21
2.2.3	Künstliche DNA-Degradierung	22
2.2.4	Polymerasekettenreaktion	23
2.2.4.1	Multiplex-PCR-Primerdesign	23
2.2.4.2	Multiplex-PCR-Amplifikation	26
2.2.4.3	PCR-Produktkontrolle	27
2.2.4.4	PCR-Optimierung	28
2.2.4.5	PCR-Produktaufreinigung	30
2.2.4.5.1	Enzymatische Aufreinigung	30
2.2.4.5.2	Vergleich verschiedener Aufreinigungsmethoden	30
2.2.5	Minisequenzierung	31
2.2.5.1	Multiplex-SBE-Primerdesign	32
2.2.5.2	SNaPshot™-Minisequenzierung	35
2.2.5.3	Kapillarelektrophorese	36
2.2.5.4	Analyse und Auswertung	37
2.2.6	Elektronische Microarray-Typisierung	39
2.2.6.1	NanoChip® Molecular Biology Workstation	39
2.2.6.2	NanoChip® 400	43
2.2.6.3	Assaydesign	45
2.2.6.3.1	Amplicon down-Assay	45
2.2.6.3.2	Capture down-Assay	47
2.2.6.3.3	Arrayed Primer Extension-Assay (APEX)	50
2.2.6.4	Oligonukleotiddesign	52
2.2.6.5	Auswertung der elektronischen Microarray Daten	57



---

<b>3</b>	<b>Ergebnisse</b>	67
3.1	Ergebnisse der Y-chromosomalen Validierungsstudie	67
3.1.1	Methodische Ergebnisse	67
3.1.2	Ergebnisse der WGA-Proben	70
3.1.3	Populationsspezifische Erkenntnisse	73
3.2	Ergebnisse der forensischen Validierungsstudie	75
3.2.1	Teilergebnisse der forensischen Validierungsstudie	75
3.2.1.1	Ergebnisse der natürlich degradierten Proben	76
3.2.1.2	Ergebnisse der Verdünnungsproben	79
3.2.1.3	Ergebnisse der Mischproben	82
3.2.1.4	Ergebnisse der künstlich degradierten Proben	83
3.2.2	Gesamtergebnisse der forensischen Validierungsstudie	85
3.3	Ergebnisse der Microarray-Technologie	90
3.3.1	Amplicon down-Assayetablierung	90
3.3.2	APEX-Assayetablierung	95
3.3.3	Capture down-Assay	96
3.3.3.1	Capture down-Assayetablierung	96
3.3.3.2	Capture down-Assayoptimierung	108
3.3.3.3	Capture down-Assayvalidierung	124
<b>4</b>	<b>Diskussion</b>	134
4.1	Diskussion zur Y-chromosomalen Validierungsstudie	134
4.1.1	Methodische Anmerkungen	135
4.1.1.1	Segmentduplikationen	135
4.1.1.2	Whole Genome Amplification	138
4.1.2	Populationsspezifische Anmerkungen	140
4.1.3	Schlussfolgerungen der Y-chromosomalen Validierungsstudie	140
4.2	Diskussion zur forensischen Validierungsstudie	142
4.2.1	Methodische Anmerkungen	142
4.2.2	Gesamtdiskussion der forensischen Validierungsstudie	143
4.2.3	Schlussfolgerungen der forensischen Validierungsstudie	144
4.3	Diskussion zur elektronischen Microarray-Technologie	146
4.3.1	Amplicon down-Assay	146
4.3.2	APEX-Assay	149
4.3.3	Capture down-Assay	149
4.3.3.1	Methodische Anmerkungen	149
4.3.3.2	Bewertung des Capture down-Assays	153
4.3.3.3	Vergleichende Beurteilung mit anderen Genotypisierungsmethoden	155
4.3.4	Ausblick	158
<b>5</b>	<b>Zusammenfassung</b>	161
<b>6</b>	<b>Literaturverzeichnis</b>	163
<b>7</b>	<b>Anhang</b>	I
7.1	Abkürzungsverzeichnis	I
7.2	Daten der Y-chromosomalen Validierungsstudie	III
7.3	Vorveröffentlichungen	IX

# 1 Einleitung

## 1.1 Grundlagen der forensischen Molekulargenetik

Die forensische Spurenkunde, deren Hauptaufgabe in der Zuordnung einer biologischen Spur zu einer Person besteht, nahm ihren Ausgangspunkt als neues Arbeitsfeld der Rechtsmedizin in der konventionellen Serologie. Die damalige Einführung der Präzipitationsreaktion zur Artbestimmung (Uhlenhuth 1901) und die Entdeckung der ABO-Blutgruppen am Anfang des 20. Jahrhunderts (Landsteiner 1901) legten das Fundament für heutige Innovationen (Geserick et al. 2001).

Mit der Entdeckung des sogenannten genetischen Fingerabdrucks (Jeffreys et al. 1985) wurden die ursprünglichen Methoden nachhaltig von innovativen DNA-Verfahren abgelöst. Es kam zu einer rasanten Entwicklung in der molekulargenetischen Grundlagenforschung und bei angewandten Verfahren der forensischen Genetik. Der genetische Fingerabdruck, der sich die Variabilität des humanen Genoms zu Nutze macht und mit immer effizienteren Detektionsmethoden das Individuum eindeutig identifiziert, hat sich kontinuierlich in der forensischen Spurenanalyse, im Strafverfahren oder auch bei Vaterschaftsgutachten im Zivilverfahren etabliert (Schneider 2007).

Hohe Sensitivität, Spezifität und Schnelligkeit sind die Kennzeichen moderner DNA-Analysemethoden. Dabei ist Material in Low Copy Number (LCN)-Menge ausreichend, um ein genetisches Profil zu erstellen, welches eine Person mit annähernd 100%-iger Genauigkeit identifiziert (Gill 2001a). Zur DNA-Gewinnung eignen sich neben Blut praktisch auch jegliche biologische Spuren, wie alle Sekrete und Gewebe, aber auch sog. totes Material, wie telogene Haare (Hellmann et al. 2001), Nägel oder abgeschilferte Corneozyten (Findlay et al. 1997), selbst in kleinsten Mengen (Wiegand und Kleiber 1997, Balogh et al. 2003). Außerdem stellt die DNA ein sehr stabiles Untersuchungsmaterial in Bezug auf Dauer, Umstände der Lagerung und Umwelteinflüsse dar. Die molekulare Anthropologie von heute beweist, dass selbst alte Knochen noch intaktes Material freigeben können (Briggs et al. 2007, d'Abbadie et al. 2007). Auch die Opfer schwerer Massenkatastrophen, wie nach den Terroranschlägen vom 11.09.2001 oder dem verheerenden Tsunami in Südostasien 2004, können mittels ihrer DNA identifiziert werden (Biesecker et al. 2005, Prinz et al. 2007).

Der genetische Fingerabdruck hat nach den Vorschriften der Strafprozessordnung (StPO) erst 1996 in die deutsche Gesetzgebung Einlass gefunden. Molekulargenetische Daten werden unter bestimmten Voraussetzungen seit Inkrafttreten des DNA-Identitätsfeststellungsgesetzes (DNA-IFG) am 07.09.1998 in einer bundesweiten DNA-Identifizierungsdatenbank, kurz DNA-Analysedatei (DAD), beim Bundeskriminalamt<sup>1</sup> (BKA) gespeichert.

Technologisch bedient sich die Datenbank der Besonderheit, dass große Teile des humanen Genoms aus hoch polymorphen repetitiven Basensequenzen bestehen. Dabei haben die sog. Short Tandem Repeats (STRs), die sich durch tandemartige Wiederholungseinheiten im nicht-kodierenden Genomabschnitt definieren, aufgrund ihrer Eigenschaften große

---

<sup>1</sup> Bundeskriminalamt Wiesbaden, <http://www.bka.de/profil/faq/dna02.html> (Stand 31.03.2008).

Bedeutung erlangt. Die unterschiedliche Anzahl von Wiederholungseinheiten aus 2-6 Basenpaaren geben die verschiedenen Allele wieder. Schon der Informationsgehalt von wenigen STRs reicht heute aus, um eine valide Aussage zur Identitätsfeststellung zu treffen. Vor diesem Hintergrund bilden heute autosomale STRs die inhaltliche Grundlage aller forensischen DNA-Datenbanken. Aktuell werden weltweit ca. 30 validierte STR-Marker in diversen nationalen Datenbanken dokumentiert (Martin et al. 2001, Schneider und Martin 2001). Die heterogene Ausgestaltung der internationalen Datenbanksysteme stellt eine Herausforderung im Hinblick auf länderübergreifende Standards dar. Europaweit setzt die EDNAP-Gruppe (European DNA Profiling Group) mit Ringversuchen und Qualitätsmanagement Maßstäbe und dient als Unterstützung der Harmonisierungsarbeit von ENFSI (European Network of Forensic Science Institutes). ENFSI definierte im Jahr 2003 sieben STRs (FGA, TH01, vWA, D3S1358, D8S1179, D18S51 und D21S11) als sog. ESS-Loci (European Standard Set of Loci). Flankiert von den alle zwei Jahre stattfindenden weltweiten ISFG-Kongressen (International Society for Forensic Genetics) bildet diese Entwicklung den notwendigen Kern einer gemeinsamen Sprache zur grenzüberschreitenden Spurenkunde (Schneider und Martin 2001).

Diese internationale Festlegung hat, nicht zuletzt unter dem Marktaspekt kommerziell etablierter STR-Multiplex-Kits, die Einführung neuer Marker stark eingeschränkt. Die Fokussierung auf STRs kann jedoch strukturbedingte Probleme bereiten; z.B. bei stark degradierter DNA als Ausgangsmaterial sind die langen Templates ein häufiger Grund für das Erreichen der Nachweisgrenze.

Dieses wissenschaftliche und gesellschaftspolitische Umfeld stellte auch eine Motivation für die Durchführung eines multizentrischen EU-geförderten Projektes dar (Kapitel 1.3.). Zunehmende Leistungsanforderungen machen es somit heute unter Anerkennung der großen Bedeutung PCR basierter STR-Typisierung notwendig, den Forschungsbereich mit verbesserten Methoden zu erweitern. Dabei fällt der Blick auf eine DNA-Analyse, die schon früh in der pharmazeutisch/medizinischen Grundlagenforschung großen Zuspruch fand: Übertragen auf die forensische Molekulargenetik stellt sich die Frage, in welcher Form Einzelnukleotid-Polymorphismen das Potential besitzen, bestehende STR-Methoden im Hinblick auf reproduzierbare, automatisierte Genotypisierungsverfahren zur Individualisierung von Personen zu ergänzen bzw. abzulösen. Zur Beantwortung dieser Frage trägt die vorliegende Dissertation „Forensisch relevante SNP-Genotypisierung mittels elektronischer Microarray-Technologie“ bei.

## 1.2 Grundlagen der Einzelnukleotid-Polymorphismen

Das gesamte menschliche Genom umfasst über drei Milliarden Basenpaare, die sich in Genbereiche und nicht-kodierende Abschnitte aufteilen. Die Gene bestehen aus kodierenden Bereichen (Exons), die Informationen für die Bildung von Strukturproteinen oder Enzymen enthalten und aus nicht-kodierenden Bereichen (Introns), deren Länge die der Exons um ein Vielfaches übertrifft. Insgesamt nehmen kodierende Bereiche nur einen geringen Prozentsatz des Gesamtgenoms ein. Die übrigen Teile, d.h. die nicht-kodierenden Bereiche, bilden den Rahmen dieser Arbeit, die sich mit dort angesiedelten Einzelbasenmutationen beschäftigt.

Unter einer Mutation versteht man die Veränderung der Nucleotidsequenz eines DNA-Moleküls, die auf den DNA-Tochterstrang übertragen wird und in der Keimbahn auftritt (Brown 2007). Eine Mutation kann entweder bei der DNA-Replikation entstehen (Spontanmutation) oder ein Mutagen verursacht einen Basenaustausch (induzierte Mutation). Betrifft die Änderung nur ein einziges Basenpaar, so ist dies eine Punktmutation. Punktmutationen lassen sich in die Gruppen der Transition (Austausch Purin zu Purin oder Pyrimidin zu Pyrimidin) und der Transversion (Austausch Purin zu Pyrimidin oder Pyrimidin zu Purin) einteilen. Weitere Einzelbasenmutationen können durch Insertion oder Deletion eines oder weniger Nucleotide entstehen. Da jede Mutation ein einzigartiges Ereignis in der Evolution des Genoms darstellt, wird sie, je nachdem ob eine Mutation stattgefunden hat, als Prä- (Wildtyp) oder Postmutationsallel (Mutante) bezeichnet. Einzelbasenmutationen, die in einer Population mit einer Allelfrequenz größer 1% auftreten, werden mit dem Ausdruck Single Nucleotide Polymorphisms (SNPs) definiert.

Nach früheren Angaben der International SNP map working group 2001 (Sachidanandam et al. 2001), dass ein SNP durchschnittlich pro 1000 Basen vorkommt, würde die Anzahl der SNPs bezogen auf das ganze menschliche Genom bei 3 Millionen liegen. Zurzeit sind jedoch in der größten, im Internet frei zugänglichen Datenbank dbSNP<sup>2</sup> des amerikanischen National Centre of Biotechnology Information (NCBI) schon mehr als 4,5 Millionen humane SNPs eingetragen, die durch unabhängige Genotypisierung und Frequenzdaten validiert sind (Stand August 2007). Schon die enorme Anzahl eindeutig definierbarer SNPs, deren Verteilung über das ganze humane Genom und zahlreiche medizinische Expertisen über die genetische Prädisposition bestimmter Krankheiten bis hin zur Pharmakogenetik, machen SNPs interessant als neue genetische Werkzeuge für die forensische Molekulargenetik.

Gewöhnlich werden menschliche Krankheiten, mit Ausnahme monogener Erbkrankheiten, nicht durch genetische Variationen eines einzelnen Gens hervorgerufen, sondern durch komplexe Einflüsse, bestehend aus einer Interaktion zwischen mehreren Genen und äußeren Faktoren, wie Umwelt und Lebensstil. Obwohl letztere Faktoren und ihre Auswirkungen auf Krankheitsprozesse bedeutend sind, ist es äußerst schwierig sie zu bestimmen, zu messen oder zu evaluieren. Mehr Aufschluss bringt die Erforschung genetischer Prädispositionen von Krankheiten unter anderem bei Herzkrankheiten (Samani

---

<sup>2</sup> <http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp>

et al. 2007), Diabetes (Elbein et al. 2002, Willer et al. 2007), und Krebs (Glinsky 2006, Gusella und Padriani 2007). Mittels sogenannter Assoziationsstudien, welche DNA-Sequenzen von Kranken und Gesunden auf Differenzen abgleichen, versucht man krankheitsbeteiligte Variationen zu bestimmen (Christensen und Murray 2007). SNPs revolutionieren die bisherigen pharmakokinetischen Betrachtungsweisen als Grundlage der Pharmakogenetik. Es wurde festgestellt, dass SNPs mit unterschiedlicher Resorption, Metabolisierung und Wirkung bestimmter Arzneistoffe assoziiert sind. Einen Schlüssel stellen SNPs im Bauplan des menschlichen Cytochrom P450 Systems dar, das mit seinen Isoenzymen vor allem für den Fremdstoffmetabolismus verantwortlich ist (Relling 1989). Je nach Mutation kann dies dazu führen, dass ein bestimmter Arzneistoff bei einem Patienten schneller (ultrarapid metabolizer) und einem Anderen langsamer (poor metabolizer) abgebaut wird. Die Vorhersage des Metabolisierungstyps dient der individuellen Arzneimittelauswahl und Dosisanpassung. Entsprechende kommerzielle Microarrays sind bereits verfügbar z.B. AmpliChip™ CYP450 (Roche Diagnostics). In Zukunft könnte die SNP-Forschung auf dem Gebiet der Pharmakogenetik zu einer sogenannten „personalisierten Medizin“ führen, die den Pharmaherstellern und Medizinern hilft, Patienten mit individuell zugeschnittenen Medikamenten zu behandeln (Nebert 1999, Oscarson 2003).

Innovationen auf dem medizinischen Sektor bieten vielversprechende Ansätze für die SNP-Analyse auf dem jüngsten Einsatzfeld der Forensik und Anthropologie. Zur Diskussion über die Verwendung von Single Nucleotide Polymorphisms als neue Marker in der forensischen Genetik kam es bereits im Jahr 1999 (Chakraborty et al. 1999). Obwohl die Arbeitsgruppe um B. Budowle den SNPs keine adäquate Relevanz gegenüber STRs abgewinnen konnte, wurden die SNP-Marker kontinuierlich Gegenstand der forensischen Forschung. Als erstes wurden die Vorteile der Y-chromosomalen SNPs durch die Arbeiten von Underhill und Jobling deutlich (Semino et al. 2000, Underhill et al. 2000, Jobling 2001, Underhill et al. 2001). Diesen folgten zahlreiche Veröffentlichungen über die Verwendung von Y-chromosomalen und mtDNA SNPs zur Beantwortung forensischer Fragestellungen (Parsons und Coble 2001, Raitio et al. 2001, Arredi et al. 2004, Cinnioglu et al. 2004, Rootsi et al. 2004, Vallone und Butler 2004). Gleichzeitig wurden erste technische Annäherungen in Form von Genotypisierungsmethoden (Pusch et al. 2002, Wise et al. 2003) oder auch Multiplexstrategien vorgestellt (Sanchez et al. 2003), um den Erwartungen der forensischen Anwendung zu entsprechen (Gill 2001b).

Für die speziellen Anforderungen in der Forensik und Anthropologie können insbesondere die Eigenschaften der SNPs mit ihren Vorteilen gegenüber anderen in der Routine eingesetzten Markern von großem Nutzen sein. Die Wissenschaft der anthropologischen und forensischen Molekulargenetik profitiert von der im Vergleich zu den STR-Systemen extrem geringen Mutationsrate der SNPs. Aufgrund der niedrigen Mutationsrate müssen SNPs über sehr viele Generationen vererbt werden, um eine relative Allelfrequenz in einer Population zu erreichen; Mit der Konsequenz, dass SNPs mit der höchsten Allelfrequenz und der weitesten Verbreitung die ältesten Mutationen darstellen, woraus sich entsprechende Abstammungs- und Populationsverhältnisse ableiten lassen (Lao et al. 2006).

Darüber hinaus liefert die rekombinationsfreie Vererbung des Y-chromosomalen SNPs und mtDNA SNPs eine patrilineare bzw. matrilineare Aufklärung zu den populationsabhängigen Fragestellungen, wie z.B. über den räumlichen und zeitlichen Verlauf einer demographischen

Expansion von Populationen. Weiterhin kann diese Informationskraft in einer forensischen Kasuistik wichtige Rückschlüsse zur Identifizierung eines Spurenlegers geben.

Besonders für die anthropologische und forensische Molekulargenetik, wo das Untersuchungsmaterial forschungsfeldbedingte Herausforderungen stellt, sind SNPs, die durch die Änderung einer einzigen Base definiert sind, von großem Vorteil. Die SNP-Analyse benötigt nur eine relativ kurze DNA-Sequenz für eine Informationsaussage, während die STR-Amplicons schon Fragmentlängen von bis zu 400bp erfordern. Das bedeutet, bei der SNP-Typisierung entfällt die Betrachtung von komplexen DNA-Sequenzen, was bei Untersuchungsmaterial mit nur ganz geringen Mengen (LCN) oder in stark degradiertem Zustand, wo gewöhnlich die STR-Analyse versagt, Erfolg versprechend ist.

Die strukturelle Konstellation der SNPs ist zusätzlich für das Design und die Validierung einer größeren Multiplexanalyse von Vorteil verbunden mit einem Benefit bei der Automatisierung. Unter dem Aspekt der Vollautomatisierung kann theoretisch die Multiplexkapazität der SNPs unter Verwendung von Hochdurchsatz-Microarrays auf 100000 gesteigert werden (Butler 2005). Dies würde wiederum die Erhöhung der Effizienz bei gleichzeitiger Reduzierung von benötigtem Untersuchungsmaterial, von zeitaufwendigem Handling und auch von Analysekosten bedeuten.

## 1.3 SNPforID-Projektübersicht und dessen Zielsetzung

Das internationale EU-Projekt mit dem Titel High Throughput Analysis of Single Nucleotide Polymorphisms for the Forensic Identification of Persons, kurz SNPforID, wurde mit Unterstützung der europäischen Kommission in Brüssel Ende 2002 im Rahmen des Competitive and Sustainable GROWTH Programms gestartet. Die vorliegende Dissertation ist Teil des Projektes.

Das SNPforID-Konsortium bildeten, neben dem Institut für Rechtsmedizin der Johannes Gutenberg-Universität Mainz, drei weitere Institute der Universitäten in Santiago de Compostela, Spanien (Instituto de Medicina Legal, Universidade de Santiago de Compostela), in Kopenhagen, Dänemark (Retsmedicinsk Institut, Københavens Universitet) und in London, Großbritannien (Institute of Health Sciences Education, Centre for Haematology, Barts and The London Queen Mary's School of Medicine and Dentistry London).

Im Folgenden werden die Zielsetzungen des Projektes und die genauen Arbeitsteilungen mit einer Zuständigkeitszuordnung beschrieben. Dabei werden nur die Teilbereiche des Projektes geschildert, die im Kontext oder methodisch im Zusammenhang stehen mit der vorliegenden Arbeit.

Der Gegenstand des dreieinhalb Jahre andauernden SNPforID-Projektes waren die Einzelnukleotid-Polymorphismen. Die übergreifende Zielsetzung bestand darin, eine Genotypisierungsmethode zu etablieren und zu validieren, die mit hoher Genauigkeit, aber geringem Aufwand mehrere SNPs simultan in einem Hochdurchsatzverfahren analysieren kann. Um das Ziel des Projektes zu erreichen wurden folgende Teilaufgaben bearbeitet:

1. Die Selektion von mindestens 50 autosomalen SNPs zur Identifikation von Personen weltweit und die Selektion von Y-chromosomal SNPs, die die Haupthaplogruppen der Welt definieren.
2. Die Entwicklung von Multiplex-PCR-Strategien für die Amplifikation der autosomalen SNPs und der Y-chromosomal SNPs in jeweils einer Multiplexreaktion.
3. Die Auswahl einer Detektionsmethode, welche in allen Laboratorien des Projektes als Referenztechnologie agieren kann und ihre Etablierung für die autosomalen und die Y-chromosomal SNPs.
4. Die Validierung der ausgewählten SNPs und der Referenzmethode in jedem Labor, an vor Ort vorhandenen Kollektiven.
5. Die Adaptation und die Etablierung der autosomalen SNPs an Hochdurchsatztechnologien, die den Mitgliedern des Projektes zur Verfügung stehen.
6. Die forensische Validierung der autosomalen SNPs und der Referenzmethode, mittels Analyse von Spuren relevantem Material, d.h. künstlich und natürlich degradierter DNA auch in geringen Mengen (LCN) und Mischproben.

### 1.3.1 SNP-Selektion

Hinsichtlich der enormen Anzahl der detektierten SNPs und ihrer zahlreichen Dokumentationen bedurfte es zunächst einer zielorientierten SNP-Selektion. Diese erfolgte zentral im Labor in Santiago de Compostela unter Berücksichtigung der Kriterien und Erwartungen, die durch das Konsortium festgelegt wurden (Phillips et al. 2004a, Phillips 2005). Differenziert werden sollten Einzelbasenmutationen, einerseits auf dem Y-Chromosom und andererseits auf autosomalen Chromosomen. Die genaue Auswahl der SNPs beider Gruppen ist im Kapitel Material und Methoden (Kapitel 2.1.1) beschrieben.

Aufgrund der patrilinearen Vererbung bei langsamer Akkumulation werden die Y-chromosomalen Marker in Haplogruppen eingeteilt. Vor diesem Hintergrund sollte die Gruppe der Y-chromosomalen SNPs so gewählt werden, dass sie die größten Haplogruppen der Welt charakterisiert (Jobling et al. 1997, Jobling 2001, Jobling und Tyler-Smith 2003).

Die allgemeine Erwartung an die Gruppe der autosomalen SNPs war, dass sie eine optimale, forensisch relevante Typisierung und damit eine sichere Individualisierung einer Person, unabhängig von ihrer Herkunft, erlaubt. Für eine genaue Identifizierung, ähnlich der konventionellen STR-Marker, sind ungefähr 50 binäre SNPs mit Allelfrequenzen von 0,5:0,5 nötig (Gill 2001b). Die ausgewählten SNPs müssen neben den Erwartungen auch angewandte Kriterien erfüllen, sich zum Beispiel für eine robuste und zuverlässige Amplifikation eignen. Die Selektion von mindestens 50 SNPs, gegenüber den adäquaten 12-16 STRs, wird durch diese Kriterien erheblich erschwert. So stellt allein schon die Anzahl der benötigten SNPs ein Hindernis dar. Es müssen mehrere SNPs von einem Chromosom gewählt werden, damit besteht die Gefahr der Kopplung mit Genen oder anderen Markern. Zusätzlich wird die Gefahr der Primerinteraktion für so viele simultan amplifizierte SNPs erheblich erhöht.

### 1.3.2 Etablierung von Multiplexstrategien

Die multiple SNP-Typisierung basiert in erster Linie auf einer Multiplex-PCR-Amplifikation, welche eine simultane Vervielfältigung aller gewünschten DNA-Fragmente erlaubt. Darüber hinaus ist es für eine multiple SNP-Analyse wichtig, Multiplexstrategien zur Typisierungsmethode und zur Detektion auszuarbeiten. Ziel war die Etablierung einer Multiplex-PCR, welche die Amplicons aller autosomalen SNPs bzw. aller Y-chromosomalen SNPs in balancierten Konzentrationen liefert. Außerdem war es für eine homogene und klare SNP-Detektion von Bedeutung, Multiplexstrategien zur Single Base Extension (SBE)-Referenz-typisierungsmethode zu etablieren.

Das Multiplex-PCR-Primerdesign für die autosomalen SNPs erfolgte im rechtsmedizinischen Institut in Kopenhagen (Sanchez et al. 2003, Sanchez et al. 2006a). Dabei wurden größte Akzente auf die Flexibilität bei der Primerauswahl, Spezifität der Primer und uniforme PCR-Amplifikation gesetzt.

Die Primer für 13 SNPs aus der Y-chromosomalen Gruppe wurden im rechtsmedizinischen Institut in Kopenhagen entworfen und bereits 2003 veröffentlicht (Sanchez et al. 2003). Die



restlichen Primer wurden nach den gleichen Kriterien im rechtsmedizinischen Institut in Santiago de Compostela ausgewählt (Brion et al. 2005).

Für die Validierung der autosomalen und der Y-chromosomalen SNPs diente die SBE-Minisequenzierung als Typisierungsmethode. Die Auswahl der Analyseprimer für diese Methode besitzt einen hohen Stellenwert. Die Analyseprimer müssen spezifisch binden, um eine balancierte und klare Detektion der SNPs zu liefern. Während die Auswahl der SBE-Analyseprimer für die autosomalen SNPs im rechtsmedizinischen Institut in Kopenhagen erfolgte, wurden die Analyseprimer für die Y-chromosomalen SNPs, nach den bereits 2003 von Sanchez et al. veröffentlichten Kriterien und Strategien, von der Arbeitsgruppe in Santiago de Compostela bereitgestellt.

Die Gesamtumsetzung des Oligodesigns ist für die Multiplex-PCR in Kapitel 2.2.4.1 und für Multiplex-Minisequenzierung in Kapitel 2.2.5.1 beschrieben.

### 1.3.3 Methoden zur SNP-Genotypisierung

In den letzten Jahren wurden zahlreiche Genotypisierungsmethoden für die SNP-Analyse entwickelt (Gut 2001). Um sich einen leichteren Überblick über die Unterschiede dieser Methoden zu verschaffen, zeigt die Tabelle 1 die beteiligten Einzelprozesse in drei Kategorien: Diskriminierungsreaktion, Reaktionsmodalität und Detektionsmethode (Kwok 2001, Syvanen 2001, Sobrino et al. 2005). Je nach Bedingung und Zielsetzung kann ein Prozess der Kategorie Diskriminierungsreaktion mit einer bestimmten Reaktionsmodalität kombiniert werden und eine passende Detektionsmethode hinzugefügt werden, um ein Genotypisierungsverfahren aufzubauen. Die Kategorie der Diskriminierungsreaktion beschreibt den molekularen Mechanismus, der bei der Reaktion abläuft. Das Ergebnis einer Diskriminierungsreaktion kann mit diversen Detektionsmethoden analysiert werden. So ist zum Beispiel eine Primer Extension-Reaktion mit allen in der Tabelle 1 aufgeführten Methoden detektierbar.

Tabelle 1: Einzelprozesse der SNP-Genotypisierungsmethoden unterschieden in drei Kategorien, ungeachtet der Kombinationsmöglichkeiten der einzelnen Kategorien.

Diskriminierungsreaktion	Reaktionsmodalität	Detektionsmethode
Allel spezifische Hybridisierung	Homogenreaktion	Fluoreszenz
Primer Extension	Festphasenreaktion	Fluoreszenz-Polarisation
Oligonucleotide Ligation	Gelelektrophorese	FRET <sup>1)</sup>
Enzymatische Spaltung	Semihomogen	Massenspektrometrie Chemilumineszenz

1) FRET, Fluorescence Resonance Energy Transfer

Die Reaktionsmodalität beschreibt den Separationsschritt. Eine Homogenreaktion läuft zum Beispiel in einer Lösung ohne weitere Separations- oder Aufreinigungsschritte ab. Sie ist meist wenig arbeitsaufwendig und eignet sich gut für ein automatisiertes Assay, ist aber nur limitiert multiplexfähig. Im Gegensatz dazu bietet die Festphasenreaktion die Möglichkeit, unter Verwendung von Glasobjektträgern, Microarrays, Beads o. Ä. mehrere Reaktionen

parallel durchzuführen und damit den Zeitaufwand zu reduzieren. Die Tabelle 2 ordnet die Genotypisierungsmethoden, die in dieser Arbeit behandelt werden, den drei Kategorien zu.

Tabelle 2: Auflistung der Genotypisierungsmethoden aus der vorliegenden Arbeit und ihre Zuordnung der drei Einzelprozesskategorien.

Assay Name (Firma)	Diskriminierungsreaktion	Reaktionsmodalität	Detektionsmethode
SNaPshot™ (AB)	Primer Extension	Gelelektrophorese	Fluoreszenz
Glasarray (Eigenbau)	Primer Extension	Festphasenreaktion	Fluoreszenz
NanoChip™ (Nanogen)	Allel spez. Hybridisierung	Festphasenreaktion	Fluoreszenz
MALDI-TOF (Sequenom)	Primer Extension	Festphasenreaktion	Massenspektrometrie
Pyrosequencing (Biotage)	Primer Extension	Semihomogen	Chemilumineszenz
SNPplex™/GenPlex (AB)	Oligonucleotide Ligation	Gelelektrophorese	Fluoreszenz

### 1.3.3.1 SNaPshot™-Minisequenzierung

Die SNaPshot™-Reaktion (Applied Biosystems) beruht auf einer SBE-Minisequenzierung und ist kommerziell als Kit erhältlich. Das Prinzip der Minisequenzierung und auch des SNaPshot™-Kits ist in Kapitel 2.2.5 ausführlich beschrieben.

Die SNaPshot™-Typisierungsmethode diente zur Validierung der autosomalen SNPs sowie der Y-chromosomalen SNPs. Zusätzlich stellte die SNaPshot™-Minisequenzierung eine Referenzmethode gegenüber anderen, in diesem Projekt benutzten Genotypisierungsmethoden dar, so auch für die elektronische Microarray-Methode, die in dieser Arbeit behandelt wird.

### 1.3.3.2 Microarray-Genotypisierung

Als Microarrays werden allgemein monolithische Träger mit zahlreichen Testorten bezeichnet, von denen jeder DNA-Sonden enthält, die mit der Erkennung eines komplementären Moleküls ein Signal liefern, das durch ein Detektionsverfahren nachzuweisen ist.

Seitdem der Einsatz von Microarrays für die DNA-Analyse das erste Mal im Jahre 1995 (Skena et al. 1995) beschrieben wurde, nimmt ihre Anwendung exponentiell zu. Anfangs wurden sie ausschließlich für Expressionsanalysen von Genen verwendet, aber ihr einfacher und günstiger Zugang zu Informationen über DNA-Gehalt und Sequenzen erweckte Aufmerksamkeit und die Methode wurde schnell für Einsätze wie die Detektion von Aberationen (Solinas-Toldo et al. 1997), SNPs (Landegren et al. 1988, Pastinen et al. 1997, Landegren et al. 1998), RNA-Interferenzeffekten (Mousses et al. 2003), oder Protein-Interaktionen (Cahill und Nordhoff 2003) adaptiert. Nicht zuletzt die wachsende Bedeutung in der medizinischen Grundlagenforschung begünstigte die ständige Weiterentwicklung und Optimierung der Microarray-Genotypisierungsmethode.

Im Rahmen des SNPforID-Projektes wurden zwei Methoden, die SNPs mittels Microarray-Typisierung analysieren, etabliert und validiert. Während die Microarray-Methode an Glasobjektträgern eine rein passive Typisierung darstellt, erfolgen die Analyseschritte mittels Nanogen-Technologie elektronisch.

### 1.3.3.2.1 Microarray an Glasobjektträgern

Die bereits oft beschriebene und weit verbreitete Microarray-Genotypisierung an Glasobjektträgern ist eine vielversprechende Methode für die Hochdurchsatzanalyse von Multiplexen (Hirschhorn et al. 2000, Pastinen et al. 2000, Lindroos et al. 2001, Lindroos et al. 2002). Das Verfahren besteht aus einer SBE-Minisequenzierung und anschließender SNP-Analyse auf dem Microarray. Nach einer Standard Multiplex-PCR-Amplifikation und Aufreinigung folgt die Minisequenzierungsreaktion (Abbildung 1). Dazu werden Analyseprimer verwendet, die an ihrem 5' Ende spezifische Tag-Sequenzen tragen und am 3' Ende unmittelbar vor den zu detektierenden SNPs enden. Die Tag-Sequenzen können beispielsweise von der Tag-Liste des  $\lambda$ -Bakteriophagen-Genoms der Firma Affimetrix gewählt werden (Hirschhorn et al. 2000).

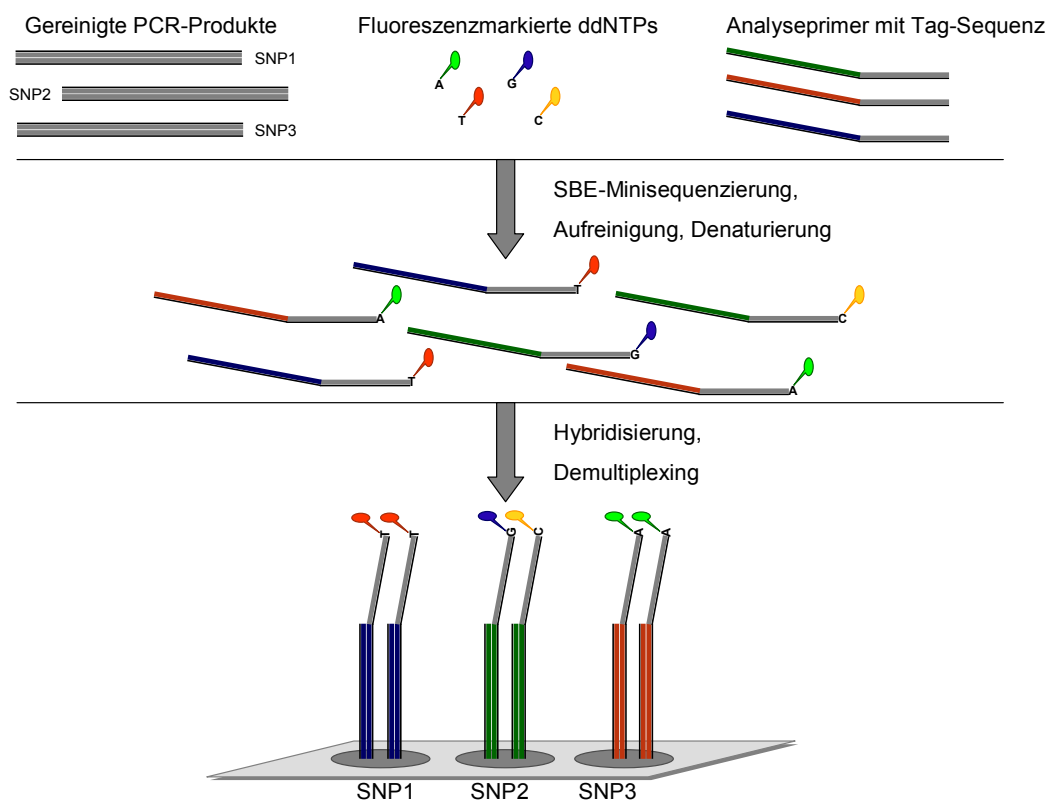


Abbildung 1: Das Prinzip der Microarray-Genotypisierungsmethode an einem Glasobjektträger.

Die dazu komplementären Sequenzen (cTag's) werden auf einem Glasobjektträger immobilisiert. Die Sequenzkopplung an der Oberfläche erfolgt über eine kovalente Bindung. Durch die Verwendung einer speziell angefertigten Kammer aus gehärtetem Silikon, die auf

den Objektträger gedrückt wird, können auf einem Objektträger parallel mehrere verschiedene PCR-Produkte aufgetragen werden (sog. array of arrays, Pastinen et al. 2000). Bei der anschließenden thermischen Hybridisierung kommt es zur Auftrennung der Multiplexe auf die einzelnen SNPs (Demultiplexing, Abbildung 1). Durch die Tag-Verlängerung hybridisieren die Minisequenzierungsprodukte spezifisch an die immobilisierten cTag's. Unter stringenten Hybridisierungs- und Waschbedingungen führt ein mismatch zu einer geringeren Bindung und resultiert in einer deutlich reduzierten Signalintensität beim Scanvorgang.

Dieses Verfahren wurde in Santiago de Compostela und Mainz parallel etabliert. Während in Santiago de Compostela nur ein einfach-Laser-Scanner zur Verfügung stand, konnte in Mainz ein vierfach-Laser-Scanner zur Validierung der Methode benutzt werden (Sobrinho et al. 2004, Bogus et al. 2006).

### **1.3.3.2 Elektronische Microarray-Technologie**

Im Gegensatz zu der passiven Bindung am Glasobjektträger erfolgt die Hybridisierung bei der elektronischen Microarray-Technologie durch elektronische Adressierung. Dabei werden die zu untersuchenden DNA-Sequenzen durch elektrische Felder gezielt an einen bestimmten Ort auf dem Microarray geleitet (Kapitel 2.2.6). Dem SNPforID-Konsortium stand als elektronische Microarray-Technologie die NanoChip™ Molecular Biology Workstation (MBW, Nanogen) in Kopenhagen und Mainz zur Verfügung. Die NanoChip™-Technologie wurde als erstes in Kopenhagen mittels kleiner SNP-Multiplexe – bestehend aus fünf Y-chromosomalen SNPs – auf Sensitivität überprüft. Weiterführende Experimente zur Etablierung, Optimierung und schließlich zur Validierung der NanoChip™-Microarray-Technologie erfolgten anschließend nur in Mainz im Rahmen dieser Arbeit (Kapitel 2.2.6).

### **1.3.3.3 MALDI-TOF-Massenspektrometrie**

Von der SNP-Genotypisierung mittels SBE-Minisequenzierung und MALDI-TOF-MS (Matrix Assisted Laser Desorption/Ionisation Time Of Flight Mass Spectrometry) als Detektionsmethode wurde bereits oft berichtet (Haff und Smirnov 1997b, Pusch et al. 2002). Die Anwendbarkeit der Plattform für Multiplexanalysen wurde früh aufgezeigt (Haff und Smirnov 1997b, Ross et al. 1998), aber nicht eingehend untersucht. Aus diesem Grund war es für das SNPforID-Projekt von Bedeutung, Multiplexassays für das MALDI-TOF-Massenspektrometer zu validieren. Diese Aufgabe wurde in dem forensischen Labor Kopenhagen bearbeitet (Mengel-Jorgensen et al. 2005).

Die MALDI-TOF-MS-Detektionsmethode nutzt die Tatsache, dass die vier verschiedenen Nukleotide spezifische Molekulargewichte aufweisen. Das Analyseprinzip basiert auf einer SBE-Minisequenzierungsreaktion, wobei an das gereinigte PCR-Produkt spezifische Primer hybridisieren und es durch eine Polymerase um genau eine Base (ddNTP) verlängert wird. Nach einer weiteren Aufreinigung werden die Proben auf eine Matrix transferiert, um eine photolytische Beschädigung und Interaktion der Proben zu verhindern. Die zu untersuchende

Probe wird bei der Desorption einem Laserpuls von wenigen Nanosekunden Dauer ausgesetzt. Die Matrix absorbiert die zugeführte Laserenergie und unterstützt damit die Ionisierung der Moleküle. Bei der Desorption/Ionisation erzeugte Ionen werden in einem elektrischen Feld beschleunigt und dabei abhängig vom Masse/Ladungsverhältnis ( $m/z$ ) aufgetrennt. Die TOF (Time Of Flight) gibt eine sehr genaue Zeitmessung von der Ionenquelle bis zum Eintreffen am Detektor an. Die in der Minisequenzierungsreaktion erzeugten DNA-Fragmente werden nach ihrer Flugzeit und der Größe unterschieden und von einer Software einem Genotyp zugeordnet.

### 1.3.3.4 Pyrosequencing-Technologie

Die für die Technologie namensgebende Firma Pyrosequencing wurde im Laufe des Projektes zur Biotage (Uppsala, Schweden) und dabei so umstrukturiert, dass sie ihr Interesse und die Unterstützung an forensischer Forschung aufgab. Aus diesem Grund kam es zum vorzeitigen Abbruch der Etablierung und Validierung dieser Technologie für die Multiplex-SNP-Analyse. Initialversuche wurden in Mainz und London erstellt (Bender et al. 2006, Harrison et al. 2006). Das Prinzip der Methode beruht auf einer Enzymkaskade. Die Grundlagen wurden erstmals 1987 von Nyren vorgestellt, danach wurde das Verfahren kontinuierlich weiterentwickelt (Ahmadian et al. 2000, Ahmadian et al. 2006, Nyren 2007).

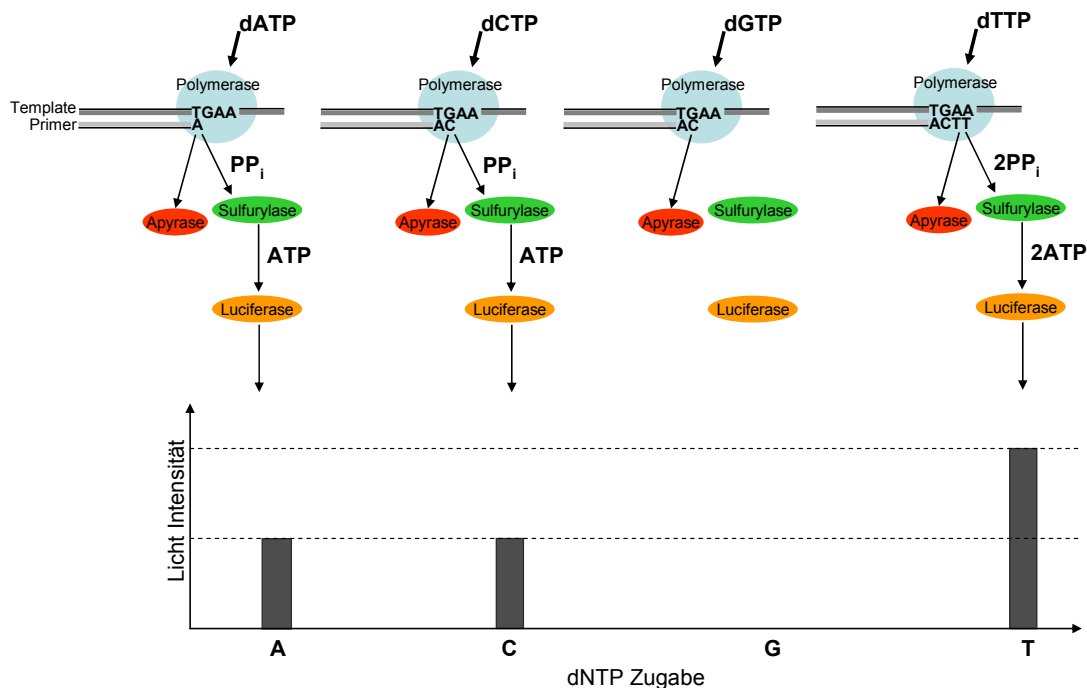


Abbildung 2: Das Prinzip der Pyrosequencing-Technologie (Ahmadian et al. 2006, modifiziert).

Die PCR-Amplifikation erfolgt mit einem biotinyliertem Primer, dadurch kann anschließend der nicht-biotinylierte DNA-Strang mit Hilfe von Streptavidin-Sacharose-Beads vollständig entfernt werden. Bei der darauf folgenden Minisequenzierung bindet ein spezifischer Primer an die ssDNA (single stranded DNA) und nacheinander werden die dNTPs in bestimmter

Reihenfolge dazugegeben. Beim Einbau eines komplementären dNTPs durch die Polymerase wird ein Pyrophosphat (PPi)-Molekül frei, das der ATP-Sulfurylase zur Erzeugung eines ATP-Moleküls dient. Dieses wandelt Luciferase in ein Lichtsignal um, dessen Intensität mit der Menge der eingebauten Nukleotide korreliert. Überschüssige dNTPs werden durch Apyrase abgebaut und der Vorgang kann mit anderen dNTPs wiederholt werden (Abbildung 2).

### 1.3.3.5 SNPplex™/Genplex

Die SNPlex™-Methode (Applied Biosystems) ist ein Genotypisierungsassay, welches bis zu 48 SNPs und unter anderem auch Insertion/Deletion automatisch typisieren kann. Die SNPlex™-Genotypisierung ist ein Multiplex Oligonucleotid Ligation-Assay (OLA), das 1998 von Landegren als eine Methode beschrieben wurde, welche auf der kovalenten Bindung zwischen zwei benachbart hybridisierten Oligonukleotiden basiert. Mittels einer thermostabilen DNA-Ligase entstehen linear in einer zyklischen Reaktion die Ligationsprodukte (Abbildung 3). Für das SNPlex™-Verfahren werden die benötigten Kits für die einzelnen Reaktionsschritte mit allen Komponenten wie Sonden, Reagenzien, Enzymen und Puffern von der Firma Applied Biosystems geliefert.

Nachdem die zu untersuchende genomische DNA denaturiert und einzelsträngig vorliegt folgt die Multiplex-OLA-Reaktion für alle SNPs in einer Multiplexreaktion. Für die OLA-Reaktion werden pro SNP drei Sonden benötigt: zwei ASOs (Allele Specific Oligos) mit einer spezifischen ZipCode-Sequenz und einer Sequenz komplementär zu dem universellen forward PCR-Primer und eine LSO (Locus Specific Oligo)-Sonde mit einer Sequenz komplementär zu dem universellen reverse PCR-Primer. Bei der OLA-Reaktion erfolgt die Phosphorylierung mit 5' Phosphat und die Ligation der Sonden parallel für alle 48 SNPs. Die SNP-flankierenden Bereiche der gereinigten Ligationsprodukte enthalten auf einer Seite die komplementären Sequenzen für einen universellen forward PCR-Primer und den ZipCode und auf der anderen eine komplementäre Sequenz für den universellen reverse PCR-Primer. Die so gewonnenen Ligationsprodukte werden im nächsten Schritt mittels eines universellen forward Primers und eines biotinylierten universellen reverse Primers amplifiziert. Anschließend werden die biotinylierten Stränge an eine streptavidinbeschichtete Platte gebunden und stehen für die ZipChute-Hybridisierung bereit. Die ZipChute-Sonde ist komplementär zu dem ZipCode-Bereich des Ligationsproduktes, enthält spezielle mobility modifier und ist fluoreszenzmarkiert. Durch die Modifizierung können die eluierten ZipChute-Sonden elektrophoretisch in einer Kapillarelektrophorese aufgetrennt und durch die Data Collection Software die Fluoreszenzintensitäten gemessen werden.

Die Genplex-Genotypisierung ist ein etwas verändertes OLA-basiertes SNPlex™-Assay, welches von Applied Biosystems nach Austausch mit dem SNPforID-Konsortium entwickelt wurde. Während beim SNPlex™-Assay die OLA-Reaktion der PCR-Amplifikation vorangestellt ist, wurde das Genplex-Assay in der Art modifiziert, dass vor der Ligationsreaktion eine Multiplex-Präamplifikation stattfindet (Abbildung 3). Das Genplex-Assay startet mit einer Multiplex-PCR-Amplifikation und enzymatischer Aufreinigung der PCR-Produkte. Danach erfolgt die OLA-Reaktion. Das Ziel war damit die Sensitivität der

OLA basierten Genotypisierung so zu erhöhen, dass viel weniger als 37ng Ausgangs-DNA (empfohlen für SNplex™) und damit forensisch relevantes Material für eine effiziente Typisierung ausreichend sind. Diese in Genplex umbenannte Methode wurde in Santiago de Compostela, London und Köln validiert (Phillips et al. 2007, Musgrave-Brown et al. 2008).

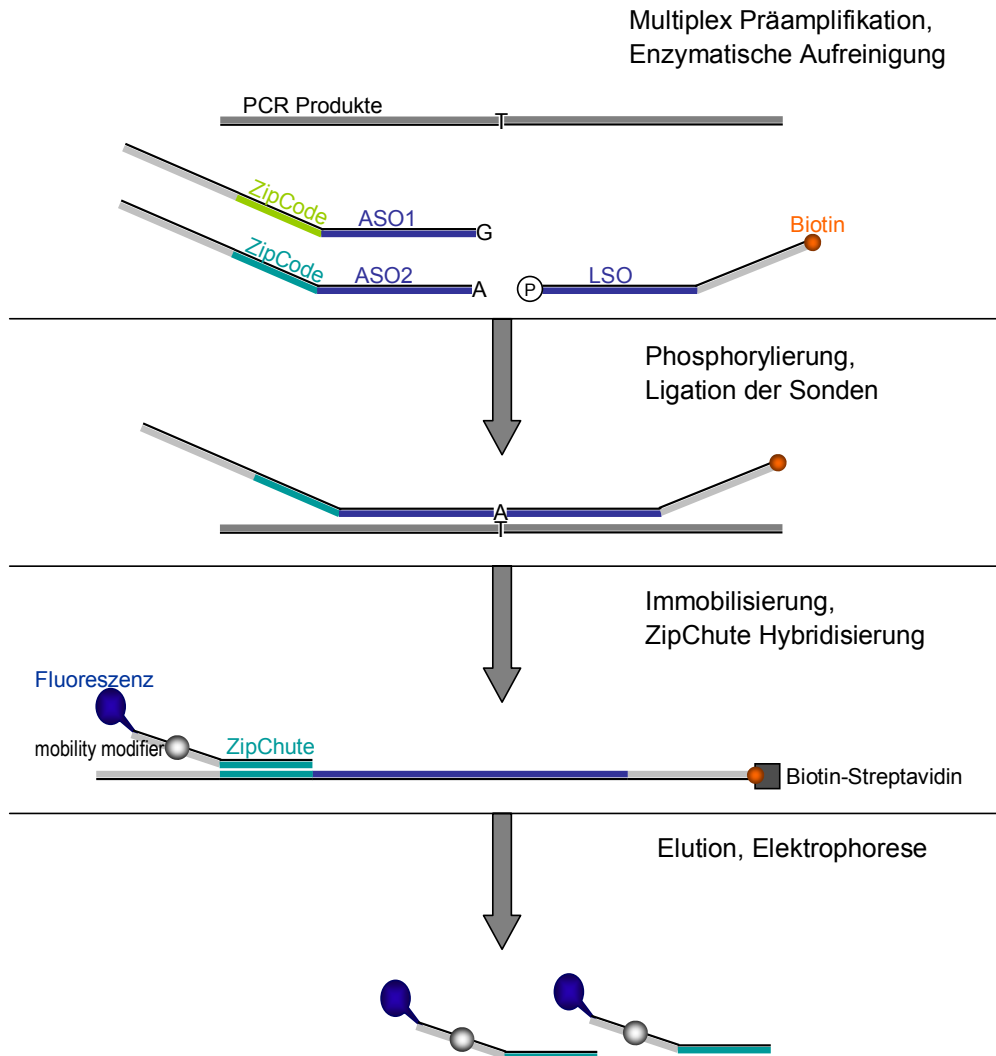


Abbildung 3: Das Prinzip der Genplex-Genotypisierungsmethode.

## 1.4 Fragestellung und Zielsetzung der Arbeit

Die Fragestellung der vorliegenden Dissertation umfasst zwei Aufgabengebiete. Dabei stellt das erste Teilgebiet mit dem Schwerpunkt Validierung die notwendige Grundlage für den eigenständigen Hauptteil zum Thema Microarray-Technologie dar.

Zunächst sollten nach einer SNP-Selektion und Etablierung von Multiplexstrategien die Gruppen der autosomalen und Y-chromosomalen SNPs validiert werden. Diese Aufgabe sollte in zwei systematischen Studien mit Beteiligung aller Arbeitsgruppen des Projektes unter Anwendung einer ausgewählten Referenzmethode gelöst werden. Für die Validierung der Y-chromosomalen SNPs wurden die einzelnen Laboratorien aufgefordert repräsentative DNA-Proben aus eigenen Kollektiven mittels der Referenzmethode zu genotypisieren und durch die ermittelten Frequenzdaten die Haplogruppenzuordnung zu bestimmen. Die Validierung der autosomalen SNPs sollte anhand forensisch relevanter Materialien, wie degradierter DNA, Proben von geringer Quantität und schlechter Qualität, oder Mischproben erfolgen. Die Aufgabe dabei war die Vorteile oder mögliche Defizite des Verfahrens aufzuzeigen und zu beurteilen.

Die zentrale Zielsetzung dieser Dissertation war die Adaptation der zuvor validierten autosomalen SNPs an die elektronische Microarray-Plattform der Firma Nanogen. In erster Linie sollte ein Hybridisierungsassay für diese Plattform neu entwickelt und etabliert werden, welches sich zur Lösung spezieller Aufgaben eignet. Dabei sollten nachhaltig Akzente auf Multiplexfähigkeit, Effizienz und Schnelligkeit des Assays gesetzt werden, sowie auf die Eliminierung von Fehlerquellen, die durch manuelle Handhabung entstehen können. Darüber hinaus war die Optimierung des für die Nanogen-Plattform neu etablierten Assays durchzuführen. Hierbei sollte die Genauigkeit und Zuverlässigkeit dieser Technologie verbessert werden, um im Anschluss die Assayentwicklung und die elektronische Microarray-Technologie mittels repräsentativer DNA-Proben aus dem Hauskollektiv zu validieren. Die Validierung sollte nicht nur die Hochdurchsatzkapazität der Methode bewerten, sondern auch die Vor- und Nachteile der NanoChip™-Microarray-Technologie gegenüber anderen SNP-Genotypisierungsmethoden verdeutlichen.



## 2 Material und Methoden

### 2.1 Untersuchungsmaterial

#### 2.1.1 Einzelnukleotid-Polymorphismen

Als Grundlage für die Selektion der Y-chromosomalen SNPs dienten populationsbezogene Veröffentlichungen (Rosser et al. 2000, Semino et al. 2000, Underhill et al. 2000) und die von dem Y-Chromosome Consortium (YCC) veröffentlichten 218 SNPs (Underhill et al. 2001). Die davon ausgewählten 29 SNPs bestimmen 31 Haplogruppen. Zwölf SNPs definieren die Haupthaplogruppen und die restlichen unterteilen diese in Untergruppen. Abbildung 4 ordnet sie dem phylogenetischen Baum zu. Als Ausdruck der damit verbundenen Erwartungen erhielt die Gruppe der 29 Y-chromosomalen SNPs den Namen Major Y-Chromosome Haplogroup Typing Kit (Brion et al. 2005).

Andererseits mussten für die Validierung der Genotypisierungsmethoden autosomale Marker bereitgestellt werden. Dazu wurden zunächst 330 autosomale Kandidaten-SNPs ausgewählt. Für die SNP-Selektion und für die Erforschung ihrer Eigenschaften wurden zwei kommerzielle<sup>3</sup> und zwei öffentliche<sup>4</sup> SNP-Datenbanken benutzt (Phillips et al. 2004a). Die Kriterien für die Selektion der Kandidaten-SNPs können in vier aufeinanderfolgenden Schritten beschrieben werden:

1. Ein SNP muss in den Datenbanken durch ein hohes und zuverlässiges Validierungsstadium charakterisiert sein und ein einzigartiges Vorkommen im humanen Genom aufweisen.
2. Um einen hohen Diskriminierungsindex (Dp) zu garantieren, müssen die SNPs eine  $\geq 30\%$ -ige Minor Allele Frequency (MAF) in mindestens zwei der drei durch Frequenzdaten belegten Populationen erreichen (Miller et al. 2005).
3. Eine Kopplung zu anderen SNPs oder zu anderen benachbarten Markern, einschließlich der etablierten STR-Systeme, muss ausgeschlossen sein.
4. Die SNP-flankierenden Sequenzen müssen anwendungsbezogene Qualitäten erfüllen, d.h. zum Beispiel keine repetitiven und im Verhältnis keine übermäßig reichen GC- bzw. AT-Sequenzen enthalten.

Mittels angewandter Kriterien wurde schließlich in der nächsten Stufe der Etablierung von Multiplexstrategien eine Gruppe von 52 autosomalen SNPs herausgefiltert. Tabelle 3 stellt die 52 ausgewählten autosomalen SNPs vor und liefert Angaben aus den Datenbanken zu ihren Chromosomenlokalisationen, Kennungsnummern bei den NCBI, dbSNP und Celera, TSC Datenbanken, möglichen Allelen und Frequenzdaten für drei verschiedene Populationen.

---

<sup>3</sup> Celera Discovery System, <http://www.vanderbilt.edu/bioinformatics/Celera/index.html> (Kerlavage et al. 2002) und Assays-on-Demand, [www.appliedbiosystems.com/products/assay](http://www.appliedbiosystems.com/products/assay).

<sup>4</sup> dbSNP, National Center of Biotechnology Information, [www.ncbi.nlm.nih.gov/entrez](http://www.ncbi.nlm.nih.gov/entrez) und TSC, The SNP Consortium, <http://snp.cshl.org>, (Thorisson und Stein 2003).

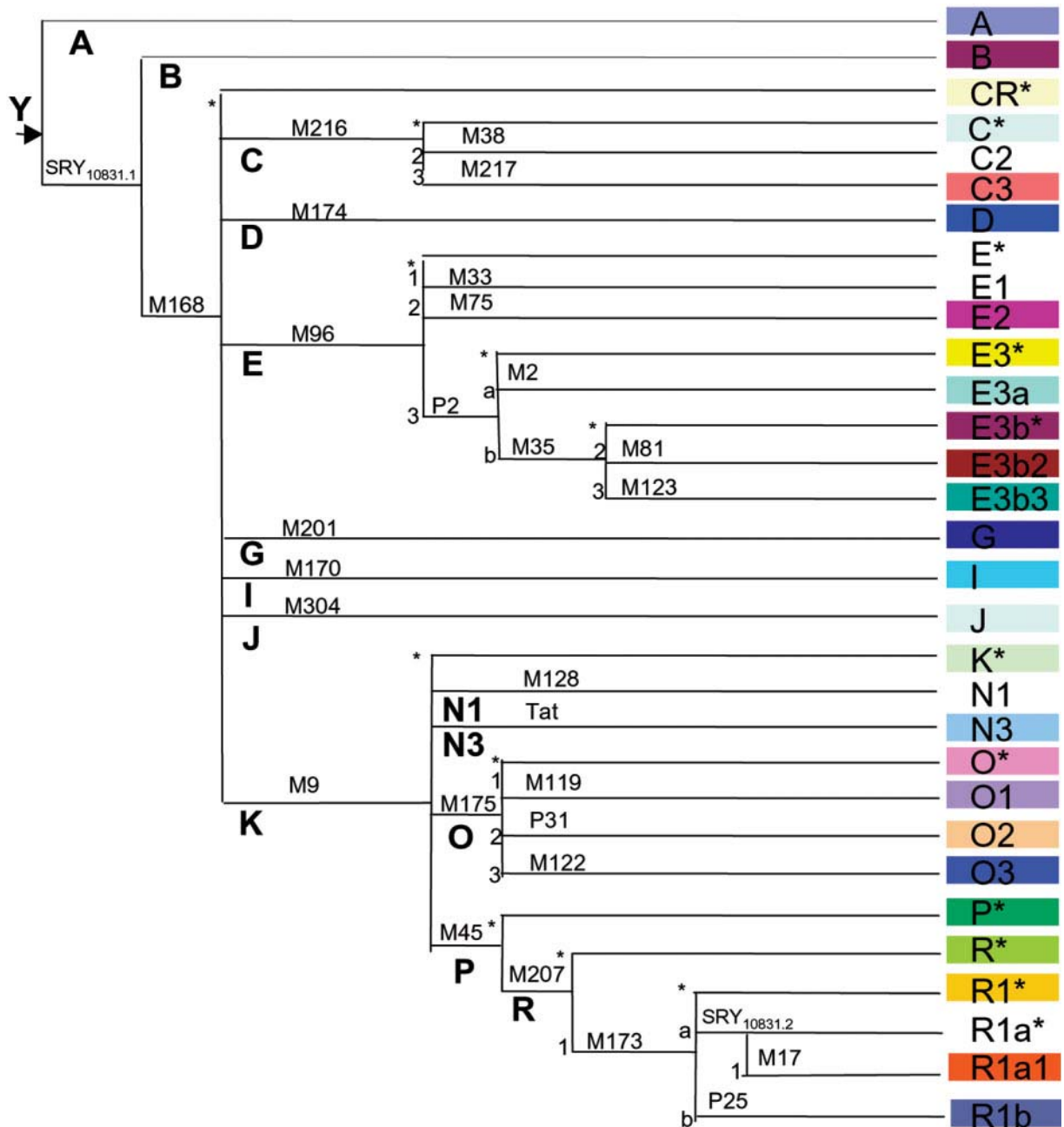


Abbildung 4: Die phylogenetische Verteilung der 31 Y-chromosomalen Haplogruppen, die durch die 29 binären SNPs des Major Y-Chromosome Haplogroup Typing Kits detektiert sind. Die Namen der Marker stehen oberhalb der Linien und weisen auf die zugehörige Haplogruppe auf der rechten Seite hin (Brion et al. 2005).

Tabelle 3: Die 52 autosomalen SNPs, sortiert nach ihren Namen. Informationen zu ihren Chromosomenlokalisationen, Datenbank-Nummern, Allelen und Frequenzdaten sind aus den Datenbanken TSC, Celera oder dbSNP, NCBI entnommen (Sanchez et al. 2006b, modifiziert).

Name	Ch.	NCBI Nr.	TSC Nr.	Allele	Amerika	Asien	Europa
SNP01	1	rs1490413	TSC0724193	A/G	0,59/0,41	0,46/0,54	0,47/0,53
SNP02	2	rs876724	TSC0208870	C/T	0,72/0,28	0,46/0,54	0,69/0,31
SNP03	3	rs1357617	TSC0496080	A/T	0,85/0,15	0,80/0,20	0,65/0,35
SNP04	4	rs2046361	TSC1065282	A/T	0,65/0,35	0,70/0,30	0,80/0,20
SNP05	5	rs717302	TSC0039610	A/G	0,92/0,08	0,83/0,17	0,45/0,55
SNP06	6	rs1029047	TSC0253802	A/T	0,61/0,39	0,27/0,73	0,37/0,63
SNP07	7	rs917118	TSC0229630	C/T	0,10/0,90	0,50/0,50	0,50/0,51
SNP08	8	rs763869	TSC0065968	C/T	0,48/0,52	k.A.	0,56/0,44
SNP09	9	rs1015250	TSC0097236	C/G	0,55/0,45	0,40/0,60	0,15/0,85
SNP10	10	rs735155	TSC0027519	A/G	0,38/0,62	0,82/0,18	0,41/0,59
SNP11	11	rs901398	TSC0177752	C/T	0,39/0,61	0,29/0,71	0,34/0,66
SNP12	12	rs2107612	TSC1108144	A/G	0,64/0,36	0,75/0,25	0,70/0,30
SNP13	13	rs1886510	TSC0904551	C/T	0,85/0,15	k.A.	0,50/0,50
SNP14	14	rs1454361	TSC0684657	A/T	0,57/0,43	0,62/0,38	0,47/0,53
SNP15	15	rs2016276	TSC0326920	A/G	0,91/0,09	0,59/0,41	0,79/0,21
SNP16	16	rs729172	TSC0028090	A/C	0,23/0,77	0,06/0,94	0,38/0,62
SNP17	17	rs740910	TSC0105771	A/G	0,92/0,08	0,92/0,09	0,57/0,43
SNP18	18	rs1493232	TSC0729796	A/C	0,44/0,56	0,39/0,61	0,56/0,44
SNP19	19	rs719366	TSC0044147	C/T	0,24/0,76	0,15/0,85	0,38/0,62
SNP20	20	rs1031825	TSC0334834	A/C	0,60/0,40	0,41/0,59	0,30/0,70
SNP21	21	rs722098	TSC0050288	A/G	0,17/0,83	0,50/0,50	0,86/0,14
SNP22	22	rs733164	TSC0023085	A/G	0,26/0,74	0,16/0,84	0,34/0,66
SNP23	10	rs826472	TSC0557086	C/T	0,70/0,30	0,66/0,34	0,44/0,56
SNP24	21	rs2831700	k.A.	A/G	0,42/0,58	0,50/0,50	0,62/0,38
SNP25	14	rs873196	TSC0202434	C/T	0,30/0,70	0,19/0,81	0,37/0,63
SNP26	16	rs1382387	TSC0544547	G/T	0,55/0,45	0,40/0,60	0,25/0,75
SNP27	12	rs2111980	TSC1113476	A/G	0,69/0,31	0,60/0,40	0,56/0,44
SNP28	8	rs2056277	TSC1082757	C/T	0,70/0,30	0,75/0,25	0,55/0,45
SNP29	18	rs1024116	TSC0247167	A/G	0,37/0,63	0,12/0,88	0,50/0,50
SNP30	6	rs727811	TSC0062764	A/C	0,30/0,70	0,66/0,34	0,61/0,39
SNP32	1	rs1413212	TSC0607362	A/G	0,61/0,39	0,47/0,53	0,41/0,59
SNP33	17	rs938283	TSC0357388	C/T	0,09/0,91	0,17/0,83	0,30/0,70
SNP34	4	rs1979255	TSC0925231	C/G	0,38/0,62	0,33/0,67	0,35/0,65
SNP35	9	rs1463729	TSC0377760	A/G	0,79/0,21	0,48/0,52	0,56/0,44
SNP36	11	rs2076848	TSC0022275	A/T	0,38/0,62	0,58/0,42	0,49/0,51
SNP37	3	rs1355366	TSC0491662	A/G	0,40/0,60	0,95/0,05	0,70/0,30
SNP38	2	rs907100	TSC0186810	C/G	0,74/0,26	k.A.	0,48/0,52
SNP39	13	rs354439	TSC0700528	A/T	0,50/0,50	0,58/0,42	0,64/0,36
SNP40	22	rs2040411	TSC1056845	A/G	0,75/0,25	0,08/0,92	0,71/0,29
SNP41	7	rs737681	TSC0033074	C/T	0,56/0,44	0,80/0,20	0,58/0,42
SNP42	21	rs2830795	k.A.	A/G	0,88/0,12	0,45/0,55	0,67/0,33
SNP43	5	rs251934	TSC0220872	C/T	0,26/0,74	0,16/0,84	0,42/0,58
SNP44	21	rs914165	TSC0197658	A/G	0,66/0,34	0,27/0,73	0,49/0,51
SNP45	1	rs10495407	k.A.	A/G	0,11/0,89	0,42/0,58	0,29/0,71
SNP46	9	rs1360288	TSC0501229	C/T	0,70/0,30	0,50/0,50	0,50/0,50
SNP48	10	rs964681	TSC0270699	C/T	0,37/0,63	0,23/0,77	0,41/0,59
SNP49	20	rs1005533	TSC0082071	A/G	0,55/0,45	0,60/0,40	0,45/0,55
SNP50	15	rs8037429	k.A.	C/T	0,58/0,42	0,42/0,58	0,62/0,38
SNP51	1	rs891700	TSC0162577	A/G	0,53/0,47	0,53/0,47	0,57/0,43
SNP52	13	rs1335873	TSC0829150	A/T	0,91/0,09	0,28/0,72	0,36/0,64
SNP53	22	rs1028528	TSC0253071	A/G	0,35/0,65	0,65/0,35	0,74/0,26
SNP54	15	rs1528460	TSC0798410	C/T	0,65/0,35	0,36/0,64	0,29/0,71

k.A. keine Angabe

## 2.1.2 Probenauswahl zur Y-chromosomalen Validierungsstudie

Für die Validierung der Y-chromosomalen SNPs wurden insgesamt 1126 DNA-Proben von nicht verwandten Männern aus zwölf Populationen weltweit (Deutschland n=150, Dänemark n=150, Grönland n=90, Türkei n=51, Thailand n=84, China n=55, Japan n=40, Galizien n=130, Mosambik n=130, Somalia n=105, Argentinien n=81 aus verschiedenen ethnischen Mischungen und Kolumbien n=60 aus verschiedenen ethnischen Mischungen) typisiert. Die Proben aus Deutschland, China, Japan und 40 Proben aus Thailand stammen aus dem Hauskollektiv der Rechtsmedizin, Johannes Gutenberg-Universität Mainz und wurden auch vor Ort analysiert. Die restlichen Proben wurden zwischen den Laboratorien der Rechtsmedizinischen Institute in Santiago de Compostela und Kopenhagen aufgeteilt. Die Proben aus Deutschland wurden aus Vaterschaftsbestimmungen ausgewählt und stellen ein Kollektiv aus nicht verwandten männlichen Personen aus dem Südwesten Deutschlands dar. Die Proben aus Chengdu/China (Zhang et al. 1994), aus Bangkok/Thailand und vom Miyazaki Medical College/Japan wurden für Forschungszwecke seit 1994 im Hauskollektiv gelagert (Schneider et al. 1998). Die genomische DNA wurde nach dem Standardprotokoll, wie von Miller et al. 1988 beschrieben, aus Vollblut extrahiert. Die Proteinbestandteile wurden mittels Proteinase-K (Roche Diagnostics) und 6M NaCl sedimentiert und die DNA im Überstand durch Zugabe von Ethanol präzipitiert. Diese DNA-Extrakte, die über Jahre bei  $-20^{\circ}\text{C}$  lagerten, wurden mit Hilfe eines Fluorometers quantifiziert (Kapitel 2.2.1) und für die Amplifikation auf 20ng/ $\mu\text{l}$  verdünnt. Einige japanische Proben erreichten keine messbaren Konzentrationen, aus diesem Grund wurden sie vor der PCR-Reaktion mittels Whole Genome Amplification (WGA) vervielfältigt (Kapitel 2.2.2).

## 2.1.3 Probenauswahl zur forensischen Validierungsstudie

Für die Validierungsstudie der 52 autosomalen SNPs wurden insgesamt 54 DNA-Proben, gemäß forensischer Herausforderungen, vorpräpariert. Qualitativ gute Standard-DNAs (Standard # 9947a und # 9948 von Promega) wurden für die Sensitivitätstests zu Konzentrationen von 1000, 500, 200, 100, 50 und 20pg/ $\mu\text{l}$  verdünnt. Aus den selben Standards wurden Mischproben von 1ng/ $\mu\text{l}$  Endkonzentration und Verhältnissen von 1:10, 1:5, 1:2 und 1:1 hergestellt. Die sog. Blindproben stellten ebenfalls Mischproben im Verhältnis von 1:10, 1:5, 1:2 und 1:1 der gleichen Standard-DNAs dar, jedoch ohne das Wissen der Bearbeiter. Die qualitativ schlechten Proben setzten sich zusammen aus Extrakten von telogenem Haar (keine Konzentrationsangabe), zersetztem Muskelgewebe (22-135pg/ $\mu\text{l}$ ) und natürlich degradiertem Material. Letzteres bestehend aus Proben, die gemäß ihrer Behandlung forensische Situationen abbilden, indem eine Umgebung für eine natürliche bakterielle Degradierung geschaffen wurde. Die Herstellung dieser Proben erfolgte, indem 2ml Blut auf einem Tupfer drei Monate lang bei  $37^{\circ}\text{C}$  inkubiert oder eine Speichelprobe frei bei Raumtemperatur einen Monat lang gelagert und periodisch mit  $\text{dH}_2\text{O}$  benässt wurde. Die natürlich degradierten Extrakte lagen nur noch in sehr geringen Konzentrationen von unmessbar bis 0,35ng/ $\mu\text{l}$  vor. Damit verbanden diese Proben zwei kritische Probleme der forensischen Spurenkunde: schlechte Qualität durch Degradierung

und geringe Quantität. Um die Auswirkung der geringen Quantität auf die Typisierung zu überprüfen wurden zwei Proben mit Ausgangskonzentrationen von 0,08 bzw. 0,21ng/µl mit Hilfe des sog. Speed-vac-Verfahrens konzentriert. Dazu wurde eine Zentrifuge benutzt, an die ein Vakuum angelegt werden konnte. Das Vakuum führt zur Siedepunktniedrigung und damit zu einem raschen Verdampfen der Flüssigkeit. Nach dem Eintrocknen wurden die Pellets in HPLC-Wasser gelöst und die Proben lagen in Konzentrationen von 2ng/µl vor. Zusätzlich wurden bei der forensischen Validierung künstlich degradierte Proben untersucht, die eine Länge zwischen 200 und 1000bp und Konzentrationen von 2-50ng/µl aufwiesen. Die künstlich degradierten Proben wurden in Mainz präpariert und das Herstellungsverfahren ist im Kapitel 2.2.3 genau geschildert.

Die 54 forensisch relevanten Proben wurden unter den beteiligten Arbeitsgruppen so aufgeteilt, dass jede Probe von mindestens zwei Arbeitsgruppen analysiert wurde. Die in Mainz genotypisierten und ausgewerteten Proben sind in der Tabelle 4 mit Angaben zu ihren DNA-Konzentrationen aufgelistet.

Tabelle 4: In Mainz analysierte Proben zur forensischen Validierungsstudie.

Name	Anzahl	Probeneigenschaften	DNA [ng/µl]
<b>Ld</b>	20	natürlich degradierte DNA	0 <sup>1)</sup> - 0,35
<b>St</b>	14	Verdünnungen aus Standard-DNA	0,02 - 10
<b>Mix</b>	8	Mischproben aus Standard-DNA <sup>2)</sup>	1
<b>Blindproben</b>	4	Mischproben aus Standard-DNA <sup>2)</sup>	1

1) DNA-Konzentration unmessbar niedrig.

2) Mischungsverhältnisse 1:10; 1:5; 1:2, 1:1.

## 2.1.4 Probenauswahl zur Microarray-Typisierung

Zur Assayetablierung und Methodenoptimierung für die elektronische Microarray-Typisierung wurden drei DNA-Proben mit bereits bekannten Genotypen für die 52 autosomalen SNPs aus dem Hauskollektiv verwendet. Bei der Auswahl der drei Proben wurde darauf geachtet, dass mindestens zwei der drei möglichen Genotypen eines SNPs (Homozygot-Wildtyp, Heterozygot oder Homozygot-Mutante) vertreten sind. Dies gewährleistete, dass die Diskriminierungsfähigkeiten der einzelnen SNPs überprüft werden konnten, was bei der Assayoptimierung von großer Bedeutung war.

Nach der Optimierung wurde die elektronische Microarray-Methode mittels 40 DNA-Proben aus dem Hauskollektiv validiert. Als Maßnahme zur Qualitätssicherung für die Bestimmung der Genauigkeit und Zuverlässigkeit der Microarray-Plattform wurden parallel die gleichen Proben auch mittels der validierten SNaPshot™-Referenzmethode genotypisiert.

## 2.2 Methoden

### 2.2.1 Probenquantifizierung

Die DNA-Extrakte wurden mittels Fluorometer (VersaFluor™, BioRad) quantifiziert. Für die Messung wurden zwei Filter, EX 480/20 (470-490nm) und EM 520/10 (515-525nm) verwendet. Die Arbeitslösung aus PicoGreen (Molecular Probes/Invitrogen life technologies) und 1xTE Puffer pH 7,5 [1:400] wurde vor der Verwendung kurz im Dunkeln inkubiert. Vor der DNA-Messung wurde der Nullwert des Gerätes auf die reine Arbeitslösung eingestellt. Im Anschluss wurde das Fluorometer mit einem Gemisch aus Standard-DNA (Calf Thymus DNA 100ng/μl, Sigma-Aldrich) und Arbeitslösung [1:100] auf 1000 geeicht. Anschließend wurde auch die zu messende DNA mit der Arbeitslösung [1:100] verdünnt, in die Messküvette überführt und die Messung vollzogen. Der angezeigte Display-Wert geteilt durch 10 ergibt die DNA-Konzentration in ng/μl. Eine Verfeinerung der Messung konnte erzielt werden, indem die oben genannte Standard-DNA im Verhältnis 1:10 verdünnt wurde, sodass die Eichung mit Standard-DNA 10ng/μl durchgeführt wurde. Bei der darauf folgenden DNA-Messung ergibt der angezeigte Wert geteilt durch 100 die DNA-Konzentration in ng/μl.

### 2.2.2 Whole Genome Amplification

Einige Proben aus dem Hauskollektiv wiesen für die Standard-Amplifikation eine zu geringe Menge an DNA auf. Aus diesem Grund wurden diese Proben vor der spezifischen PCR mittels einer unspezifischen Präamplifikation namens Multiple Displacement Amplification (SDA) oder Whole Genome Amplification (WGA) vervielfältigt.

Bei der unspezifischen Präamplifikation, die auf einem isothermalen, linearen Reaktionsprinzip beruht, kommt es zu einer ständigen Strangverdrängung (Dean et al. 2002, Hosono et al. 2003). Dabei entstehen kontinuierlich ohne Denaturierung, neue Einzelstränge, die als Ausgangsmaterial für neue Replikationen dienen (Abbildung 5). Da bei der WGA keine zyklischen Denaturierungsschritte notwendig sind, kann hierfür die effiziente, nicht thermostabile, Bakteriophage Phi29 DNA-Polymerase verwendet werden, welche bei einer konstanten Temperatur von 30°C aktiv bleibt. Nach der Reaktion kann die vervielfältigte DNA ohne Aufreinigung für spezifische Amplifikationen verwendet werden (Pask et al. 2004, Tzvetkov et al. 2005). Für die WGA-Reaktion wurde der kommerzielle GenomiPhi™ Kit (Amersham Biosciences) laut Herstellerprotokoll verwendet. Dementsprechend wurde 1μl Template mit 9μl Sample Buffer, der die Random Hexamer Primer enthält, vermischt, drei Minuten bei 95°C denaturiert und anschließend auf Eis abgekühlt. Parallel dazu wurde der Enzym-Mix auf Eis hergestellt, indem 9μl Reaction Mix, bestehend aus Puffer und dNTPs, mit 1μl Phi29 Polymerase vermischt wurden. Danach wurde der Enzym-Mix zu dem Template dazugegeben und 16 Stunden bei 30°C inkubiert. Das Enzym wurde anschließend 10 Minuten bei 95°C inaktiviert. Nach der Abkühlung der Probe wurde die DNA-Konzentration mittels einer fluorometrischen Quantifizierung (Kapitel 2.2.1) ermittelt und auf 20ng/μl verdünnt.

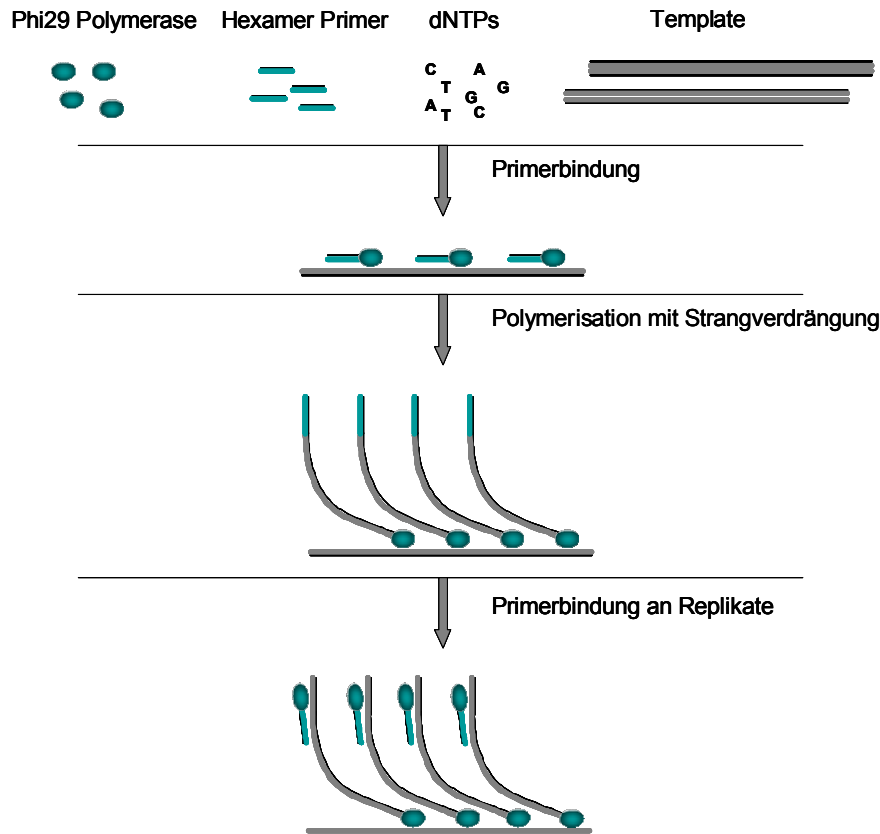


Abbildung 5: Schematische Darstellung des WGA-Prozesses. Hexamer Primer binden zufällig an multiple Stellen der einzelsträngigen DNA. Die multiple Replikation mittels Phi29 Polymerase beginnt. Während der Polymerisation kommt es zur Strangverdrängung, wobei kontinuierlich einzelsträngige DNA synthetisiert werden. Folglich binden erneut die Hexamer Primer an die Replika und der Strand Displacement Amplification Prozess setzt sich fort.

### 2.2.3 Künstliche DNA-Degradierung

Für die künstliche Degradierung wurde DNA verwendet, die aus der männlichen Hepatomen Zelllinie HepG2 mittels Standard Phenol/Chloroform/Isoamylalkohol 25:24:1 (Carl Roth) Prozedur nach Proteinase-K Verdau [10mg/ml] (Roche Diagnostics) extrahiert wurde. Die fluorometrische Konzentrationsbestimmung, Verdünnung auf 500ng/μl und RNase-A Verdau [100μg/ml] für eine Stunde bei 37°C, gehörten zur DNA-Vorbereitung.

Die künstliche DNA-Degradierung erfolgte mittels Ultraschall-Sonifikation durch Branson Sonifier B-12/15 (Heinemann) bei einer Abgabeleistung 4 und einem Minimalausschlag um 20 (Bender 2004). Es wurden zwei unterschiedliche Degradierungsstufen präpariert; d.h. eine Probe wurde nur 10 Sekunden und die zweite Probe 40 Sekunden im Dreier-Intervall mit 40 Sekunden Abkühlphase dazwischen sonifiziert. Während des Beschallungsvorgangs befanden sich die Proben in -20°C kühlem Alkohol 30%. Um DNA-Fragmente mit kontrollierter Basenpaarlänge herzustellen, wurden die Proben nach der Sonifikation elektrophoretisch aufgetrennt und aufgrund bestimmter Längen aus dem Gel extrahiert. Hierfür wurde die Elektroelutionsmethode mittels eines Dialyseschlauchs verwendet. Je 50μl der Sonifikationsproben wurden auf ein TAE-Agarosegel 1% mit Ethidiumbromid aufgetragen

und elektrophoretisch bei 100V aufgetrennt. Unter UV-Licht konnten mit Hilfe einer DNA-Fragmentlängenleiter (Ready-Load™ 100bp DNA Ladder, Invitrogen life technologies) vier Bereiche aus dem Gel geschnitten werden. Die Gelstücke wurden in einen Dialyseschlauch gegeben, mit 1xTAE-Puffer bedeckt und der Schlauch luftblasenfrei verschlossen. Unter einer 50V Elektrophorese wanderten die DNA-Fragmente in zwei Stunden vollständig aus dem Gel in den Puffer. Nachdem die Polarität für eine Minute umgedreht wurde, damit sich die DNA von der Schlauchwand löst, wurde der Puffer mit den DNA-Fragmenten in einem Gefäß aufgefangen. Anschließend erfolgte eine Ethanolfällung. Dafür wurden 500µl Probe mit 1ml EtOH 99,8% (-20°C) und 100µl NaAcetat 3M gemischt und für 10 Minuten bei Raumtemperatur inkubiert. Nach 25-minütiger Zentrifugation bei 13000rpm wurde der Überstand verworfen und 300µl EtOH 70% dazugegeben. Ohne zu mischen erfolgte ein dreiminütiger Zentrifugationsschritt und der Überstand wurde vollständig entfernt. Vor der Elution in 60µl HPLC-Wasser wurde die Probe 45 Minuten bei 37°C getrocknet. Die DNA-Konzentration der Eluate wurde mit Hilfe des Fluorometers bestimmt (Kapitel 2.2.1). Die vier Proben wurden mittels SGMplus™ STR-Kit (Applied Biosystems) nach Standardprotokoll, mit 2ng DNA in 25µl Endvolumen mit 28 Zyklen amplifiziert und die Fragmentlängen durch kapillarelektrophoretische Auftrennung detektiert.

## 2.2.4 Polymerasekettenreaktion

### 2.2.4.1 Multiplex-PCR-Primerdesign

Nach abgeschlossener Selektion der 52 autosomalen Marker und der 29 Y-chromosomalen Marker (Kapitel 2.1.1) wurden die SNPs enthaltenden DNA-Sequenzen mit Hilfe des Suchmoduls BLAST<sup>5</sup> (Basic Local Alignment Search Tool) der frei zugänglichen Datenbank GenBank® von NCBI identifiziert. Das Design der PCR-Primer erfolgte anschließend mit Hilfe der im Internet frei zugänglichen Programme Primer 3.0 v. 0.2<sup>6</sup> oder Integrated DNA Technologies IDT® SciTools, Oligo Analyzer 3.0<sup>7</sup> computergestützt (Brion et al. 2005 und Sanchez et al. 2006b).

Um die Spezifität und ganz besonders die Anwendbarkeit der Primer in einer Multiplex-PCR für die simultane Amplifikation von 52 bzw. 29 SNPs zu gewährleisten, wurden für das Design aller Primer die gleichen Kriterien befolgt (Butler et al. 2001). Die Primeroligos besitzen eine theoretische Schmelztemperatur um 60°C, (Toleranz  $\pm$  5°C) bei einer Salzkonzentration von 180mM und einem Verhältnis von nahezu 1:1 zwischen Purinen und Pyrimidinen. Die Primerlängen liegen zwischen 18 und 34 Nukleotiden. Um Artefakte zu vermeiden wurden die Primer verworfen, die am 3' Ende vier oder mehr mit sich selbst komplementäre Basen aufwiesen, oder Primer, die mit einem anderen Primer aus dem 52 bzw. 29 Multiplex-Kit interagiert haben und Primer, die Homologien zu repetitiven Sequenzen oder anderen Loci im Genom zeigten.

---

<sup>5</sup> [www.ncbi.nlm.nih.gov/blast/](http://www.ncbi.nlm.nih.gov/blast/)

<sup>6</sup> [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)

<sup>7</sup> <http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/>



Für das Primerdesign der 29 SNPs aus dem Major Y-chromosome Haplogroup Typing Kit dienten die DNA-Sequenzen aus der Datenbank und die bereits veröffentlichten Primer als Ausgangspunkt. Die Primer der 29 SNPs definieren PCR-Fragmente von Längen zwischen 79 und 186bp (Tabelle 5). Um den Erwartungen eines forensisch relevanten Multiplex-Kits gerecht zu werden, wurden die PCR-Produktlängen für die 52 autosomalen SNPs möglichst kurz gewählt (von 59 bis 115bp, Tabelle 6).

Tabelle 5: Die 29 SNPs des Major Y-chromosome Haplogroup Typing Kits, ihre Bestimmungsnummer in den Datenbanken, die Substitutionen und die dazugehörigen Primerpaare mit Konzentrationsangaben für die Multiplex-PCR-Amplifikation und die Längen der PCR-Produkte (Brion et al. 2005, modifiziert).

Locus	Nr. <sup>1)</sup>	Mutation	PCR Primerpaar (5'-3' Ende)		$\mu\text{M}^{2)}$	PCR bp
			forward primer	reverse primer		
<b>M2</b>	rs3893	A>G	acggaaggaggtctaaaattcagg	aaaatacagctccccctttatcct	1,63	128
<b>M9<sup>4)</sup></b>	rs3900	C>G	aggaccctgaaatacagaactg	aaataittcaacatttcacaaaggaa	6,5	186
<b>M17<sup>4)</sup></b>	rs3908	1bp del <sup>6)</sup>	cctggcatacactggaatc	agctgaccacaaactgatgtaga	1,95	170
<b>M33<sup>4)</sup></b>	AC009977	A > C	tgaccgtcatagctgagaca	ttgaagcccccaagagagac	3,58	160
<b>M35<sup>4)</sup></b>	AC009977	G > C	agggcatggtcccttctat	tccatgcagacttccggagt	3,9	96
<b>M38</b>	AC009977	T > G	agatctgttgctactgtacccta	gagctggcacatctgtcataata	1,3	177
<b>M45<sup>4)</sup></b>	rs2032631	G > A	gagagaggatataaaaattggcagt	tgacagtgccacaaaggctc	0,39	138
<b>Tat<sup>4)</sup></b>	AC002531	T > C	tatatggactctgagttagactgtga	ggtgccgtaaaagtgtgaaataatc	0,52	115
<b>M75</b>	AC010889 <sup>5)</sup>	G > A	cttgcaaaagccaaacagatttc	ttgaacagaggcattgtggaat	5,2	120
<b>M81<sup>4)</sup></b>	rs2032640	C > T	catctcttaacaaaagaggtaaattttgtcc	cattgtgttacatggcctataatattcagt	2,6	179
<b>M96<sup>4)</sup></b>	AC010889 <sup>5)</sup>	G > C	tgccctctcacagagcactt	ccaccactctgttgccttg	7,15	143
<b>M119</b>	AC010889 <sup>5)</sup>	A > C	gatgtagaaaaatggtatgggtattcc	gaaatgccaaaggtaaatgactcac	0,78	188
<b>M122<sup>3)</sup></b>	AC010889 <sup>5)</sup>	T > C	ggtgccctttggaatgaataaatc	gaattgcatttctaggtattcagg	0,65	175
<b>M123<sup>3)</sup></b>	AC010889 <sup>5)</sup>	G > A	ggtgccctttggaatgaataaatc	gaattgcatttctaggtattcagg	0,65	175
<b>M128</b>	AC010889 <sup>5)</sup>	2bp del <sup>6)</sup>	gcaaaagagaactgcctctataaaaatc	gcagaattatctacctcttcaaaactg	2,6	139
<b>M168</b>	rs2032595	C > T	gtggagatgtgtggagggtga	cctctatcagaccatggtatctca	2,6	145
<b>M170<sup>4)</sup></b>	rs2032597	A > C	cagctcttattaagttatgtttctatattctgtg	gtcctcattttacagtgagacacaac	1,63	119
<b>M173<sup>4)</sup></b>	rs2032624	A > C	tttcttacaattcaaggcatttag	ctgaaaacaaaacactggcctatca	0,98	81
<b>M174</b>	rs2032602	T > C	gcttctctgaatacctctggagt	tcttcaaggaaaagtgtgc	0,46	94
<b>M175<sup>4)</sup></b>	rs2032678	5bp del <sup>6)</sup>	gatttaactctctgaatcaggccat	ttctactgataccttcttctgttcattc	0,65	79
<b>M201</b>	rs2032636	G > T	tcagatctaataatccagatcaactgag	gttcaaatccatataccagca	0,65	93
<b>M207</b>	AC006376	A > G	ggggcaaatgtaagtcaagc	tcacttcaacctctgttggaa	0,33	83
<b>M216</b>	AC010877	C > T	ccaatggaattttatacccaca	tgacactgctagtatgtatacctgttg	1,63	160
<b>M217</b>	rs2032668	A > C	tctgtttcagatcattctcaattactg	ctgctgtggcttcatcaaaaata	1,95	145
<b>M304</b>	AC007678	A > C	caaagaaaagcaggagagttgtaa	aaacgtcttataccaaaatataccacag	0,65	90
<b>P2</b>	AC010137 <sup>5)</sup>	C > T	agccatctttccttaagagaggt	gcttgagttctctaggttttggga	0,26	104
<b>P25<sup>4)</sup></b>	rs150173	C > A <sup>7)</sup>	tgaccatcacctgggttaaagt	ggcagataaaggtgtcacatcacat	0,26	109
<b>P31</b>	AC002992	T > C	aacaggtaggtgattttgtgtaca	gaccogatatcgtgccatt	5,85	122
<b>SRY10831<sup>4)</sup></b>	rs2534636 <sup>5)</sup>	A > G	tcatccagtccttagcaaccatta	ccacataggtgaacctgaaaatg	1,63	150

- 1) Bezeichnungsnummer in GenBank oder dbSNP.
- 2) Endkonzentration in der 29plex-PCR.
- 3) SNPs auf dem gleichen PCR-Fragment.
- 4) Primer, die von Sanchez et al. 2003 veröffentlicht wurden.
- 5) Komplementärstrang im Vergleich zu den YCC Informationen von 2002.
- 6) Deletion.
- 7) PSV (Paralogous Sequence Variant, Sanchez et al. 2004).

Die Primeroligos wurden zentral von der Firma DNA Technology A/S synthetisiert, durch HPLC (high performance liquid chromatography) gereinigt, ihre Qualität durch MALDI-TOF geprüft und an die Teilnehmer geliefert, um in allen beteiligten Laboratorien die gleichen Konditionen zu sichern.

Tabelle 6: PCR-Primer für die 52 autosomalen SNPs mit Konzentrationsangaben für die Multiplex-PCR-Amplifikation und die Längen der PCR-Produkte (Sanchez et al. 2006b, modifiziert).

Locus	Mutation	PCR Primerpaar (5'-3' Ende)		$\mu\text{M}^{1)}$	PCR bp
		forward primer	reverse primer		
SNP01	A/G	GTGTGGACTGGGCTGATGT	TTCTCACTAGTGCCCTGCTCTG	0,06	68
SNP02	C/T	GCAGGCTCCATTTTATACCACT	GAATATCTATGAGCAGGCAGTTAGC	0,02	83
SNP03	A/T	AGCTGATGCAGACCCTTACAC	GGATAGCTGATAAGAAACATGACCA	0,02	90
SNP04	A/T	CCTATTTGTATGTATCTATTGTCTATGAACG	GTCATTGTTGACACTTCACCTTCTA	0,03	79
SNP05	A/G	CTTTAGAAGGCATATCGTATTAAGTGTG	AACACAGAAAGAGGTTTCATATGTTGG	0,04	86
SNP06	A/T	CATAACGTGGATTGTGACGA	GGAATAAACTGAAGGCTAAAGAAAAG	0,06	100
SNP07	C/T	GCCCTTAGGGTCGGTTC	GTAAGAGATGACTGAGGTCAACGAG	0,03	87
SNP08	C/T	ATCAAGTGCTTTCTGTTGACATTTG	GGCTACTCCCTCATAATGTAATGC	0,09	100
SNP09	C/G	AAGTGATGGAGTTAGGAAAAGAACC	AAGACATTAGGTGGATTCATAGCTG	0,03	95
SNP10	A/G	GGAGAAAACCGGAGAGCTG	GAGTGCACCGAATTCAACG	0,05	100
SNP11	C/T	CTGGGTGCAAACCTAGCTGAATATC	GTTGGAATGTACTAGGCAAGAAACTAA	0,02	70
SNP12	A/G	GAGCATTCTCTTCTGTTAAAATTGC	TGAGTACATTATTCAACTGTTTTGGAG	0,05	93
SNP13	C/T	GTCCTTGCAATCTTTCTACCAGAG	GGATTTTCACAACAACACTTGC	0,01	86
SNP14	A/T	AGGGAAATACACCCTGAGCTG	AGCTGTCCATCATCAGTAAGACAC	0,02	73
SNP15	A/G	TGCATCCCAGCCTCCACT	ATTGTACCTTGCCACTTTGTGTG	0,17	90
SNP16	A/C	CATTAATATGACCAAGGCTCCTCT	ACATTTCCCTCTTGCGGTTAC	0,02	60
SNP17	A/G	GTATAACAGTTTGCTAAGTAAGGTGAGTG	AGATAGGTTGAGTTTTGGCTTTA	0,08	87
SNP18	A/C	CTATTCTCTCTTTTGGGTGCTAGG	CAAACCTGTTATTGTGAGGCCTGT	0,07	59
SNP19	C/T	CCACAGCATCTTTAACTCTTTTATTATCC	GTAAGGACTTATAGTGAGTAAAGGACAGG	0,06	105
SNP20	A/C	CTTATCTTTCCACATTATGGTCCT	AAGATATAATCACTGCTTTCAAGTATGC	0,03	98
SNP21	A/G	GGAAATCACATCTGTTGACAGTAATGA	GGGTAAGAAATATTCAGCACATCC	0,12	81
SNP22	A/G	AGCTTTGAGCCCAAGTGC	CGGCTCAGGAATGTCAGG	0,02	68
SNP23	C/T	TGAATTTTGTCTCTGTTATATTAGTCACC	TGTAATTGAAATTTGTAGGCAATAGAC	0,04	85
SNP24	A/G	GGCTAAACTATTGCCGAGA	TTCCCTAGAACCACAATTATCTGTC	0,04	62
SNP25	C/T	GCATTCAAATCCCAAGTCT	GCAGGAGTTGGAGTCAATCAG	0,03	63
SNP26	G/T	ACGAAGGAGAAACACCTGAACT	TGGAGTACTTAATAAGACGCTGCAT	0,02	69
SNP27	A/G	AGCATCTTGGCAGCATCC	AGCAAGATCTTTGCCAGTGAGT	0,01	72
SNP28	C/T	CCAAACTGGGTGTTAGGGAGAC	TCATTATCTCGTCATACTTCCCTGT	0,03	73
SNP29	A/G	CCATGTGTTCTAATAAAAAGGATTGC	TGGGAAGTGAGCAAAAGTAAATACA	0,03	76
SNP30	A/C	GTGTTCTTTTCTCTTACCAGAAC	GTGAATGAAATCATGAGATTGCTG	0,03	78
SNP32	A/G	AACCTCCTTTGGAACACTGAC	CAACATTCCATTATCCAGGAGAC	0,05	84
SNP33	C/T	CATTGAAGTCTAACCCTAGTACG	GGATGAGGCCCAACCCATA	0,03	85
SNP34	C/G	TCAGAGACTATGGATGGTATTTAGGTC	CATGGAACGTTGGAACCTTG	0,03	86
SNP35	A/G	ACTATCAGTCTCTGCCCTATTCTG	CACATGTGCATGCTTTTGG	0,04	87
SNP36	A/T	GCCTCACCACCAGAAATCAG	GACATCAGAAACTCCCATGAAACT	0,04	89
SNP37	A/G	CCATGATTTCTTGTGGTGAGA	CACATGTGCTTAGGCCACAAC	0,05	90
SNP38	C/G	GGAGTTCTGATAACGATTCTGAAG	ACAGAAAAGAAGCcCAGTTGGA*	0,04	91
SNP39	A/T	GGCTTCTCTTCCCTTATGTATCTC	CAGGTTGCGATAGAAAACAGTGAAT	0,05	93
SNP40	A/G	TCTGGAATGCCAGTCTTTTGT	CAGAACGCCTATGAAAACCAGT	0,03	94
SNP41	C/T	ACATGTGAGGCCATCTCCAC	CCTTACTGTGATGTAGGCACTGTTT	0,03	96
SNP42	A/G	CACCTCTATAGACATAGGACACACCAT	ATCTAGGCTCTGAATCAGGATGAG	0,02	97
SNP43	C/T	AGAGGGCAGTGAGGCTTTAAGTAG	TGCTAGAATCCAGACTTAACTACCAG	0,01	98
SNP44	A/G	AGCAGCAGAGCCTGGATG	AGACCAGTCACCTCTTTTGCAC	0,07	100
SNP45	A/G	AGATCTCCACTTCCCTTGGTTG	CTCCCAAATTTACATTGCCACT	0,05	102
SNP46	C/T	AGACTCTCTGTGTGGCTTTG	GAGGGGGCATCTGTTGAG	0,04	103
SNP48	C/T	GTACCTGGAGGTGATTTCTGTGAG	GTTATGGAGGATTGGAAGAACCAG	0,03	106
SNP49	A/G	GGTTTGTGTGTGAGTGTTCAGAT	CCTTATGCCTCCCCTGAAC	0,03	107
SNP50	C/T	TTCATTTGCTACACCTCCATAGTA	TGCTACGTAAGAGGTCATTGCTATC	0,03	108
SNP51	A/G	TTTTCAGAGGTGGTATTTCTAGCTG	GCTATGACACTCCTTAGAACTATGCAA	0,04	109
SNP52	A/T	GTGGATGATATGGTTTCTCAAGG	TTCAACAAACGTGTGATGCTCT	0,03	110
SNP53	A/G	ACAGCTGATGCCTCCCTGA	GAGGATGAAGGTTAGAGCCAGAC	0,02	113
SNP54	C/T	TCCTGGAGATCAATATTTAGCCTTA	GGGTGACCAGTAGTTCTATGAGC	0,06	115

1) Endkonzentration in der 52plex-PCR.

### 2.2.4.2 Multiplex-PCR-Amplifikation

Der PCR-Ansatz nach Standardprotokoll zur Multiplexamplifikation erfolgte mit einem Volumen von 25µl (Tabelle 7) oder, unter Beibehaltung der Konzentrationen, in doppeltem Volumen. Zur Erzielung möglichst gleicher Produktausbeute für alle SNPs, bei gleichzeitig geringer Entstehung von Primer-Dimeren oder unspezifischen Nebenprodukten, wurden die optimalen Primerkonzentrationen für jeden Marker (Tabelle 5 und Tabelle 6) und die optimale Menge des Magnesiumchlorids empirisch ermittelt (Tabelle 7).

Tabelle 7: Standardprotokoll einer Multiplex-PCR-Amplifikation.

Komponenten (Firma)	Lieferkonzentration	Endkonzentration	Volumen [µl]
10xBuffer (AB)	10x (15mM MgCl <sub>2</sub> )	1x (1,5mM MgCl <sub>2</sub> )	2,5
MgCl <sub>2</sub> (AB)	25mM	6,5mM	6,5
dNTPs (PeqLab)	jeder 10mM	700µM	1,75
Primermix <sup>1)</sup>	jeder 25µM	0,01-7,15µl	13
AmpliAq Gold™ (AB)	5U	2U/25µl	0,4
gDNA	~20ng/µl <sup>2)</sup>	~6ng/25µl	0,3
H <sub>2</sub> O			0,55
<b>Total</b>			<b>25</b>

1) Der Primermix beinhaltet alle forward und reverse Primer für alle SNPs in unterschiedlicher Konzentration.

2) Stellt die empfohlene Konzentration dar.

Für die Y-chromosomale und ebenso für die forensische Validierungsstudie wurden die PCR-Amplifikationen unverändert nach vorgegebenen Primerkonzentrationen erstellt. Abweichungen vom Standardprotokoll fanden nur bei Proben mit einer DNA-Konzentration <20ng/µl statt, indem 3,5µl MgCl<sub>2</sub> [50mM] verwendet wurden und damit ein entsprechend größeres Volumen an DNA eingesetzt werden konnte.

Für die Microarray-Methoden-etablierung wurden die Amplifikationen, angefangen von einzelnen Singleplexen kumulativ bis zu 24plex, in verschiedenen Konstellationen angesetzt. Im Umkehrschluss wurde für die Microarray-Methodenoptimierung von 48 SNPs ausgegangen und je nach Erfolg der einzelnen SNPs Schritt für Schritt auf ein 35plex reduziert. Schließlich stand für die Validierung der elektronischen Genotypisierung ein Multiplex aus 35 autosomalen SNPs fest. Im Rahmen der PCR-Produktkontrolle (Kapitel 2.2.4.3) wurden die 48 SNPs zusätzlich in bestimmte 3plex- bzw. 4plex-Amplifikationen eingeteilt. Abgesehen von den Veränderungen, die unter dem Optimierungsaspekt erfolgten (Kapitel 2.2.4.4), wurden die Primerkonzentrationen unabhängig der Multiplexgröße beibehalten. Alle Amplifikationen fanden in einem Thermocycler (GeneAmp PCR System 9600, Perkin Elmer oder Thermal Cycler 2720, Applied Biosystems) statt. Der Aktivierung der hitzeinduzierten Polymerase unter den Bedingungen 10 Minuten bei 95°C folgten 35 Zyklen aus je drei Schritten: 30 Sekunden Denaturierung bei 95°C, 30 Sekunden Annealing bei 60°C und 30 Sekunden Elongation bei 65°C. Nach einer letzten Extension von 6 Minuten bei 65°C wurde die PCR gestoppt und auf 6°C abgekühlt.

Der Erfolg der PCR konnte stichprobenartig im Vergleich zur Negativkontrolle mittels einer Agarosegelelektrophorese 2% (1x TBE Puffer mit Ethidiumbromid) überprüft werden.

Anschließend wurden 15µl von den PCR-Produkten für die Minisequenzierungsreaktion enzymatisch aufgereinigt (Kapitel 2.2.4.5.1).

### 2.2.4.3 PCR-Produktkontrolle

Die Effizienz der PCR-Amplifikation von 52 autosomalen SNPs für 48 Loci wurde mittels Agilent 2100 Bionalyzer Chiptechnologie bestätigt (Abbildung 6). Dafür wurde ein DNA LabChip® Kit verwendet, welches einen Chip und Reagenzien beinhaltet, um eine Längenbestimmung und Quantifizierung der DNA-Fragmente durchzuführen. Der DNA LabChip® besteht aus 16 Mikrobehältern, davon werden drei für einen Gel/Farbstoff-Mix und ein weiteres für die DNA-Leiter verwendet, die verbliebenen zwölf Behälter werden mit unterschiedlichen PCR-Produkten befüllt (Abbildung 6). Weiterhin besteht der Chip aus einem Netz von Mikroleitungen, das die elektrophoretische Auftrennung der DNA-Fragmente nach ihrer Fragmentlänge ermöglicht. Die Befüllung des Chips mit Gel-Dye Mix wird durch eine spritzenbasierte Chip Priming Station erleichtert. Die Agilent-Plattform ist eine vollautomatische Elektrophorese-Technologie, die in 30 Minuten genaue Angaben über Fragmentlängen und Konzentrationen von zwölf verschiedenen PCR-Produkten liefert.

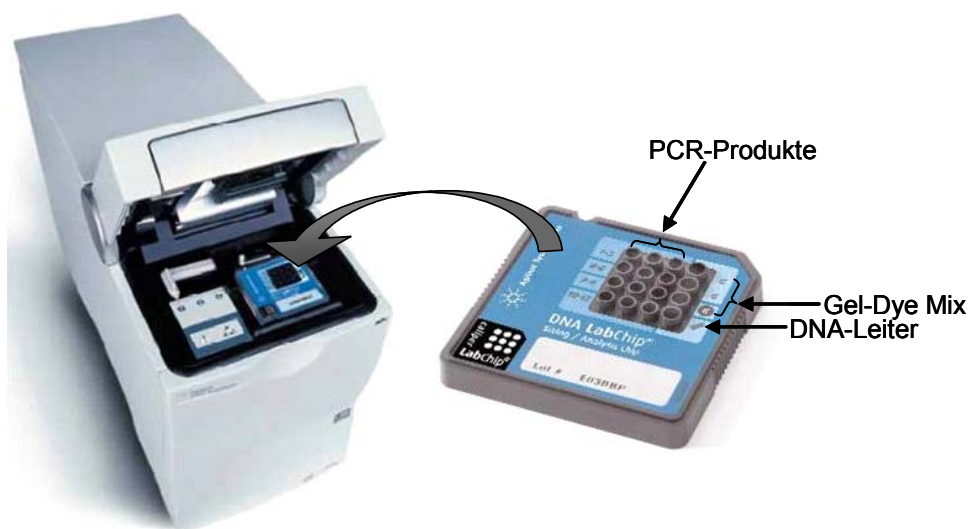


Abbildung 6: Agilent 2100 Bionalyzer Chiptechnologie zur Überprüfung der PCR-Produkte.

Da sich das Multiplexdesign von 52 SNPs im Bereich ähnlicher Fragmentlängen bewegte, mussten für die Überprüfung kleinere Multiplexe erstellt werden, denn die Technologie fordert einen Längenunterschied von mindestens 5bp zwischen den einzelnen Loci. Bestimmte Multiplexe von drei bzw. vier SNPs wurden für die Agilent-Analyse erstellt (Tabelle 8), die auch mittels Nanogen typisiert wurden, um die Frage nach eventuellen Assoziationen zwischen der PCR-Konzentration in Agilent und der Signalstärke in Nanogen zu beantworten.

Tabelle 8: 15 Multiplexe aus drei oder vier SNPs zur Überprüfung der PCR-Produkte. Durch Fettschrift sind die Namen der SNPs hervorgehoben, daneben stehen die entsprechenden Fragmentlängen der PCR-Produkte.

	<b>1 SNP</b>		<b>2 SNP</b>		<b>3 SNP</b>		<b>4 SNP</b>	
<b>PCR 1</b>	<b>18</b>	59bp	<b>02</b>	83bp	<b>12</b>	93bp	<b>52</b>	110bp
<b>PCR 2</b>	<b>16</b>	60bp	<b>32</b>	84bp	<b>40</b>	94bp	<b>53</b>	113bp
<b>PCR 3</b>	<b>24</b>	62bp	<b>23</b>	85bp	<b>48</b>	96bp	<b>54</b>	115bp
<b>PCR 4</b>	<b>25</b>	63bp	<b>33</b>	85bp	<b>42</b>	97bp		
<b>PCR 5</b>	<b>22</b>	68bp	<b>05</b>	86bp	<b>20</b>	98bp		
<b>PCR 6</b>	<b>01</b>	68bp	<b>13</b>	86bp	<b>43</b>	98bp		
<b>PCR 7</b>	<b>26</b>	69bp	<b>34</b>	86bp	<b>08</b>	100bp		
<b>PCR 8</b>	<b>11</b>	70bp	<b>07</b>	87bp	<b>10</b>	100bp		
<b>PCR 9</b>	<b>27</b>	72bp	<b>17</b>	87bp	<b>44</b>	100bp		
<b>PCR 10</b>	<b>14</b>	73bp	<b>35</b>	87bp	<b>45</b>	102bp		
<b>PCR 11</b>	<b>28</b>	73bp	<b>03</b>	90bp	<b>46</b>	103bp		
<b>PCR 12</b>	<b>29</b>	76bp	<b>15</b>	90bp	<b>19</b>	105bp		
<b>PCR 13</b>	<b>30</b>	78bp	<b>36</b>	90bp	<b>41</b>	106bp		
<b>PCR 14</b>	<b>04</b>	79bp	<b>37</b>	90bp	<b>49</b>	107bp		
<b>PCR 15</b>	<b>21</b>	81bp	<b>38</b>	91bp	<b>50</b>	108bp		

Die Agilent-Analyse erfolgte laut Herstellerprotokoll. Alle Reagenzien wurden 30 Minuten vor dem Benutzen bei Raumtemperatur äquilibriert. Als erstes wurde ein Gel-Dye Mix hergestellt, indem 25µl Dye zur Gel Matrix gegeben, gemischt und 15 Minuten zentrifugiert wurden. Es wurden 9µl Gel-Dye Mix in den vorgesehenen Behälter (**C** und **G** markiert) auf den Chip gegeben und mittels Chip Priming Station in die Mikroleitungen gefüllt (Abbildung 6). Weiterhin wurden 5µl Gel-Dye Mix in den Behälter für die DNA-Leiter gegeben. Jeweils 5µl des DNA-Markers wurden in die zwölf Behälter für die PCR-Produkte gefüllt. Anschließend wurde 1µl DNA-Leiter und je 1µl PCR-Produkt in den dafür vorgesehenen Behälter pipettiert. Der präparierte Chip wurde im Agilent 2100 Bionalyzer so platziert, dass die Elektroden direkt in die Mikrobehälter auf dem Chip passten. 30 Minuten nach dem Start konnten die Ergebnisse digital in vorprogrammierter Reihenfolge als Elektropherogramm, Gelbilder und Datentabelle mittels der 2100 Expert Software angezeigt werden.

#### 2.2.4.4 PCR-Optimierung

Es wurden keine PCR-Optimierungen vorgenommen, weder für die Amplifikation der 29 Y-chromosomalen SNPs Validierungsstudie noch für die 52 autosomalen SNPs der forensischen Validierungsstudie. Es erfolgten aber mehrere Maßnahmen, um die Effizienz der PCR-Amplifikation speziell für die Microarray-Technologie zu steigern.

Als erstes wurde der Versuch gestartet, den für das Microarray benötigten DNA-Strang gegenüber dem Rest im Überschuss zu amplifizieren, indem eine asymmetrische Amplifikation, oder asymmetrische Reamplifikation erfolgte, bis schließlich eine physikalische Strangtrennung unternommen wurde. Die asymmetrische Amplifikation lief unter Verwendung bestimmter Primermix-Verhältnisse (5:1 oder 10:1) zwischen den Primern, die den Targetstrang und den unbenötigten DNA-Strang amplifizieren. Die asymmetrische Reamplifikation wurde nach einer Standard-PCR durchgeführt. Für die Reamplifikation wurde ein Primermix aus Primeroligos zur Synthetisierung der Targetstränge verwendet. Zu

33,5µl des PCR-Produktes kamen in der Endkonzentration 0,6x Reaction Buffer, 3,25mM MgCl<sub>2</sub>, 0,5mM dNTPs, 8µM je Primer und 0,05U AmpliTaq™ Gold hinzu. Die Reamplifikationen sind unter gleichen Konditionen wie die Standardprotokolle über 15 Zyklen gelaufen.

Um den benötigten DNA-Strang abzutrennen, wurde der Primer für den unbenötigten Strang mit Biotin versehen und durch Streptavidin-Beads nach der Amplifikation aufgefangen und dadurch die Targetstränge separiert.

Außerdem wurde die Wirkung des PCR-Additivs BSA (Bovines Serumalbumin) und Glycerin getestet, indem 100ng BSA (Roche Diagnostics GmbH) und 0,58µl Glycerin 86% (Carl Roth) das 50µl Standardprotokoll ergänzten.

Abschließend wurden die Konzentrationen der Primer für die einzelnen SNPs variiert und je nach Quantifizierung mittels Agilent (Kapitel 2.2.4.3) und nach Signalintensität bei der Genotypisierung erhöht oder reduziert (Tabelle 9).

Tabelle 9: Primerkonzentrationen der Standard-Multiplex-PCR von 52 SNPs und der optimierten PCR von 48 SNPs. N bedeutet, dass diese SNPs nicht die Designkriterien des elektronischen Microarrays erfüllten (Kapitel 2.2.6.4) und aus dem Multiplex entfernt wurden.

SNPs	Endkonzentrationen [ng/µl]		SNPs	Endkonzentrationen [ng/µl]	
	Standard 52plex	Optimiert 48plex		Standard 52plex	Optimiert 48plex
SNP01	0,06	0,11	SNP27	0,01	0,03
SNP02	0,02	0,04	SNP28	0,03	0,06
SNP03 <sup>1)</sup>	0,02	0,10	SNP29	0,03	0,05
SNP04	0,03	0,05	SNP30	0,03	0,05
SNP05	0,04	0,07	SNP32	0,05	0,09
SNP06	0,06	N	SNP33	0,03	0,05
SNP07 <sup>1)</sup>	0,03	0,06	SNP34	0,03	0,05
SNP08	0,09	0,17	SNP35	0,04	0,08
SNP09	0,03	N	SNP36	0,04	0,08
SNP10 <sup>1)</sup>	0,05	0,24	SNP37 <sup>1)</sup>	0,05	0,10
SNP11	0,02	0,04	SNP38 <sup>1)</sup>	0,04	0,24
SNP12	0,05	0,09	SNP39	0,05	N
SNP13 <sup>1)</sup>	0,01	0,04	SNP40	0,03	0,10
SNP14	0,02	0,03	SNP41 <sup>1)</sup>	0,03	0,12
SNP15	0,17	0,31	SNP42	0,02	0,04
SNP16	0,02	0,04	SNP43	0,01	0,03
SNP17	0,08	0,14	SNP44 <sup>1)</sup>	0,07	0,08
SNP18	0,07	0,12	SNP45	0,05	0,09
SNP19	0,06	0,10	SNP46 <sup>1)</sup>	0,04	0,12
SNP20	0,03	0,06	SNP48	0,03	0,05
SNP21	0,12	0,23	SNP49	0,03	0,06
SNP22 <sup>1)</sup>	0,02	0,06	SNP50 <sup>1)</sup>	0,03	0,04
SNP23	0,04	0,08	SNP51	0,04	N
SNP24	0,04	0,07	SNP52 <sup>1)</sup>	0,03	0,02
SNP25	0,03	0,05	SNP53	0,02	0,04
SNP26 <sup>1)</sup>	0,02	0,02	SNP54	0,06	0,10

1) Aus der 35plex-Amplifikation entfernt.

## 2.2.4.5 PCR-Produktaufreinigung

### 2.2.4.5.1 Enzymatische Aufreinigung

Jeder Minisequenzierung der Y-chromosomalen und autosomalen Multiplexe vorangeschaltet war eine enzymatische Aufreinigung mittels SAP (Shrimp Alkaline Phosphatase) und Exo I (E. coli Exonuclease I). Hierbei wurden in einer Reaktion die PCR-Produkte mit den Enzymen Exo I und SAP versetzt, um überschüssige Primer und dNTPs, sowie einzelsträngige DNA aus der PCR-Reaktion zu entfernen (Werle et al. 1994) (Tabelle 10). Die Exo I degradiert einzelsträngige DNA und Oligonukleotide in 3'>5' Richtung und SAP löst die Phosphatgruppen der freien Nukleotide auf, sodass diese in der Nachfolgereaktion zum Beispiel das Verhältnis von dNTPs und markierten ddNTPs nicht stören. Der Reaktionsansatz wurde im Thermocycler für 60 Minuten bei 37°C inkubiert und anschließend wurden die Enzyme 15 Minuten bei 75°C inaktiviert.

Tabelle 10: Standardprotokoll der enzymatischen Aufreinigung.

Komponenten (Firma)	Lieferkonzentration	Endkonzentration	Volumen [µl]
Reaction Buffer (USB)	10x	0,17x	0,36
SAP (USB)	1U	2U/21µl	2,00
Exo I (USB)	10U	4U/21µl	0,40
PCR Produkt			15,00
H <sub>2</sub> O			3,24
Total			21

### 2.2.4.5.2 Vergleich verschiedener Aufreinigungsmethoden

Für die elektronische Microarray-Genotypisierung ist die Entsalzung der PCR-Produkte eine wichtige Voraussetzung. Die Schwierigkeit bestand darin, eine schonende Aufreinigungsmethode zu finden, welche gründlich genug ist für die Microarray-Technologie, gleichzeitig aber keinen Produktverlust, gerade von kurzen PCR-Fragmenten (59-115bp) einer Multiplex, verursacht. Dafür wurden folgende Aufreinigungs-Kits getestet und ihre Eignung für die Genotypisierung verglichen: MultiScreen® Separations System (Millipore Corporation), MSB® Spin PCRapace (Invitex), MinElute PCR Purification Kit (Qiagen) und ChargeSwitch® Forensic DNA Purification Kit (Invitrogen life technologies). Alle Aufreinigungsschritte richteten sich nach den Protokollen der Hersteller.

Während die ChargeSwitch®-Aufreinigung mittels magnetischer Beads erfolgt, beruhen die anderen drei Methoden auf einem Filtermembran-Trennprinzip durch Säulenzentrifugation (Invitex und Qiagen) oder Vakuum (Millipore).

Das Millipore MultiScreen® Separations System besteht aus einer Manifoldkammer als Verteiler, mit einer Halterung für eine 96-Mikrotiterplatte und den nötigen Schläuchen mit einem Druckventil für die Vakuumregulation. Für die Aufreinigung mittels Millipore

Separations System wurden die PCR-Produkte 1:1 mit HPLC-Wasser verdünnt und in die Behälter einer 96-er Mikroplatte mit Filtermembranboden (MultiScreen<sup>®</sup> PCR<sub>96</sub> Filter Plate, Millipore) von 0,2µm Mikroporen überführt. Das Manifold wurde an eine Vakuumpumpe angeschlossen und für 10-50 Sekunden Vakuum (500mbar) angelegt, bis die ganze Flüssigkeit durch das Filter gezogen wurde. Nach einem Waschschrift mit 100µl HPLC-Wasser wurden 45µl 100mM Histidin (His) in die Behälter pipettiert und 15 Minuten auf einem Schüttler eluiert.

Bei den zwei Zentrifugationssystemen wurden die PCR-Produkte mit einem speziellen Puffer (Binding Buffer) 1:4 gemischt, auf Filtersäulen geladen und zentrifugiert. Der Puffer garantiert die optimalen pH-Bedingungen, damit die doppelsträngige DNA an der Filtermembran der Säule haftet. Der Unterschied zwischen den zwei Aufreinigungsprozessen liegt in dem entfallenen Waschschrift bei dem Invitek-Kit. Das Qiagen-Kit entfernt mittels eines Waschpuffers auf Ethanolbasis in einem zusätzlichen Zentrifugationsschritt überschüssige Reagenzien wie z.B. Puffer, Salze, ungebundene Primer und dNTPs. Im letzten Elutionsschritt wurden die gereinigten Produkte mit HPLC-Wasser ca. fünf Minuten bei Raumtemperatur inkubiert und durch Zentrifugation von der Filtermembran gelöst.

Für die ChargeSwitch-Aufreinigungsmethode wurden die PCR-Produkte zu einem Gemisch von magnetischen Beads (ChargeSwitch<sup>®</sup> Magnetic Beads, Invitrogen life technologies) und Clean-up-Buffer gegeben. Der pH-Wert wird auf <6,5 abgesenkt und dadurch die Oberfläche der Beads positiv beladen. Während einer kurzen Inkubationszeit wurden die negativ geladenen PCR-Produkte an der Oberfläche der Beads gebunden. Anschließend wurden die Gefäße mit dem PCR/Beads-Gemisch in einen magnetischen Ständer (MagnaRack<sup>™</sup>, Invitrogen life technologies) eingebracht. Mit Hilfe der magnetischen Anziehung konnten die Beads sehr effizient von der Flüssigphase separiert werden. Im anschließenden Waschschrift wurde der Vorgang wiederholt, dafür wurden die Gefäße vom Ständer genommen, die Beads mittels eines Puffers gewaschen, dabei überschüssige Primer und Salze gelöst und die Beads mittels Ständer wieder separiert. Ein darauf folgender Elutionsschritt neutralisiert mittels eines salzarmen Puffers mit einem höheren pH-Wert (>8,5-9,0) die Oberflächenladung. Die ursprünglich an die positiv geladene Oberfläche gebundenen PCR-Produkte verlieren ihre Affinität zur zunehmend neutralen Oberfläche und befinden sich bei der letzten magnetischen Separierung im Elutionspuffer.

## 2.2.5 Minisequenzierung

Für die Minisequenzierung wird ein Verfahren benötigt, das zum einen einzelsträngige DNA-Fragmente unterschiedlicher Länge erzeugt, zum anderen eine Markierung entsprechend der zu analysierenden Basen Adenin (A), Cytosin (C), Guanin (G) oder Thymin (T) liefert, damit sie später detektiert und voneinander unterschieden werden können. Die Minisequenzierung löst diese Aufgaben durch enzymatische Reaktionen, die sich thermozyklisch wiederholen.

Durch Anwendung der Minisequenzierung wird, ausgehend von dem gereinigten, amplifizierten Target, eine sogenannte Single Base Extension SBE-Reaktion durchgeführt. Hierbei bindet der Analyseprimer exakt eine Base vor dem polymorphen SNP an die einzelsträngige



DNA. In der Reaktion werden anschließend fluoreszenzmarkierte ddNTPs mittels einer DNA-Polymerase je nach Sequenz komplementär am 3' Ende eingebaut (Abbildung 7). Nach einem basenspezifischen Einbau der ddNTPs kommt es zum Kettenabbruch, da die fehlende 3' OH-Gruppe eine weitere Strangsynthese unmöglich macht.

Die Analyse sowie Visualisierung der Reaktion kann anschließend mittels Kapillarelektrophorese nach laserinduzierter Fluoreszenzdetektion erfolgen (Kapitel 2.2.5.3). Die Minisequenzierung ist eine Multiplexreaktion, welche sich für die parallele Analyse von mehreren SNPs eignet, indem Analyseprimer unterschiedlicher Längen gewählt werden (Abbildung 7).

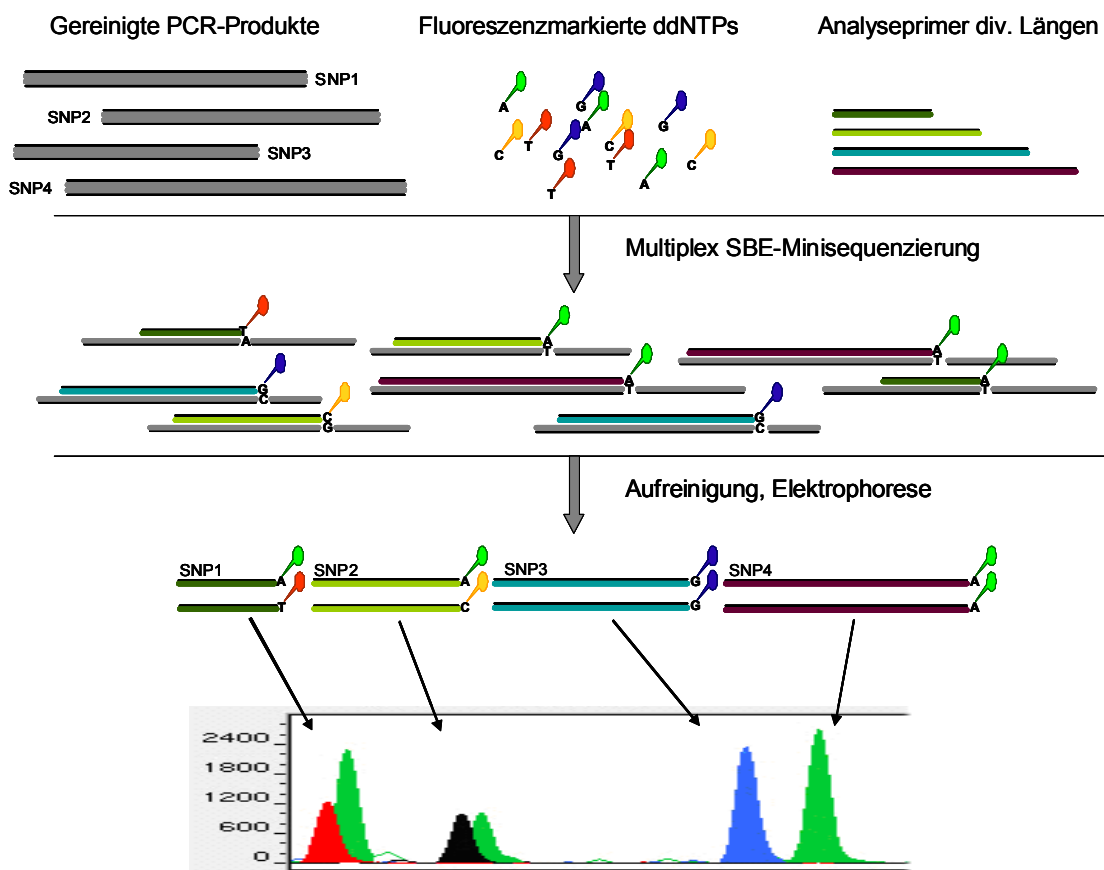


Abbildung 7: Das Prinzip der SBE-Minisequenzierung. Mittels Analyseprimer von unterschiedlichen Längen können die Produkte einer Multiplex elektrophoretisch aufgetrennt werden.

### 2.2.5.1 Multiplex-SBE-Primerdesign

Die Auswahl der Analyseprimer für die SBE-Minisequenzierung erfolgte analog des vorhergehenden PCR-Primerdesigns (Kapitel 2.2.4.1) mit Hilfe der frei zugänglichen Primer 3.0 v. 0.2 Software. Die Entscheidung über die Längen der Analyseprimer wurde danach getroffen, dass Fragmente mit gleichen Substitutionsmöglichkeiten vier Basen auseinanderliegen. Gleichzeitig bedeutet das, dass zwei Analyseprimer mit unterschiedlichen Substitutionsmöglichkeiten die gleiche Länge besitzen können. Zum Beispiel werden

SNP M2 mit A/G und SNP Tat mit T/C Substitutionen von zwei Analyseprimern von 26nt Länge detektiert (Tabelle 11).

Die Längen der Primer für die 29 Y-chromosomalen SNPs liegen zwischen 17 und 94nt. Davon sind die targetspezifischen Sequenzen zwischen 17 und 28nt. Um die gewünschte Länge zu erreichen, wurde am 5' Ende ein neutraler Schwanz angefügt. Dafür wurde eine Sequenz gewählt, welche in der NCBI-Datenbank keine Übereinstimmung mit humaner DNA zeigt (5'-AACTGACTAACTAGGTGCCACGTCGTGAAAGTCTGACAA-3', Lindblad-Toh et al. 2000). Wenn diese Randomsequenz aus 40 Nukleotiden nicht ausreichend war, wurde der Analyseprimer zusätzlich mit Poly-T bzw. Poly-C verlängert (Tabelle 11).

Tabelle 11: Analyseprimer der Minisequenzierung zur Detektion von 29 Y-chromosomalen SNPs in einer einzelnen Reaktion. Die Target-Sequenz ist fragmentspezifisch und endet unmittelbar vor dem SNP. Die Target-Sequenz, 5' Anhang und Poly-dNTP zusammen ergeben die Länge der Analyseprimer (Brion et al. 2005, modifiziert).

SNP	Poly-dNTP <sup>1)</sup>	5' Anhang <sup>1)</sup>	Target-Sequenz <sup>1)</sup>	$\mu\text{M}$ <sup>2)</sup>	nt <sup>3)</sup>
M174	/	/	tacctctggagtgccc	0,06	17
M45	/	caa	ctcagaaggagcttttgc	0,14	22
M2	/	gacaa	ctttatcctccacagatctca	0,2	26
Tat	/	/	gctctgaaatattaaatcaaac	0,08	26
M170	/	t	ctattttacttaaaatcattgttc	0,05	30
M217	/	acaa	ttatgtattttcctctgaagagtt	0,07	30
P25	/	tcgtgaaagtctgacaa	tgccgaaacctgcctg	0,41	34
M201	/	aagtctgacaa	taataatccagatcaactgagg	0,02	34
M38	/	gtgaaagtctgac	cagtactattatggaaaaccaact	0,1	38
M304	/	gaaagtctgacaa	tgtcaattgaaagtaactgtga	0,14	38
M123	/	acgtgtgaaagtctgacaa	atttctaggtattcaggcgatg	0,68	42
M207	/	cgtgaaagtctgacaa	caaatgtaagcaagcaagaattta	0,24	42
M35	/	gccacgtcgtgaaagtctgacaa	cccaatttctttgggacactg	0,54	46
M128	/	cgtgaaagtctgacaa	cctcttcaactgtaaatgaaataactg	0,11	46
P31	/	gccacgtcgtgaaagtctgacaa	tacataaataaggttttttgggtg	0,08	50
M216	/	ggtgccacgtcgtgaaagtctgacaa	gctagtattgtatactgttgaat	0,08	50
M17	/	aactaggtgccacgtcgtgaaagtctgatct	ccaaaattcacttaaaaaacc	0,07	54
M175	/	ctaaactaggtgccacgtcgtgaaagtctgacaa	cacatgccttctcacttctc	0,24	54
M173	/	actaaactaggtgccacgtcgtgaaagtctgacaa	tacaattcaaggcatttagaac	0,07	58
M119	/	gactaaactaggtgccacgtcgtgaaagtctgacaa	tattccaattcagcatacaggc	0,2	58
M96	/	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	ggaaaacaggctctcataata	0,36	62
M122	dC+5(dTdC)+dT	ctaaactaggtgccacgtcgtgaaagtctgacaa	tcagattttcccctgagagc	0,1	66
M75	4(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	aaagacaattatcaaacacatcc	0,02	70
SRY10831	5(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	ttgtatctgacttttccacagct	0,06	74
M168	dC+6(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	ctattgttttaattctcagctagc	0,03	78
M8	7(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	cttggtttgtgtgagtatactctatgac	0,08	82
M9	9(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	catgtctaaattaaagaaaaataagag	0,88	86
P2	16(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	gtgccctaggaggagaa	0,1	90
M33	dC+12(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	cagttacaaaagtataatgtctgagat	0,27	94

1) Sequenzrichtung 5' Ende zu 3' Ende.

2) Endkonzentration der Analyseprimers in 29 Multiplex-Minisequenzierung.

3) Länge der Analyseprimers in Nukleotiden.

Tabelle 12: Analyseprimer zur Detektion von 52 autosomalen SNPs in zwei Reaktionen. Die Längen der Analyseprimer des Auto1-Kits ergeben die Target-Sequenz, 5' Anhang und Poly-dNTPs. Die Längen der Analyseprimer des Auto2-Kits ergeben sich aus den Target-Sequenzen und Poly-dNTPs (Sanchez et al. 2006b, modifiziert).

SNP	Poly-dNTP <sup>1)</sup>	5' Anhang <sup>1)</sup>	Target-Sequenz <sup>1)</sup>	µM <sup>2)</sup>	nt <sup>3)</sup>
<b>AUTO1</b>					
SNP1	/	a	TCTGAGGCCAGCCAGTT	0,01	18
SNP2	/	caa	CACTGCACTGAAGTATAAGTA	0,04	24
SNP3	/	tctgacaa	TGATAAGAAACATGACCAAGC	0,02	29
SNP4	17(dC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	TTGTTGACACTTCACCTTCTA	0,27	78
SNP5	13(dC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	GGCATATCGTATTAACGTGT	0,02	74
SNP6	/	aaactaggtgccacgtcgtgaaagtctgacaa	GTAAGAATTCAAGATGGTATTT	0,04	54
SNP7	/	/	TGACTGAGGTCAACGAGC	0,02	18
SNP8	/	cgtgaaagtctgacaa	GTTTGTATATTATTTCTAACTCAA	0,03	42
SNP9	/	gtctgacaa	ATGAAAAATCACTGGGACAG	0,02	29
SNP10	/	gtgaaagtctgacaa	CACCGAATTCAACGGGAAG	0,02	34
SNP11	/	aggtgccacgtcgtgaaagtctgacaa	ACTAGCTGAATATCAGCCC	0,04	46
SNP12	/	taaactaggtgccacgtcgtgaaagtctgacaa	GTGTTTTTCTAAATCATATTGTCT	0,02	58
SNP13	/	acaa	GATTTTCACAACAACACTTGC	0,05	25
SNP14	4(dC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	AAATACACCCTGAGCTGC	0,02	62
SNP15	14(dC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	TGAGAGAGAAAAATGTGC	0,02	74
SNP16	10(dC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	TCTTGC GGTTACATACATTC	0,02	70
SNP17	/	cgtcgtgaaagtctgacaa	TAAGGTGAGTGGTATAATCATAT	0,01	42
SNP18	5(dC)	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	TTGTGAGGCCTGTTTATTTTG	0,01	66
SNP19	/	actgactaaactaggtgccacgtcgtgaaagtctgacaa	TTTCCTCTCCCATTCAG	0,04	58
SNP20	/	aaactaggtgccacgtcgtgaaagtctgacaa	TTTCAAGTATGCTCGGGG	0,02	50
SNP21	/	cgtgaaagtctgacaa	TGACAGTAATGAAATATCCTTG	0,01	38
SNP22	/	cgtgaaagtctgacaa	GGCCATCCACTTGAAA	0,01	34
SNP23	/	tgccacgtcgtgaaagtctgacaa	GACTAATTAATCTGCTATCTG	0,01	46
<b>AUTO2</b>					
SNP24	dT + 16(dCdT)	/	TATCTGTCTTTGGCCTTCTAACT	0,03	56
SNP25	dT + 14(dCdT)	/	AATCAGGCATGTTACATTACAA	0,03	52
SNP26	dT + 21(dCdT)	/	GAGAAACACCTGAACCTTCAA	0,03	64
SNP27	2(dCdT)	/	CATCTTGGCAGCATCCTTC	0,03	23
SNP28	dT + 37(dCdT)	/	CCCTGTCTTGAACCTCC	0,02	92
SNP29	dT + 15(dCdT)	/	TGTTCTAATAAAAAGGATTGCTCAT	0,03	56
SNP30	dT + 26(dCdT)	/	CCGGAACCTCAACGACTTA	0,04	72
SNP32	9(dCdT)	/	AGACATTTGTTTCATATAAGTGAAATG	0,05	44
SNP33	dT	/	CTAGTACGTTAGATGTGACCG	0,05	22
SNP34	dT + 10(dCdT)	/	GAACGTTGGAACCTTTGCT	0,05	40
SNP35	32(dCdT)	/	CAGCATACACTCATAGCCAC	0,05	84
SNP36	5(dCdT)	/	CCACCAGAAATCAGGGC	0,05	27
SNP37	dT + 15(dCdT)	/	CTGGCTTTGGCAAGTCC	0,03	48
SNP38	dT + 4(dCdT)	/	CCAGTTGGAGCCTTCTTT	0,03	27
SNP39	dT + 28(dCdT) + dC	/	CTCTCTCATGTATCACATTCT	0,04	80
SNP40	21(dCdT) + dC	/	CTCTGTATTTTCTTACTCTAAGTGC	0,05	68
SNP41	/	/	TGAGGCCATCTCCACC	0,03	16
SNP42	dT + 9(dCdT)	/	GGACACACCATTTTATTGTCTAAAG	0,03	44
SNP43	dT + 33(dCdT)	/	TAGTAGATATCTGGCTGTCCC	0,03	88
SNP44	dT + 7(dCdT)	/	GAGCTGCCTGAGGGTGT	0,05	32
SNP45	22(dCdT)	/	GAAAAACAACAGAATGGAATAGGA	0,04	68
SNP46	/	/	AGGGGATGCAGCCAGT	0,03	16
SNP48	dT + 9(dCdT)	/	TGTGAGGAACGTCGACA	0,03	36
SNP49	dT + 9(dCdT)	/	GCAAGAGCCGTGGAATT	0,03	36
SNP50	dT + 26(dCdT)	/	TACAGAAATAACTCTTGCGCTAC	0,03	76
SNP51	dT + 3(dCdT) + dC	/	TCCATTTCTTTTTTTTGAAGCCT	0,02	32
SNP52	dT + 7(dCdT)	/	AGGTACCTAGCTATGTACTCAGTAT	0,04	40
SNP53	dT + 13(dCdT)	/	GAAAGGTCCTTACTCGACATC	0,03	48
SNP54	dT + 17(dCdT)	/	AAAACCTTTTACTGTAACATTTTCA	0,03	60

1) Sequenzrichtung 5' Ende zu 3' Ende.  
 2) Endkonzentration der Analyseprimer in der 23plex- bzw. 29plex-Minisequenzierungen.  
 3) Länge der Analyseprimer in Nukleotiden.

Die Analyseprimer für die Typisierung der 52 autosomalen SNPs wurden in zwei Multiplexe geteilt, in Auto1 für 23 SNPs und in Auto2 für die restlichen 29 SNPs (Tabelle 12). Demnach wurden die PCR-Produkte der 52plex mittels zweier SBE-Reaktionen mit unterschiedlichen Analyseprimer-Kits typisiert. Die Längen der Analyseprimer für den Auto1-Kit liegen zwischen 17 und 78 Nukleotiden. Die passenden Primerlängen wurden wie bei den Y-chromosomalen SNPs mit Hilfe der Randomsequenz und Poly-T bzw. Poly-C erreicht (Tabelle 12). Die Längen der Analyseprimer für den Auto2-Kit liegen zwischen 16 und 92 Nukleotiden und zur Verlängerung wurde Poly-T bzw. Poly-C herangezogen (Tabelle 12). Alle Analyseprimer besitzen eine theoretische Annealingtemperatur um 55°C. Zur Erzielung möglichst gleicher Produktausbeute für alle 52 bzw. 29 SNPs bei gleichzeitig geringem Aufkommen von Primer-Dimeren oder unspezifischen Nebenprodukten wurden die optimalen Primerkonzentrationen für jeden Marker empirisch ermittelt (Tabelle 11 und Tabelle 12). Wie bereits die PCR-Primer, wurden auch die SBE-Analyseprimer zentral von der Firma DNA Technology A/S geliefert und an die teilnehmenden Arbeitsgruppen verteilt. Die Oligos waren HPLC-gereinigt und ihre Reinheit mit MALDI-ToF MS überprüft. Zur Lagerung wurden sie in Tris/EDTA Puffer auf 50µM verdünnt (10mM Tris, 100mM EDTA, pH 7,5, Sigma-Aldrich).

### 2.2.5.2 SNaPshot™-Minisequenzierung

In der vorliegenden Arbeit wurde das kommerzielle ABI Prism® SNaPshot™ Multiplex-Kit (Applied Biosystems), das auf dem Prinzip der Minisequenzierung beruht, für die Y-chromosomale und die autosomale forensische Validierungsstudien sowie auch als Referenzmethode für die Microarray-Validierungsstudie eingesetzt.

Der Ready Reaction Mix beinhaltet alle Reagenzien, die für die Minisequenzierung nötig sind. Die Komponenten Reaktionspuffer, fluoreszenzmarkierte ddNTPs und AmpliTaq® DNA-Polymerasen sind in unbekanntem Mengenverhältnissen gemischt. Die Fluoreszenzmarkierungen der ddNTPs sind in Tabelle 13 aufgeführt.

Tabelle 13: Fluoreszenzmarkierung der einzelnen ddNTPs des ABI Prism® SNaPshot™ Multiplex-Kit und ihre optischen Eigenschaften.

ddNTP	Fluoreszenz	Emmission	Absorption	Farbe
<b>A</b>	dR6G	545nm	528nm	Grün
<b>C</b>	dTAMRA	580nm	546nm	Gelb <sup>1)</sup>
<b>G</b>	dR110	535nm	485nm	Blau
<b>T</b>	dROX	605nm	574nm	Rot

1) Im Elektropherogramm schwarz dargestellt.

Der SNaPshot-Ansatz zur SNP-Typisierung erfolgte stets in einem Volumen von 8µl (Tabelle 14). Die Komponenten jeder SNaPshot-Reaktion waren die Exo/SAP aufgereinigten PCR-Produkte, die sequenzspezifischen Analyseprimer und der SNaPshot Ready Mix. Die Minisequenzierung wurde in einem Thermocycler (GeneAmp PCR System 9600, Perkin

Elmer oder Thermal Cycler 2720, Applied Biosystems) in 25 Zyklen in drei Schritten durchgeführt (10 Sekunden 96°C, 5 Sekunden 55°C und 30 Sekunden 60°C). Anschließend wurden durch Zugabe von 1U SAP und einer Stunde Inkubation die Phosphatgruppen der freien ddNTPs aufgelöst. Nachdem das Enzym 15 Minuten bei 75°C inaktiviert wurde, waren die Minisequenzierungsprodukte fertig für die Analyse in einer Kapillarelektrophorese.

Tabelle 14: Standardprotokoll der SNaPshot™-Minisequenzierung.

Komponenten	Lieferkonzentration	Endkonzentration	Volumen [µl]
<b>SNaPshot Ready Mix</b>	unbekannt		4,0
<b>Primermix<sup>1)</sup></b>	jeder 100µM	0,01-0,88µM	1,0
<b>Aufgereinigtes Produkt</b>			1,0
<b>H<sub>2</sub>O</b>			2,0
<b>Total</b>			8,0

1) Der Primermix beinhaltet alle Analyseprimer einer Multiplexreaktion in unterschiedlicher Konzentration.

### 2.2.5.3 Kapillarelektrophorese

Zur Genotypbestimmung wurden die Minisequenzierungsprodukte ihrer Größe nach in einer Kapillarelektrophorese aufgetrennt, die eingebauten fluoreszenzmarkierten ddNTPs detektiert und anschließend den verschiedenen Polymorphismen zugeordnet.

Der Trennungsmechanismus beruht auf der unterschiedlichen elektrophoretischen Mobilität der Moleküle. Die Mobilität ist durch die Ladung und die Größe des Moleküls definiert. Die Fragmente wandern im elektrischen Feld, unterstützt von dem elektroosmotischen Fluss, der durch die Wandladung der Kapillare hervorgerufen wird und sie ihrer Größe nach von dem viskosen Polymer auftrennt. Die Matrix des Gels aus Linear-Polymerketten ist derart effizient, dass sie eine theoretische Trennleistung von nur einem Nukleotid besitzt. Erreicht ein markiertes DNA-Fragment das Detektionsfenster, so wird dieses von einem Argonlaserstrahl zur Fluoreszenz angeregt. Das emittierte Licht wird auf einem Hohlspiegel prismatisch aufgespalten. Die Signale werden von einer CCD-Kamera (Charged Coupled Device) aufgenommen und an eine Analysesoftware weitergeleitet. Die Ergebnisse lassen sich in Form eines Elektropherogramms als verschiedenfarbige Signale für die einzelnen ddNTPs darstellen oder die Fragmentlängen, Peakhöhen und -flächen und Farben tabellarisch angeben.

Zur Analyse wurden 4µl SNaPshot-Produkte mit 19µl deionisiertem Formamid Hi-Di (Applied Biosystems) und 0,5-1,5µl GeneScan™ 120 LIZ™ Size Standard (Applied Biosystems) gut gemischt und bei 95°C 5 Minuten in einem Thermocycler denaturiert.

Der interne Längenstandard besteht aus DNA-Fragmenten definierter Länge zwischen 15 und 120 Nukleotiden (15, 20, 25, 35, 50, 62, 80, 110 und 120nt). Die einzelsträngigen DNA-Fragmente sind mit 5<sup>th</sup> Dye LIZ® Fluorophore markiert. Zur Längenbestimmung wird die mathematische Global Southern Methode herangezogen, dabei wird die reziproke Verwandtschaft zwischen Fragmentlänge und Mobilität, abgewandelt nach E.M. Southern, berechnet (Southern 1979).

Für die Analyse der SNaPshot-Produkte in der Y-chromosomalen Validierungsstudie wurde die Kapillarelektrophorese ABI PRISM® 310 Genetic Analyzer (Applied Biosystems) benutzt, während für die autosomalen Multiplexe, sowohl für die forensische Validierungsstudie als auch für die Methodvalidierung, das vier Kapillarelektrophorese System ABI PRISM® 3100-*Avant* Genetic Analyzer (Applied Biosystems) eingesetzt wurde.

Die Lauf- und Analyseparameter unterscheiden sich je nach verwendetem System bzw. Polymer und sind in der Tabelle 15 zusammengefasst.

Tabelle 15: Lauf- und Analyseparameter der Kapillarelektrophorese-Systeme PRISM® 310 Genetic Analyzer und PRISM® 3100-*Avant* Genetic Analyzer.

Plattform	Polymer	Injektion		Lauf			Analyse Software
		Zeit	Volt	Volt	Temp.	Zeit	
AB 310	Pop-4™	22s	2kv	15kv	60°C	30min	GeneScan® v3.1.2
AB 3100	Pop-6™	25s	1kv	15kv	60°C	2x23min	GeneMapper® ID v3.1

#### 2.2.5.4 Analyse und Auswertung

Die Automatisierung in der genetischen Analyse und deren Anwendung in der Routine erzeugt eine immer größere Datenflut. Wichtig ist es, die Teilprozesse in der automatischen Analyse aufeinander abzustimmen und in das Gesamtverfahren zu integrieren. Dabei muss aber die Möglichkeit bestehen, auf die Teilbereiche im Rahmen der Datenverarbeitung zurückgreifen zu können, um deren Qualität und Eindeutigkeit zu überprüfen. Eine Arbeitserleichterung und Effizienzsteigerung bei der Datenanalyse konnte durch die Einrichtung verständlicher und flexibler Makros erreicht werden.

Die kapillarelektrophoretischen Plattformen, die die automatische Längenbestimmung und Quantifizierung von fluoreszenzmarkierten DNA-Fragmenten erlauben, sind an entsprechende Softwarepakete gekoppelt (z.B. Data Collection Software, GeneScan® und GeneMapper®). Die Ergebnisse werden als elektronische Files gespeichert, damit bleiben sie editierbar und können als Graphik oder Textdateien exportiert werden, um weiterführende statistische Auswertungen vorzunehmen.

Eine automatische Allelbestimmung der Y-chromosomalen SNPs erfolgte mit Hilfe der 310GeneScan® v3.1.2 Analyse Software, wobei die minimale Signalstärke für alle Farben auf 100RFUs (Relative Fluorescence Units) festgelegt wurde. In seltenen Fällen musste eine manuelle Korrektur vorgenommen werden. Für die in Mainz genotypisierten Proben wurden die Allelfrequenzen und ihre Zuordnung zu einer Haplogruppe vor Ort ermittelt. Weitere populationsbezogene Auswertungen wurden nach der Sammlung der Daten aus allen Laboratorien zentral in Santiago de Compostela mittels der Arlequin 2.0<sup>8</sup> Software vorgenommen. Um die Frage der Y-chromosomalen Validierungsstudie zu klären, inwieweit sich populationsgenetische Unterschiede in den Haplogruppen manifestieren, wurde die Methode Analysis of Molecular Variance (AMOVA, Excoffier et al. 1992) verwendet, die die Distanz zwischen zwei Haplogruppen ausdrückt. Der Differentiationsindex  $F_{ST}$  (Weir und Hill

<sup>8</sup> <http://anthro.unige.ch/arlequin>

2002) gibt das Verhältnis der Inter-Populations-Varianz  $\sigma_a^2$  zur gesamten Varianz  $\sigma^2$  an. Je höher der  $F_{ST}$ -Wert ist, umso stärker differieren die Haplogruppen. Mit Hilfe von SPSS Version 11.5 wurde das MDS (Multidimensional Scaling) der  $F_{ST}$ -Distanz erstellt.

Die Allelbestimmung der 52 autosomalen SNPs, welche bei der SNaPshot-Reaktion in zwei Multiplexe aufgeteilt wurde (Kapitel 2.2.5.1, Tabelle 12), erfolgte automatisch mit Hilfe der GeneMapper® ID v3.1 Software. Die minimalen Signalstärken wurden je nach Farbe unterschiedlich festgelegt, für Blau 100RFUs, Grün 60RFUs und Rot bzw. Gelb 30RFUs. Die Ergebnisse wurden, falls notwendig, nach vorheriger manueller Editierung in elektronische Textdateien umgewandelt. Anschließend wurden die Ergebnisse der zwei Multiplexanalysen mittels Access Software Makros zusammengeführt, sodass letztlich ein Genotyp aus 52 SNPs pro Person gegeben war.

## 2.2.6 Elektronische Microarray-Typisierung

Das Prinzip der Microarray-Genotypisierung beruht auf dem reversiblen Prozess der Hybridisierung. Als Hybridisierung wird die Reassoziaton von einzelsträngiger DNA zu einem Doppelstrang bezeichnet, welche von einem thermodynamischen Gleichgewicht abhängig ist. Bei der Microarray-Analyse kommt es zur selektiven Hybridisierung von komplementären Einzelsträngen, die anschließend durch Markierungen, z.B. mit Fluoreszenzfarbstoffen, detektiert werden. Die Eigenschaften der selektiven Hybridisierung und der multiplen Markierungen der Microarray-Genotypisierung ermöglichen die gleichzeitige Analyse einer großen Anzahl von DNA-Molekülen auf kleinstem Raum. Anwendungsgebiete der Microarrays sind vor allem Genexpressionsanalysen (Dhiman et al. 2002, Weidenhammer et al. 2002), aber auch die Identifizierung von Molekülen in Stoffwechsel- oder Signaltransduktionskaskaden (Hughes et al. 2000), die Tumorklassifizierung und die Targetsuche nach Zielgenen für Medikamente in der Pharmaindustrie oder auch die STR-Typisierung (Radtky et al. 2000) und SNP-Analyse zur Detektion von Mutationen (Sosnowski et al. 1997).

Als feste Phase für die Hybridisierung werden neben Kunststoffträgern, Membranen oder Glasobjektträgern, in jüngster Zeit auch elektronische Mikrochips verwendet. Die Firma Nanogen Inc. leistete 1997 Pionierarbeit mit ihrer fortschrittlichen Nanotechnologie. Sie verband zeitaufwendige und arbeitsintensive Prozeduren in einer Lab-On-A-Chip Technologie. Die elektronische Microarray-Technologie mittels Nanogen-Plattform erreicht eine hohe Effizienz, indem die Analyseschritte durch eine elektrische Feldaktivierung unterstützt werden. Somit werden nicht nur die zu immobilisierende ssDNA oder die Oligonukleotidsonden elektronisch adressiert, sondern es kann auch die anschließende Hybridisierung gezielt durch ein elektronisches Feld erfolgen. Zusätzlich ermöglicht die Nanogen-Technologie vollautomatische, vorprogrammierte, thermische Zyklen, die bei der Diskriminierung ausschlaggebend sind.

Für die vorliegende Arbeit standen zwei elektronische Microarray-Plattformen der Firma Nanogen zu Verfügung, die NanoChip<sup>®</sup> Molecular Biology Workstation (MBW) und die weiterentwickelte NanoChip<sup>®</sup> 400 (NC400). Die Assayetablierung, Optimierung und schließlich die Validierung der elektronischen Microarray-Technologie erfolgte an einer NanoChip<sup>®</sup> MBW-Plattform. Diese wurde der Mainzer Arbeitsgruppe von der Firma Nanogen zu Forschungszwecken zur Verfügung gestellt. Da die Firma Nanogen die Ablösung der NanoChip<sup>®</sup> MBW durch die weiterentwickelte Technologie der NanoChip<sup>®</sup> 400 anstrebt, wurde im Rahmen dieser Arbeit das etablierte Assay zusätzlich auch an diese Plattform adaptiert.

### 2.2.6.1 NanoChip<sup>®</sup> Molecular Biology Workstation

Die NanoChip<sup>®</sup> MBW integriert die Mikroelektronik in die molekulargenetische Analyse. Das System besteht aus zwei voneinander getrennten Instrumenten mit unterschiedlichen



Funktionen, dem sogenannten Loader und dem Reader (Abbildung 8). Die Handhabung und Kontrolle beider Instrumente erfolgt über einen integrierten Computer.

Das wichtigste Werkzeug der elektronischen MBW-Technologie ist das NanoChip™ 100 Cartridge, das einen elektronischen Silizium-Mikrochip von 2x2mm Größe umschließt. Die aktive Tragfläche des Mikrochips besteht aus 100 individuellen Platin-Mikroelektroden (Pads) von je 80µm Größe, die durch einzelne Platindrähte individuell und beliebig elektronisch aktiviert werden können (Abbildung 9). Die Oberfläche des Mikrochips kann, mittels eingebauten Mikroleitungen, mit Oligonukleotid- bzw. Pufferlösungen überspült werden.

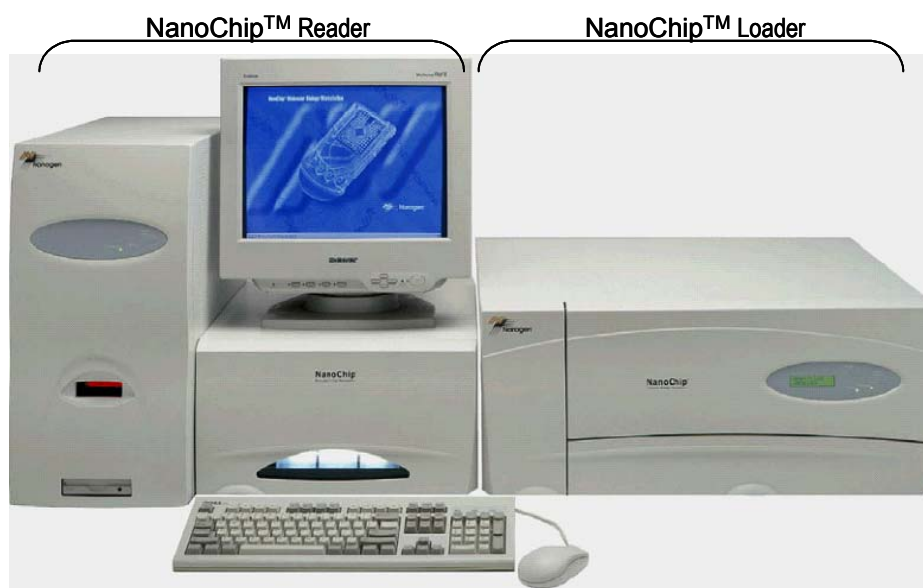


Abbildung 8: Die NanoChip® MBW-Plattform.

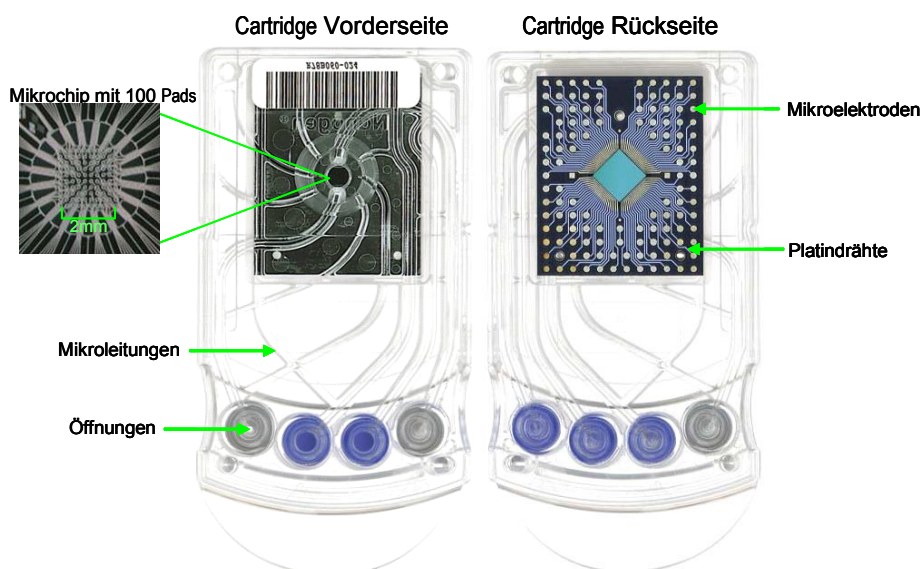


Abbildung 9: Das NanoChip® 100 Cartridge. Der Mikrochip an der Vorderseite und die Mikroelektroden mit den Platindrähten an der Rückseite sichtbar.

Der automatische Loader wird für die elektronische Beladung (Adressierung) des Cartridges oder für die elektronische Hybridisierung verwendet. Dafür werden die Proben mit einer entsprechenden Pufferlösung (His, Histidin oder Arg, Arginin) in eine 96er Mikrotiterplatte überführt und in den Loader gestellt. Ein computergesteuerter Roboterarm saugt die Probe aus der Platte auf und leitet die Flüssigkeit über die Öffnungen und Mikroleitungen des Cartridges über den Mikrochip (Abbildung 9). Der Loader einer MBW-Plattform macht mit vier Steckleisten (Pogo Pin Board) die parallele Beladung von ein bis vier Cartridges möglich. Die Stromversorgung gelangt über Schnittstellenkabel und Anschlussstifte (Pogo Pins) zu den Elektroden des Microarrays.

Der Mikrochip ist mit einer dünnen Hydrogelschicht überzogen. Diese sogenannte Permeation Layer hat bei einer elektronischen Hybridisierung gleichzeitig eine physikalische und eine chemische Funktion. Sobald die Mikroelektrode positiv geschaltet wird, werden Wasserstoff-Ionen ( $H^+$ ) freigesetzt. Die neutralen His-Moleküle der Pufferlösung, welche bei der Hybridisierung die Oligonukleotide beinhaltet, nehmen ein Proton auf und werden positiv geladen (Abbildung 10). Diese kationischen His-Moleküle unterstützen die Hybridisierung (Edman et al. 1997, Sosnowski et al. 2002).

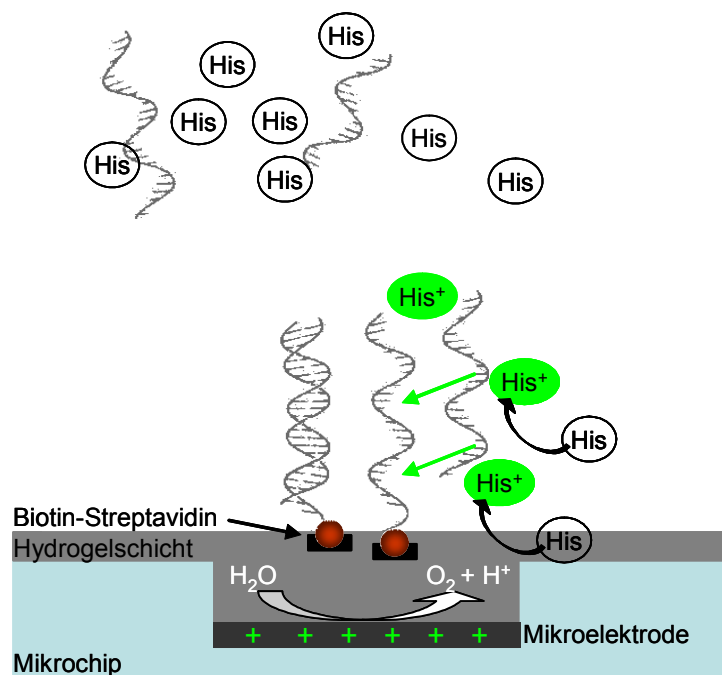


Abbildung 10: Die Wirkung des His-Puffers bei der Hybridisierung. Bei Positivschaltung einer Elektrode oxidiert das Wasser zu Sauerstoff und  $H^+$  Ionen an der Oberfläche. Dabei wird das His-Molekül positiv geladen. Das kationische Histidin neutralisiert die negative geladene Phosphatgruppe der DNA, wodurch die Hybridisierung erfolgt.

Bei der Adressierung spielt neben der mikroelektronischen Manipulation von Molekülen eine weitere Komponente eine entscheidende Rolle. Die Hydrogelschicht beinhaltet jeweils im Bereich der Mikroelektroden Streptavidin, das der Immobilisierung von ssDNA oder Oligonukleotidsonden an der Chipoberfläche dient. Dank der hohen Bindungskonstante der Biotin-Streptavidin-Kopplung können biotinylierte Oligonukleotide an die beschichtete Oberfläche binden (Abbildung 10). Die Biotin-Streptavidin-Bindung ist eine stark affine Bindung (Abbildung 11) (Gonzalez et al. 1997). Das deglykosylierte Streptavidin, das aus

Bakterien gewonnen wird, weist einen isoelektrischen Punkt (IP) von 6,3 auf und hat vier Bindungsstellen für Biotin ( $C_{10}H_{16}N_2O_3S$ ) (Abbildung 11). Wegen seiner geringeren Gesamtladung und aufgrund der Tatsache, dass Streptavidin kein Glykoprotein ist und deshalb nicht an Kohlenhydrat-Rezeptoren bindet, sind unspezifische Bindungen in Streptavidin-Systemen selten (Rico und Moy 2007).

Bei der Verwendung der Nanogen-Plattform wird die Effizienz der Biotin-Streptavidin-Bindung durch Anlegen einer elektrischen Spannung erhöht. So werden die negativ geladenen Oligonukleotide zunächst gezielt mobilisiert und durch die Anreicherung an bestimmten Orten die Biotin-Streptavidin-Kopplung beschleunigt. Damit ist die Möglichkeit gegeben, biotinylierte DNA-Fragmente gezielt an die Chipoberfläche zu binden und ungebundene Moleküle restlos zu entfernen. Aufgrund dieser Gegebenheiten lassen sich verschiedene Analyseassays je nach Fragestellung an der Chipoberfläche aufbauen, ohne sich vorher durch Oligospotting oder Ähnliches festlegen zu müssen (Sosnowski et al. 1997, Gilles et al. 1999, Sosnowski et al. 2002, Weidenhammer et al. 2002).

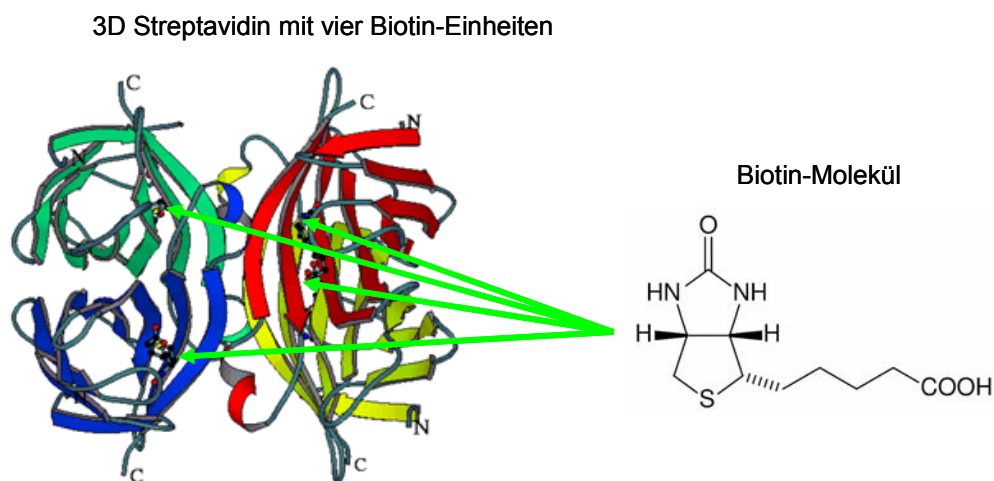


Abbildung 11: Dreidimensionale Struktur des Streptavidins. Streptavidin ist ein aus vier identischen Untereinheiten aufgebautes Protein (Tetramer) aus Streptomyceten (*Streptomyces avidinii*), bestehend aus 254 Aminosäuren und besitzt vier affine Bindungsstellen für Biotin.

Zusätzlich enthält der Loader Gefäße für einen Waschpuffer, der computergesteuert durch Schläuche über den Chip gepumpt wird und in einem Behälter als abgesaugter Abfall gesammelt wird. Nach der Beendigung der automatischen Adressierung und Hybridisierung wird das Cartridge aus dem Loader genommen und ist bereit für diskriminatorische Prozesse, die im Reader erfolgen.

Der Reader beinhaltet den Computer, der die automatische Bedienung und Steuerung der einzelnen Arbeitsschritte und die vollautomatische Datenauswertung ermöglicht. Zusätzlich ist der Reader einer MBW-Plattform zuständig für die Diskriminierung der Proben und für das Scannen der Fluoreszenzfarben. Die Diskriminierung erfolgt nach einer elektronischen, thermalen oder chemischen Stringenz, die die idealen Schlüsselverhältnisse für eine Hybridisierung bzw. keine Diskriminierung darstellt. Dafür sorgt das Pogo Pin Board bzw. ein Heizblock im Reader, über dem das Cartridge automatisch platziert wird. Die Grenzen einer elektronischen Stringenz liegen zwischen 0nA/Pad und 10000nA/Pad; für die thermale Stringenz können Temperaturen von 20°C bis 60°C gewählt werden. Zusätzlich ist eine

chemische Stringenz durch Waschung des Cartridges mit selektiven Pufferlösungen (Tabelle 16) über ein Pumpsystem möglich. Es sind dafür zwei Behälter vorgesehen, einmal für einen Hochsalzpuffer (High Salt Buffer, HSB) und zum anderen für einen Niedrigsalzpuffer (Low Salt Buffer, LSB).

Tabelle 16: Diskriminierungspuffer für die chemische Stringenz.

Diskriminierungspuffer		Komponenten
Arbeitslösungen	LsgA	NaH <sub>2</sub> PO <sub>4</sub> [50mM]
	LsgB	Na <sub>2</sub> HPO <sub>4</sub> [50mM]
Low Salt Buffer	LSB	LsgB + LsgA (pH 7,0)
High Salt Buffer	HSB	LsgB + LsgA (pH 7,4) + NaCl [500mM]

Nach der Diskriminierung detektiert der Reader die Ergebnisse, indem die Fluoreszenzfarbstoffe Grün oder Rot mittels zweier Laser (Abbildung 12) angeregt und die emittierte Lichtintensität gemessen wird. Dafür können flexible und für verschiedene Pads unterschiedliche Parameter gewählt werden (Photomultiplier-Stärke, PMT niedrig, mittel oder hoch und Akkumulationszeit von 128 bis 30000 Millisekunden, oder automatische Akkumulation). Anschließend können die Resultate mittels der Onboard Software (nLAB NanoChip™ Data Viewer & Analysis) ausgewertet werden.

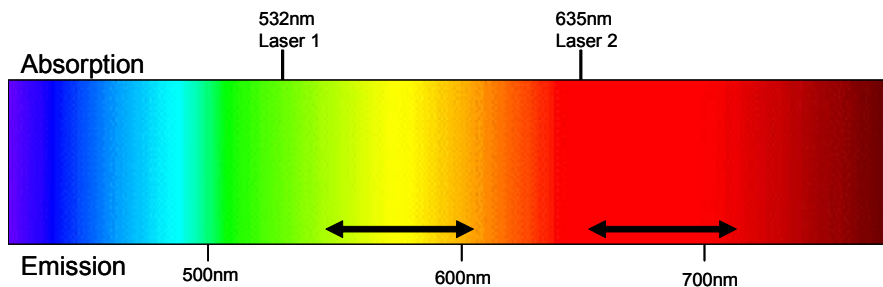


Abbildung 12: Laser Absorption und Emissionsspektrum.

### 2.2.6.2 NanoChip® 400

Vom Prinzip her beruht die NanoChip® 400-Technologie auf dem der NanoChip® MBW, bietet aber bedeutende Entwicklungsfortschritte. Mit dieser Plattform ist die Firma Nanogen ihrem Ziel, ein integriertes Sample-To-Answer System zu entwickeln, ein Stück näher gekommen. Denn hier ist es nur noch nötig eine PCR-Amplifikation voranzuschalten, um dann alle weiteren Schritte vollautomatisch und vorprogrammiert in wenigen Stunden ablaufen zu lassen. Die NanoChip® 400-Plattform stellt eine Systemeinheit für die komplette Analyse ohne manuelle Zwischenschritte dar, indem sie den Loader und den Reader in einem Instrument verbindet (Abbildung 13). Das bedeutet, alle Arbeitsschritte des ganzen

Genotypisierungsprozesses, wie Adressierung, Hybridisierung, Diskriminierung und Detektion, erfolgen in einem einzigen Instrument.

Der Roboterarm garantiert eine präzise Bewegung der Spitze, die die Probe aus einer Mikrotiterplatte aufnimmt und zum Cartridge transportiert. Der Barcode Scanner identifiziert ein bestimmtes Cartridge und ermöglicht die spätere Zuordnung der gespeicherten Informationen. Die Plattform ist über Ethernet mit einem Computer verbunden, der zur Programmierung und Durchführung der Protokolle dient und eine automatische Datenauswertung ermöglicht.

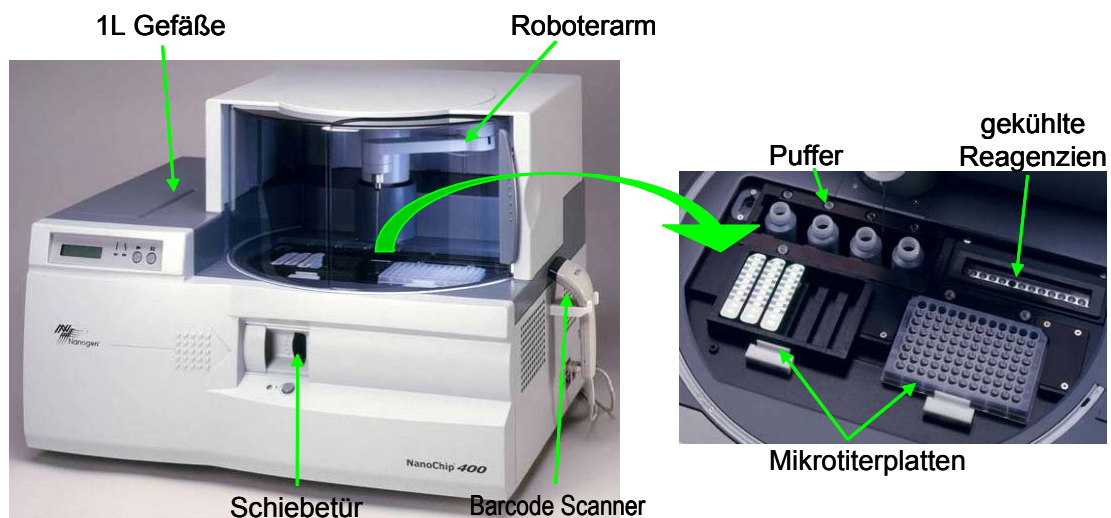


Abbildung 13: NanoChip® 400- Plattform und das Instrumentendeck.

Die Pufferbehälter beinhalten 1L Puffer- oder Waschlösungen und werden durch einen Füllsensor kontrolliert. Das Cartridge wird in die Öffnung gestellt, die durch eine Schiebetür geschlossen wird. An dem Instrumentendeck sind die Plätze für die nötigen Reagenzien vorgesehen, davon zwei für Mikrotiterplatten, die alle zum Protokoll verwendeten Lösungen, wie Proben, Diskriminatoren, Reporter etc. beinhalten. Kleine Behälter stellen verschiedene Pufferlösungen bereit. Die optische Vorrichtung besteht aus einer CCD-Kamera mit LED-Farbfilter für die Detektion von zwei Fluoreszenzfarben.

Die Datenanalyse ist sehr benutzerfreundlich, die Ergebnisse können mittels NanoChip® 400 Data Analysis Software direkt analysiert werden; zusätzlich können die Rohdaten für weitere Auswertungen exportiert werden.

Im NC400 findet ein Complementary Metal Oxide Semiconductor (CMOS) Mikrochip Anwendung. Die CMOS-Technologie verleiht dem Cartridge die Eigenschaft, die Identitätsdaten und alle Verlaufsdaten über die Verwendung, wie z.B. Betriebszustände, Protokolle usw. zu managen (Onboard Memory). Zusätzlich ist durch einen Onboard Thermistor die präzise Temperaturkontrolle gegeben. Das NC400-Cartridge wird nach demselben Prinzip benutzt und ist mit derselben Oberfläche ausgestattet, wie das der MBW-Plattform. Einziger Unterschied ist, dass der Chip des NC400-Cartridges in 400 Pads strukturiert ist, die einzeln oder in verschiedenen Kombinationen aktiviert werden können.

### 2.2.6.3 Assaydesign

Als Ausgangspunkt für die Assayetablierung im Zusammenhang mit der elektronischen Microarray-Technologie dienten die 52 autosomalen Marker (Kapitel 2.1.1). Im Folgenden werden drei Assays (Amplicon down, Capture down und Arrayed Primer Extension) beschrieben, die sich durch ihren Aufbau voneinander unterscheiden. Als erstes wurde das sog. Amplicon down-Assay (Kapitel 2.2.6.3.1) beispielhaft für einige SNPs etabliert, gefolgt von dem Capture down-Assay (Kapitel 2.2.6.3.2), welches schließlich auch für die Validierung mittels autosomaler SNPs verwendet wurde. Das Arrayed Primer Extension Assay (Kapitel 2.2.6.3.3) stellt nur einen kleinen Exkurs dar, welches aus den bereits vorhandenen Oligos entstand, aber nicht weiter verfolgt wurde.

Im Laufe der Amplicon down- und Capture down-Assayetablierung und der Validierung des letzteren wurden zahlreiche Optimierungsmaßnahmen getroffen hinsichtlich einer PCR-Produktbehandlung, gerätespezifischen Abläufen und Microarray-Analysemethoden:

- Zur Optimierung der PCR-Produktbehandlung gehören bestimmte Schritte der PCR-Amplifikation (Kapitel 2.2.4.4) und die Untersuchung der verschiedenen PCR-Aufreinigungsmethoden (Kapitel 2.2.4.5.2), um ein geeignetes, verlustfreies Verfahren für die Analyse der kurzen Fragmente zu ermitteln.
- Unter der Optimierung gerätespezifischer Abläufe sind die Maßnahmen zusammengefasst, die die Funktionalität der Plattform beeinflussen, wie Wasch-, Lade-, Hybridisierungs- und Diskriminierungspuffer, gerätespezifische Einstellungen und das Loader- und Readerprotokoll.
- Die Optimierungsmaßnahmen der Microarray-Analysemethode beinhalten alle Schritte, die zur Verbesserung der Assays unternommen wurden, wie die Konzentrationen der verschiedenen Reagenzien und das Oligodesign bzw. Oligoredesign.

#### 2.2.6.3.1 Amplicon down-Assay

Das Amplicon down-Assay wird ausgehend von dem PCR-Produkt auf dem Mikrochip aufgebaut (Abbildung 14). Dafür wurde für die PCR-Amplifikation ein am 5' Ende biotinylierter Primer verwendet. Das PCR-Produkt (50µl oder zwei Produkte gemischt 100µl) wurde nach der Amplifikation mittels verschiedener Aufreinigungsmethoden (Kapitel 2.2.4.5.2) von überschüssigen Oligos und Salzen befreit. Für die elektronische Adressierung wurden die Proben in His [50mM] (Sigma-Aldrich) gelöst und vier Minuten bei 95°C denaturiert und nach dem Abkühlen in eine 96er Mikrotiterplatte (Nunc) überführt. Zwischenzeitlich wurden die Schläuche des Loaders mit His-Puffer [50mM] automatisch durchspült und ein Loaderprotokoll geschrieben, das den automatischen Ablauf der Adressierung kontrolliert. Die biotinylierten PCR-Produkte wurden meistens doppelt, d.h. jeweils auf zwei Pads 90-120 Sekunden lang bei 2V adressiert. Die biotinylierten Einzelstränge, welche durch Elektrodenaktivierung über bestimmten Pads konzentriert wurden, gehen mit dem Streptavidin in der Hydrogelschicht eine permanente Verbindung



ein. Die restlichen Oligonukleotide wurden mittels eines Waschpuffers (His 50mM und 0,01% v/v TritonX-100, Sigma-Aldrich) automatisch restlos entfernt. Nach der elektronischen Adressierung wurde das Cartridge aus dem Loader genommen und mit Hochsalzpuffer manuell mehrmals gewaschen. Für die anschließende passive Hybridisierung mit einem sogenannten Stabilizer, d.h. einem Oligo, das komplementär zum PCR-Produkt ist und eine stabilisierende Funktion hat, wurde dieser in einer Endkonzentration von 0,2 $\mu$ M in HSB gelöst und dem PCR-Produkt zugesetzt. Zusätzlich wurden die zwei Reporteroligos in einer Konzentration von je 0,5 $\mu$ M dazugegeben und manuell auf den Mikrochip übertragen (Abbildung 14 und Tabelle 17). Die beiden Reporter tragen am 3'-Ende die für das „Wildtyp“- bzw. „Mutante“-Allel spezifische Base und am 5'-Ende unterschiedliche Fluoreszenzfarbstoffe (z.B. Cy3 bzw. Cy5) zur Unterscheidung (Kapitel 2.2.6.4). Das Cartridge wurde mit der Stabilizer/Reporter-Mischung 5-10 Minuten bei Raumtemperatur inkubiert. Die Allel spezifische Bindung der Reporteroligos wird durch die zwischen dem 3'-Ende des Reporters und dem 5'-Ende des Stabilizers bestehende Base Stacking Energy (BSE) aufrechterhalten.

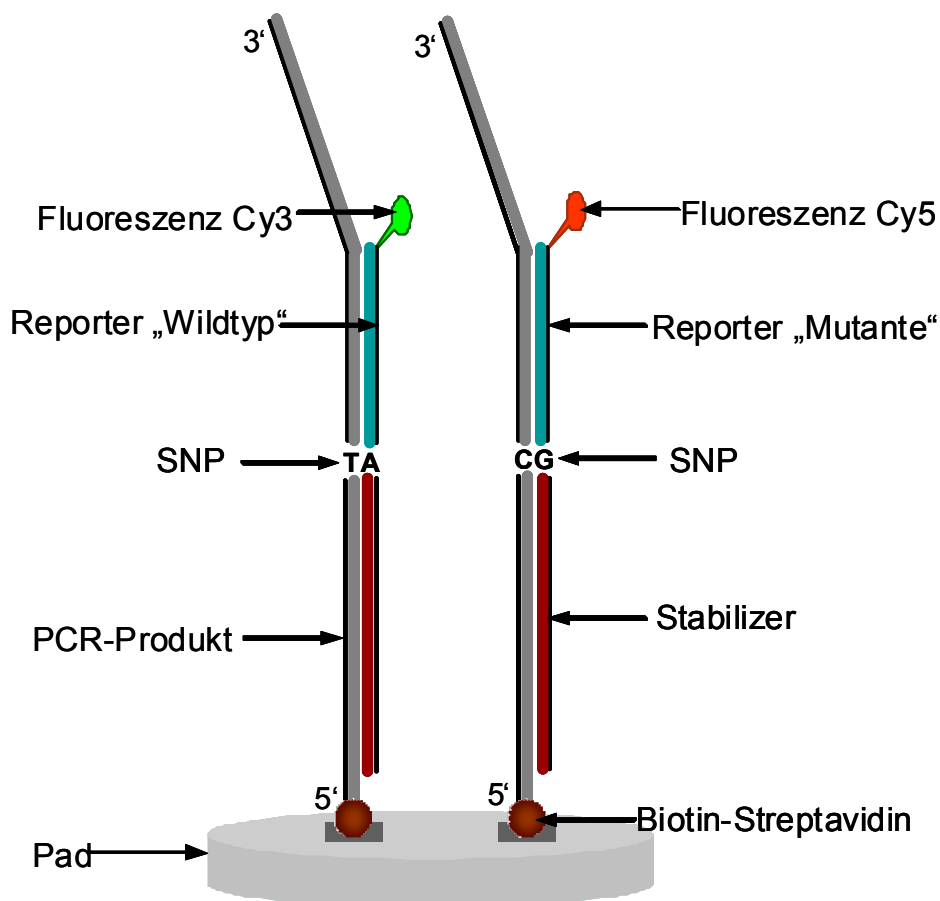


Abbildung 14: Schematische Darstellung des Amplicon down-Assays.

Schließlich wurde das Cartridge zur Diskriminierung in den Reader gestellt, wo das vorprogrammierte Readerprotokoll automatisch abläuft (Tabelle 17). Nach kurzer Inkubation folgte die thermische Diskriminierung, indem die Temperatur von 35°C ausgehend graduell erhöht wurde. Die Temperaturschritte wurden von Waschschritten und Scannprozessen unterbrochen. Wurde ein SNP auf diese Weise bei mehreren Diskriminierungstemperaturen

detektiert, wurde das Readerprotokoll mit einem thermischen Stripping bei 48-52°C beendet (Tabelle 17). Nachdem das Cartridge zusätzlich mit NaOH und einem Niedrigsalzpuffer gewaschen wurde, lagen die immobilisierten PCR-Produkte in Einzelsträngen vor und waren bereit für die Diskriminierung und die Detektion des nächsten SNPs der Multiplex, durch die Zugabe von neuen für den nächsten SNP spezifischen Stabilizer und Reportern und Wiederholung des Reportingschrittes.

Tabelle 17: Ablauf des Amplicon down-Assays mit Angaben zu den Einzelschritten und Optimierungsmaßnahmen bei der Assayetablierung.

Assayablauf	Einzelschritte und Optimierungsmaßnahmen
<b>PCR Ampifikation</b>	---- Singleplex, Duplex und Pentaplex SNPs 01, 02, 05, 09 und 11 ---- nach Standardprotokoll ---- Endvolumen 50µl
<b>PCR-Produkt Aufreinigung</b>	---- ODER mittels Millipore MultiScreen® Separations System: - Mikrotiterplatte mit Filterboden neu bzw. wiederverwendet - Elutionspuffer His [50mM] bzw. HPLC-Wasser mittels Invitex MSB® Spin PCRapace mittels Qiagen MinElute PCR Purification Kit mittels Invitrogen ChargeSwitch® Forensic DNA Purification Kit ---- 50-100µl PCR-Produkt auf 35µl konzentriert
<b>PCR-Produkt Adressierung</b>	---- elektronisch 90-120 Sekunden bei 2V ---- aufgereinigtes PCR-Produkt in 70µl His [50mM]
<b>Stabilizer/Reporter Hybridisierung</b>	---- ODER sukzessiv für jeden SNP passiv 5-10 Minuten bei Raumtemperatur ---- Reporter [0,5µM] und Stabilizer [0,2µM] in 100µl HSB
<b>Reporter Diskriminierung</b>	---- thermisch nach Readerprotokoll: 180" bei RT inkubieren 35°C HSB waschen 36°C HSB waschen und scannen 37°C HSB waschen und scannen 38°C HSB waschen und scannen 39°C HSB waschen und scannen
<b>Stripping</b>	---- ODER thermisch bei 48-52°C und Waschen mit LSB passiv mit 0,1N NaOH und LSB
<b>Auswertung</b>	---- automatisch nach Einstellung: Heterozygote < 1:2 < No Designation < 1:5 < Homozygote Signal-to-noise ratio for Pad Exclusion or No Designation: 5:1 Minimum background-subtracted signal für Pad Inclusion: 50 ---- Background Pad mit His ---- Normalisierung mittels Heterozygotenkontrolle pro SNP

### 2.2.6.3.2 Capture down-Assay

Der Aufbau des Capture down-Assays basiert auf einer sogenannten Capture-Sonde. Diese entspricht der Stabilizer-Oligosequenz mit dem Unterschied, dass sie an ihrem 3' Ende durch einen Spacer aus 6 dNTPs und einem Biotinmolekül verlängert ist (Abbildung 15 und Tabelle 18). Es wurden 0,8µl Capture-Sonde [10µM] für jeden SNP einzeln zu 79,2µl His [50mM]



gegeben und in die 96er Mikrotiterplatte überführt, bis sich alle Capture-Oligos für einen Multiplex in der Platte befanden. Zusätzlich wurden die zu untersuchenden PCR-Produkte separat in die Mikrotiterplatte pipettiert.

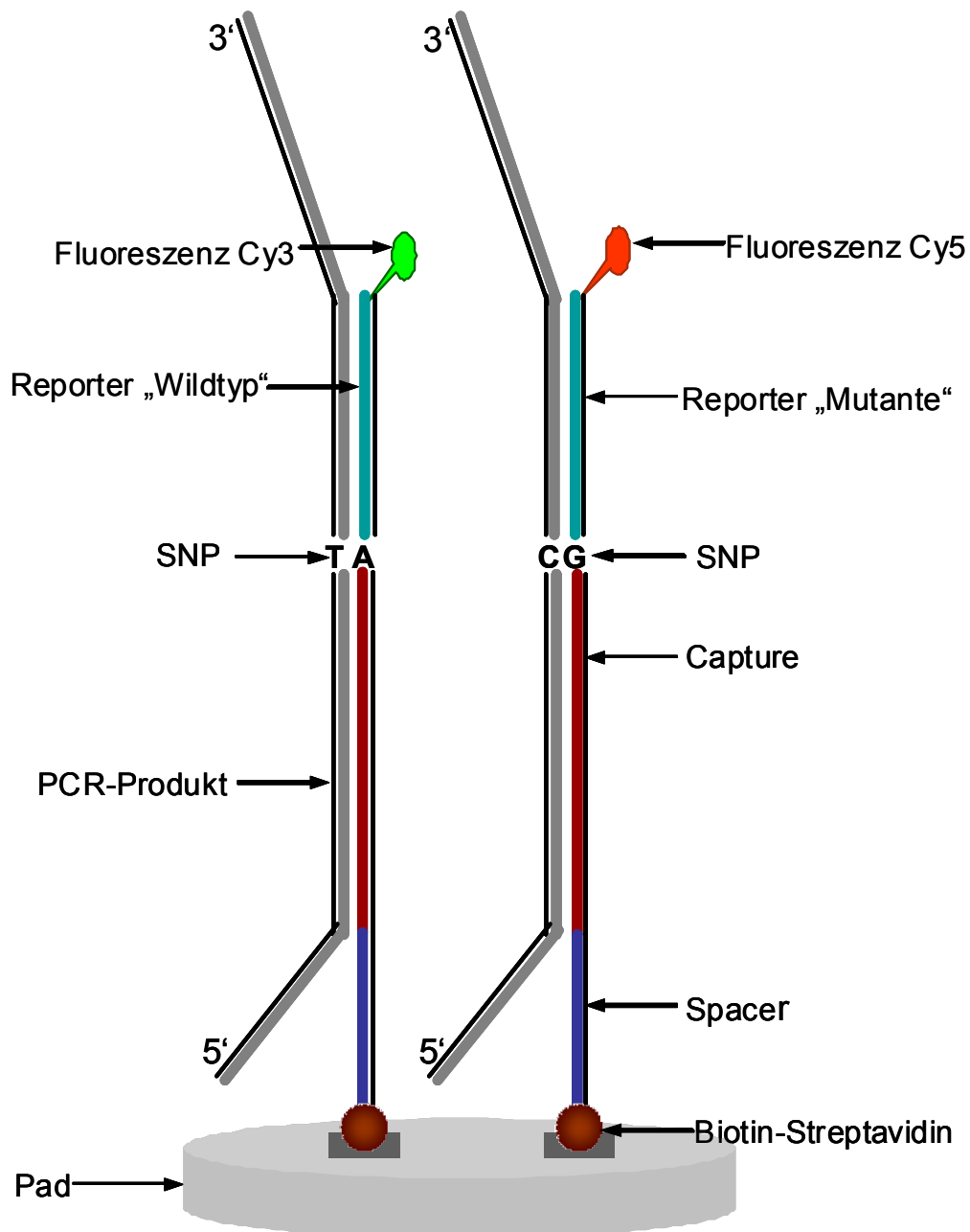


Abbildung 15: Schematische Darstellung des Capture down-Assays.

Zur Vorbereitung der PCR-Produkte gehörte immer die Lösung in His [50mM] (Sigma-Aldrich) oder Arg [100mM] (Sigma-Aldrich) und die Denaturierung bei 95°C für vier Minuten. Für das Capture down-Assay kamen ursprünglich PCR-Produkte zum Einsatz, die mittels Invitek-Kit (Kapitel 2.2.4.5.2) aufgereinigt wurden. Im Verlauf der Validierung wurden dann aufgrund der gewonnenen Erkenntnisse auch ungereinigte Produkte eingesetzt. Die Menge der eingesetzten Produkte wurde von 50µl sukzessiv bis auf 1µl verringert (Tabelle 18).

Tabelle 18: Ablauf des Capture down-Assays mit Angaben zu den Einzelschritten bei der Etablierung und Maßnahmen bei der Optimierung.

Assayablauf	Einzelschritte und Optimierungsmaßnahmen
<b>PCR Ampifikation</b>	<p>---- Singleplex, Duplex bzw. div. Multiplexe (5plex, 13plex, 24plex, 35plex, 48plex)</p> <p>---- Endvolumen 25µl bzw. 50µl</p> <p>nach Standardprotokoll</p> <p>nach veränderten Primerkonzentrationen (Tabelle 9)</p> <p>---- ODER</p> <p>asymmetrische Amplifikation</p> <p>asymmetrische Reamplifikation</p> <p>anschließend Strangtrennung mittels Biotinyliertem Primer</p> <p>PCR-Zusätze Glycerin und BSA</p>
<b>PCR-Produkt Aufreinigung</b>	<p>---- ODER</p> <p>mittels Millipore MultiScreen® Separations System</p> <p>mittels Invitex MSB® Spin PCRapace</p> <p>keine Aufreinigung</p>
<b>Capture Adressierung</b>	<p>---- elektronisch 90-120 Sekunden bei 2V</p> <p>---- Puffervolumen 80µl His [50mM]</p> <p>---- ODER</p> <p>Capture div. Konzentrationen [0,1µM; 0,15µM oder 0,2µM]</p> <p>Capture Redesign</p>
<b>PCR-Produkt Hybridisierung</b>	<p>---- ODER</p> <p>passiv bei Raumtemperatur</p> <p>elektronisch 120 Sekunden bei 2V</p> <p>---- Puffervolumen 70µl HSB [50mM], His [50mM] oder Arg [100mM]</p> <p>---- ODER</p> <p>Invitex gereinigtes Produkt, 10µl-50µl</p> <p>enzymatisch gereinigtes Produkt, 10µl</p> <p>ungereinigtes Produkt, 1µl-10µl</p> <p>ungereinigtes Produkt, sukzessive Hybridisierung von 1µl</p>
<b>Diskriminator Hybridisierung</b>	<p>---- ODER</p> <p>elektronisch 60 Sekunden bei 2V</p> <p>passiv in zwei Schritten: 1. Diskriminatoren und 2. Universal Reporter</p> <p>passiv in einem Schritt, 10 Minuten bei Raumtemperatur</p> <p>---- Hybridisierungsvolumen 35µl HSB [50mM]</p> <p>---- Konzentration Universal Reporter [1µM]</p> <p>---- ODER</p> <p>Diskriminatoren div. Konzentrationen [0,1-0,5µM]</p> <p>Diskriminatoren Redesign</p>
<b>Diskriminierung</b>	<p>---- ODER</p> <p>Diskriminierungsprotokoll, gradient 36-40°C</p> <p>Touch down Protokoll, gradient 52-38°C</p> <p>Kombination Touch down und Diskriminierungsprotokoll: 52°C 30s, 48°C-40°C 10s, 38°C 30s, 25°C 8xHSB 39°C 8xHSB, 40°C 8xHSB, 41°C 8xHSB, 4xLSB, 25°C scan</p>
<b>Auswertung</b>	<p>---- automatisch mittels Onboard Software nach Einstellungen: Heterozygote &lt; 1:2 &lt; No Designation &lt; 1:3,5 &lt; Homozygote Signal-to-noise ratio for Pad Exclusion or No Designation: 3:1 Minimum background-subtracted signal für Pad Inclusion: 10</p> <p>---- statistisch mittels Konvertierungs-, Berechnungs- und Genotypisierungsmakro</p> <p>---- ODER</p> <p>Background Pad mit His</p> <p>Background Pad mit His und PCR</p> <p>Heterozygotenkontrolle</p> <p>---- ODER</p> <p>Normalisierung mittels Tag-Sequenzen, kompl. zu Universal Reporter</p> <p>statistisch mittels Excelmakros</p>
<b>Stripping</b>	<p>---- ODER</p> <p>thermisch bei 52-54°C und waschen mit LSB</p> <p>passiv mit 0,1N NaOH</p> <p>elektronisch 60 Sekunden bei -1,5V</p>

Nachdem die 96er Mikrotiterplatte vorbereitet und der Loader automatisch mit His-Puffer [His 50mM] durchgespült war, erfolgte die Programmierung eines Loaderprotokolls aus zwei Schritten. Als erstes wurden alle Capture-Sonden nacheinander elektronisch adressiert und durch die affine Biotin-Streptavidin-Kopplung permanent an bestimmten Pads gebunden. Nachdem die Adressierung aller Capture-Oligos abgeschlossen war, folgte der zweite Schritt mit der elektronischen Hybridisierung der einzelsträngigen PCR-Produkte. Wenn es sich zum Beispiel um eine Multiplex-PCR für 48 SNPs handelte, wurden 48 einzelne Capture-Adressierungen und eine Hybridisierung einer Multiplex-Probe auf 48 Pads vorgenommen. Nach der elektronischen Hybridisierung wurde das Cartridge aus dem Loader genommen und mit HSB manuell mehrmals gewaschen. Zwischenzeitlich wurde ein 35µl Gemisch aus Wildtyp- und Mutante-Reportern aller SNPs in Konzentrationen von 0,2nM-1nM in HSB hergestellt. Dieses Reportergemisch wurde manuell auf das Cartridge gegeben. Im Verlauf der Experimente wurde das Readerprotokoll für das Capture down-Assay mit verschiedenen Testprotokollen optimiert (Tabelle 18); das Diskriminierungsprotokoll basierte, wie schon beim Amplicon down-Assay, auf einer graduellen Temperaturerhöhung, bis die richtige Temperatur für eine sichere Diskriminierung gefunden wurde. Davon zu unterscheiden ist das Touch down-Protokoll, das mit einer sehr hohen Temperatur startet und diese schrittweise reduziert, bis die Oligos hybridisieren. Während das Prinzip des Diskriminierungsprotokolls darin besteht, die nicht perfekt passenden Oligos zu diskriminieren, bis nur perfekt gebundene Oligos verbleiben, wird bei dem Touch down-Protokoll erwartet, dass nur die perfekt passenden Oligos hybridisieren. Letztendlich wurde ein Readerprotokoll bestehend aus einer Kombination von Touch down- und Diskriminierungsprotokoll entwickelt (Tabelle 18). Das Readerprotokoll sah drei Scans vor, bei 38°C, 40°C und 41°C, für die Validierung reichte jedoch nur ein einziger Scan bei 41°C für alle SNPs einer Multiplex-Analyse.

### **2.2.6.3.3 Arrayed Primer Extension-Assay (APEX)**

Die Arrayed Primer Extension Methode beruht auf dem Prinzip der Minisequenzierung, indem eine Einzelbasenverlängerung stattfindet (Jaakson et al. 2003). Dazu wurde für jeden SNP eine Capture-Sonde elektronisch an einen Pad adressiert, welche mit ihrem 3'-Ende genau vor der SNP-Position endet, woran im nächsten Schritt die einzelsträngigen PCR-Produkte elektronisch hybridisieren (Abbildung 16 und Tabelle 19). Die PCR-Vorbereitung, Capture-Adressierung und Proben-Hybridisierung wurden gemäß eines Standard Capture down-Assays (Kapitel 2.2.6.3.2) präpariert. Danach wurde das Cartridge aus dem Loader genommen und manuell mit Na<sub>2</sub>HPO<sub>4</sub> Puffer [50mM] 10 Minuten bei Raumtemperatur inkubiert und gewaschen. Anschließend wurde ein Gemisch aus DNA-Polymerase (ThermoSequenase, Amersham Biosciences) und fluoreszenzmarkierten ddNTPs (ddUTP mit Cy5 und ddATP und ddCTP mit Cy3, PerkinElmer Life and Analytical Sciences) hergestellt und manuell auf das Cartridge überführt (Tabelle 19). Während eines 30-minütigen automatischen Inkubationsschritts im Reader bei 48°C konnten die ddNTPs an den freien 3' Enden der Capture-Sonden eingebaut werden. Schließlich folgte nach mehreren Waschschritten das Scannen. Mittels APEX wurden acht autosomale SNPs in einer Multiplex typisiert.

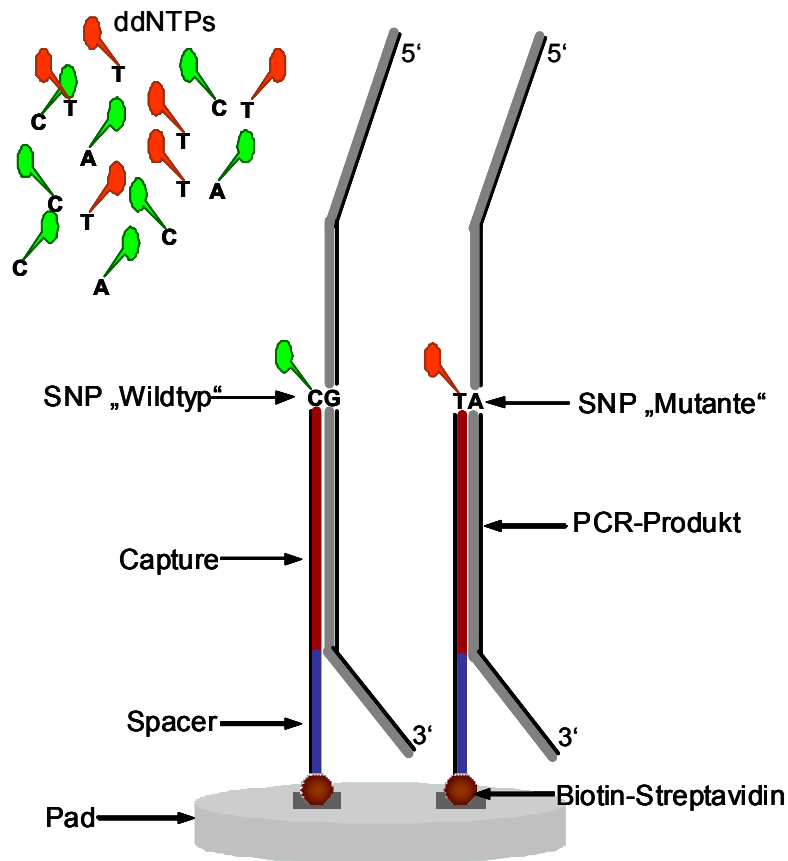


Abbildung 16: Schematische Darstellung des APEX-Assays.

Tabelle 19: Ablauf des APEX-Assays mit Angaben zu den Einzelschritten bei der Etablierung.

Assayablauf	Einzelschritte
<b>PCR Amplifikation</b>	<ul style="list-style-type: none"> <li>---- 8plex-PCR (SNPs 24, 25, 27, 28, 29, 33, 35 und 40)</li> <li>---- nach Standardprotokoll</li> <li>---- Endvolumen 50µl</li> </ul>
<b>PCR-Produkt Aufreinigung</b>	<ul style="list-style-type: none"> <li>---- mittels Invitex MSB® Spin PCRapace</li> </ul>
<b>Capture Adressierung</b>	<ul style="list-style-type: none"> <li>---- elektronisch 120 Sekunden bei 2V</li> <li>---- Puffervolumen 80µl His [50mM]</li> <li>---- Capture-Sonde [0,1µM]</li> </ul>
<b>PCR-Produkt Hybridisierung</b>	<ul style="list-style-type: none"> <li>---- elektronisch 120 Sekunden bei 2V</li> <li>---- Puffervolumen 70µl His [50mM]</li> <li>---- Invitex gereinigtes PCR-Produkt, 50µl</li> </ul>
<b>Primer Extension</b>	<ul style="list-style-type: none"> <li>---- Enzym-Mischung Endvolumen 30µl: 6µl 5x1st Buffer + 3µl 0,1M DTT + 1,5µl ddNTP Mix [je 100µM] + 1,5µl Polymerase [200U/µl] + 18µl HPLC</li> <li>---- 30 Minuten bei 48°C inkubieren</li> </ul>
<b>Detektion</b>	8xHSB, scannen
<b>Auswertung</b>	<ul style="list-style-type: none"> <li>---- automatisch nach Einstellung: Heterozygote &lt; 1:2 &lt; No Designation &lt; 1:3,5 &lt; Homozygote Signal-to-noise ratio for Pad Exclusion or No Designation: 5:1 Minimum background-subtracted signal für Pad Inclusion: 50</li> <li>---- Background Pad mit His</li> </ul>

### 2.2.6.4 Oligonukleotiddesign

Der unterschiedliche Assayaufbau benötigte individuell angepasste Oligonukleotide und Sonden, deren Design nach assayspezifischen Kriterien erfolgte. Die Voraussetzung für das Oligodesign waren die bereits validierten PCR-Primer der 52 autosomalen SNPs, die das Amplifikationsprodukt und damit die Ausgangssequenzen für jeden SNP bestimmten. Für die ersten Versuche wurden Oligonukleotide und Sonden für fünf SNPs (01, 02, 05, 09 und 11) entworfen, um diese mittels Amplicon down- sowie Capture down-Assay zu typisieren. Mit zunehmender Erfahrung und der sich abzeichnenden Vor- und Nachteile der einzelnen Assays richtete sich die Konzentration auf das Oligodesign für den Capture down-Assay. Zwischenzeitlich wurden für acht SNPs (24, 25, 27, 28, 29, 33, 35 und 40) auch Capture-Sonden für das APEX-Assay bestimmt.

Während die PCR-Primer der 52 SNPs für das Capture down- und APEX-Assay unverändert blieben, wurde für das Amplicon down-Assay ein PCR-Primer je SNP am 5' Ende biotinyliert.

Alle weiteren Oligonukleotide und Sonden wurden mittels der kostenlosen Online Software Primer 3.0 v. 0.2 unter Berücksichtigung der Assayvorgaben (Kapitel 2.2.6.3) ausgewählt.

Der Stabilizer ist ein Oligonukleotid für das Amplicon down-Assay, dessen 5'-Ende direkt eine Base vor dem Polymorphismus endet. Die wichtigste Assayvorgabe war, dass die Schmelztemperatur des Stabilizers mindestens 50°C, idealerweise größer als 55°C sein sollte, damit die Hybride beim Reporting nicht denaturieren. Gleichzeitig wurde die Länge nicht unnötig vergrößert, um die Dimerisierungsgefahr niedrig zu halten. Bei einigen SNPs reichte in eine Richtung die Länge der flankierenden Sequenz nicht aus, um einen Stabilizer zu platzieren. In solchen Fällen musste die Analyse auf den anderen Strang umgelegt werden. Die für das Amplicon down-Assay gewählten Stabilizer konnten als Capture-Sonden für das Capture down-Assay eingesetzt werden, indem an ihren 3'-Ende ein 6Poly dTTP oder 6Poly dCTP, als Spacer und ein Biotin angehängt wurden. Für das APEX-Assay mussten die Capture-Sonden umgekehrt an den 5'-Enden Biotin besitzen, um ihre 3'-Enden für eine ddNTP-Extension frei zu halten.

Die Reporter sind jeweils zwei Oligos pro SNP, die sich durch eine Base und ihre Fluoreszenzmarkierungen unterscheiden. An dem 3'-Ende der Reporteroligos befindet sich der Polymorphismus, entweder als „Wildtyp“- oder als „Mutante“ und dementsprechend an ihrem 5'-Ende die Fluoreszenzmarkierung durch Cy3<sup>TM</sup> für den Wildtyp und Cy5<sup>TM</sup> für die Mutante (Tabelle 20).

Tabelle 20: Die Laser Absorption und Fluoreszenz Emission der verschiedenen Farbstoffe.

Fluoreszenzfarbstoff (Firma)	Absorption [nm]	Emission [nm]	Farbe
Cy3 <sup>TM</sup> (Amersham Bioscience)	550	570	Grün
Alexa Fluor <sup>®</sup> 532 (Molecular Probes)	532	554	Grün
Cy5 <sup>TM</sup> (Amersham Bioscience)	649	670	Rot
Alexa Fluor <sup>®</sup> 647 (Molecular Probes)	650	665	Rot

Größter Wert wurde darauf gelegt, dass die Schmelztemperaturen der zwei Reporter eines SNPs nah beieinander und nach Möglichkeit für alle Reporter um die 40°C liegen. Damit sollten die gleichen Voraussetzungen für beide Reporter eines SNPs bei der Diskriminierung geboten werden. Zusätzlich wurde darauf geachtet, dass die Reportersequenzen keine homopolymeren Sequenzen oder Dimeren aufweisen und der AT-Anteil nicht zu hoch ist, damit die Längen der Oligos nach Möglichkeit 9-12nt nicht überschreiten. Vor dem Hintergrund, dass alle Reporter, die für eine Multiplex-Typisierung mittels Capture down-Assay benötigt werden, in einer Lösung zusammengemischt werden, wurden die Interaktionen aller Oligos untereinander überprüft. Dafür wurde die Online zur Verfügung stehende FastPCR<sup>9</sup> Software benutzt, mit der es gelingt, alle Kombinationen der Oligos (z.B. insgesamt 104, für 52 SNPs) aufzuzeigen und nach einem vorher definierten Maß einem Interaktionstest zu unterwerfen.

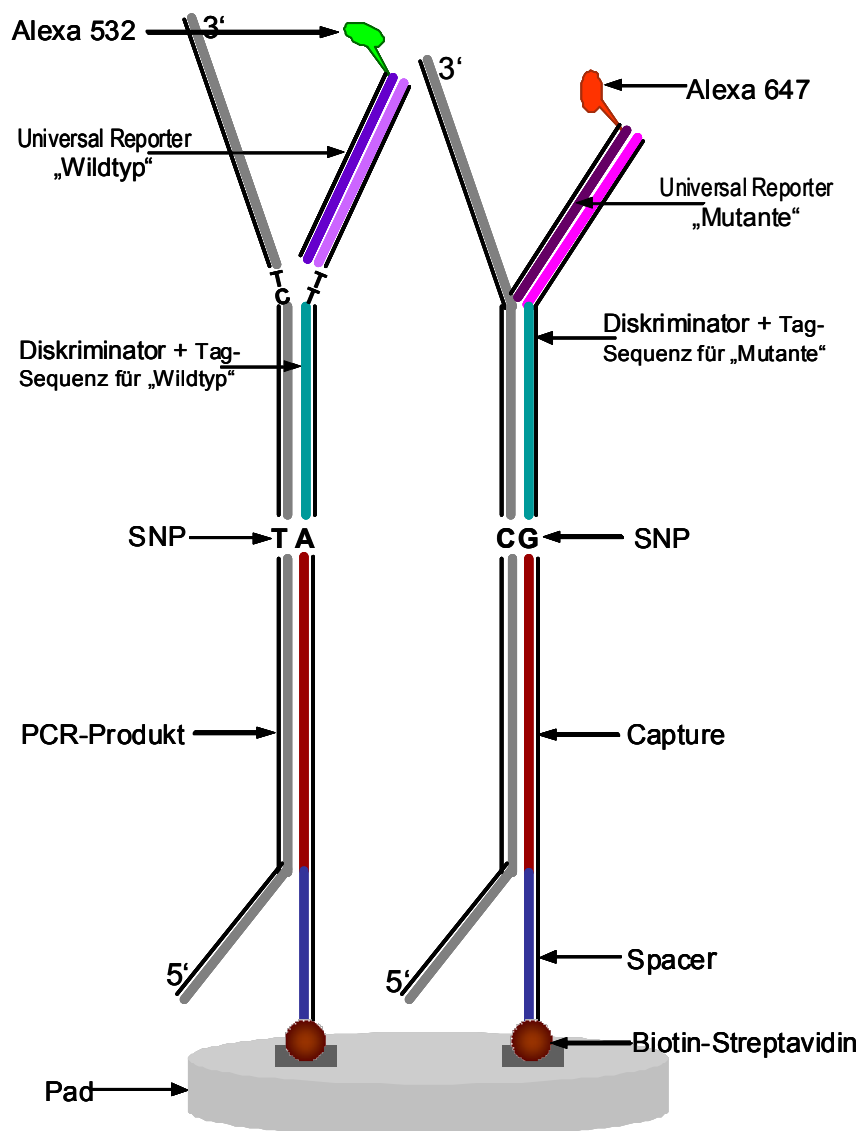


Abbildung 17: Capture down-Assay nach der Einführung der Universal Reporter-Oligonukleotide.

<sup>9</sup> <http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm>

Nach einer kurzen Versuchsphase wurde das Reporting des Capture down-Assays in zwei Schritte aufgeteilt, um damit die hohen Kosten zu vermeiden, die die Fluoreszenzmarkierung jedes einzelnen Reporteroligos verursacht (Abbildung 17). Die Reporteroligos, die nach oben genannten Vorgaben ausgewählt wurden, erhielten statt einer Markierung je nach Allel eine von zwei universellen Tag-Verlängerungen an ihrem 5'-Ende. Diese wurden Diskriminatoren genannt. Um die Designvorgaben eines Reporters auch für die verlängerten Diskriminatoren beizubehalten, wurde darauf geachtet, dass die Hybridisierung des PCR-Produktes nicht über die Diskriminatoren hinaus bis in die Tag-Sequenz hinein möglich war, da das eine Änderung der Diskriminierungstemperatur zur Folge gehabt hätte. In solchen Fällen wurden zwischen Diskriminator und Tag-Sequenz zwei Basen eingefügt, die nicht komplementär waren zu dem PCR-Strang (siehe Wildtyp in Abbildung 17). Zusätzlich zu den Diskriminatoren wurden zwei mit Alexa Fluor® 532 für den Wildtyp bzw. mit Alex Fluor® 647 für die Mutante markierte (Tabelle 20) und zur Tag-Sequenz komplementäre Universal Reporter für die Detektion benötigt (Tong und Ju 2002) (Abbildung 17). Dafür wurden zwei beliebige Oligosequenzen verwendet (Goering 2003), die keine natürlich vorkommenden Sequenzen beinhalten und aus 18 Basen bestehen. Ihre Schmelztemperatur liegt um 50°C und damit höher als die Diskriminierungstemperatur während des Readerprotokolls.

Nachdem das Oligonukleotiddesign für alle SNPs und für den forward-, sowie den reverse-Strang abgeschlossen war, wurden gleich vier SNPs (06, 09, 39 und 51) aus dem 52 autosomalen Multiplex ausselektiert, die die Designvorgaben nicht erfüllen konnten, weil z.B. viele homopolymeren Sequenzen vorlagen, oder bedingt durch die Länge des Amplikons das Design nicht möglich war (z.B. SNP06 und SNP51). Damit standen 48 autosomale SNPs für die Etablierung der elektronischen Microarray-Technologie bereit. Anschließend wurde die Entscheidung über die Analyserichtung für diese 48 SNPs nach drei Kriterien getroffen (Abbildung 18):

- Als wichtigstes Kriterium zählte, dass die Base Stacking Energies des „Wildtyps“ und der „Mutante“ eines SNPs zwischen dem 3'-Ende der Reporter/Diskriminatoren und dem 5'-Ende des Stabilizer/Capture-Oligos möglichst klein und gleich sind. Nach Möglichkeit wurde ein Base Stacking zwischen TA, TT oder AA vermieden (Tabelle 21).
- Als zweites Kriterium galt die Minimierung des Abstandes Delta zwischen den Schmelztemperaturen der zwei Reporter/Diskriminatoren, um eine zuverlässige Diskriminierung der match und mismatch Oligos zu gewährleisten.
- Als Drittes wurde die Sekundärstruktur des PCR-Produkts unter Berücksichtigung der Na<sup>+</sup>, Mg<sup>++</sup> Konzentrationen und bestimmter Temperaturbedingungen mittels der HTML-Software mfold Version 3.2<sup>10</sup> und nearest neighbor Thermodynamik in einem Bildformat als .gif Datei dargestellt (Rouillard et al. 2003, Zuker 2003). Auf diese Weise konnte festgestellt werden, in welcher Richtung vom SNP aus störende Schleifen in der Sekundärstruktur entstehen und ein Stabilisator nötig ist, bzw. in welcher Richtung keine Schleifen die Hybridisierung der Diskriminatoren stören.

---

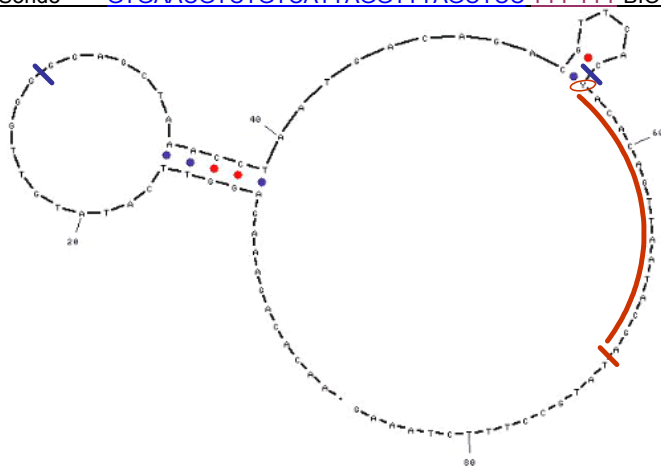
<sup>10</sup> <http://frontend.bioinfo.rpi.edu/zukerm/cgi-bin/rna-index.cgi>

Nach zahlreichen theoretischen und experimentellen Optimierungen wurden schließlich die in der Tabelle 22 und Tabelle 23 aufgeführten Oligos und Sonden für die drei Assays verwendet.

Forward Strang		Mutation R G>A	BSE
5' (A)CAGACGTTACAC		C/TACACAGTTAATACGATATGCCTTTCTAAAG 3'	-6.57 -3.82
<b>Oligonukleotide</b>		<b>5' --- 3'</b>	<b>T<sub>m</sub></b>
Reporter_Wildtyp	CAGACGTTACAC		38,7°C
Reporter_Mutante	ACAGACGTTCACT		40,8°C
Capture-Sonde	ACACAGTTAATACGATATGCCTTTCTAAAG	TTTTT BIO	57,3°C
Reverse Strang		Mutation Y C>T	BSE
5' (T)CGTATTA		ACTGTGTG/A GTGAACGTCTGTCATTAGGTTAGCTCC 3'	-8.26 -6.78
<b>Oligonukleotide</b>		<b>5' --- 3'</b>	<b>T<sub>m</sub></b>
Reporter_Wildtyp	CGT ATT AAC TGT GTG		39,8°C
Reporter_Mutante	TCG TAT TAA CTG TGT A		40,4°C
Capture-Sonde	GTGAACGTCTGTCATTAGGTTAGCTCC	TTT TTT BIO	60,8°C

BSEs Minimierung, homolog für die Reporter

Delta T<sub>m</sub> Minimierung



Sekundärstruktur in Reportersequenz vermeiden

Abbildung 18: Oligodesign für ein SNP und das Capture down-Assay. In der Tabelle sind die Oligos für forward bzw. reverse Strang aufgelistet. Anhand der drei Oligodesignkriterien wurde in diesem Fall der reverse Strang ausgewählt. Dieser zeigt homologe BSEs und Delta T<sub>m</sub> Werte und es befinden sich keine störenden Schleifen in der Sekundärstruktur.

Tabelle 21: Base Stacking Energies.

Base Stacking Energies 5' --- 3'			
GC	-14.59	AG	-6.78
GT	-10.51	CT	-6.78
AC	-10.51	AT	-6.57
GA	-9.81	CA	-6.57
TC	-9.81	TG	-6.57
CG	-9.69	AA	-5.37
GG	-8.26	TT	-5.37
CC	-8.26	TA	-3.82



Tabelle 22: Oligonukleotide für eine Amplicon down-Typisierung von fünf SNPs.

SNP	Stabilizer 5' --- 3'	Tm °C	Reporter <sup>1)</sup> 5' --- 3'	Tm °C	BSE
01	TGCAAAGAACCCACATCAGCCC	61,7	GCCAGCCAGTTT	42,5	-5,37
			GCCAGCCAGTTC	43,2	-6,78
02	TACTTATACTTCAGTGCAGTGGTATAAAAAATGGAG	62,2	GTGTGACAAAAATG	39,5	-10,5
			AGTGTGACAAAAATA	39,7	-6,57
05	GTGAACGTCTGTCATTAGGTTTAGCTCC	60,8	CGTATTAAGTGTG	39,8	-8,26
			TCGTATTAAGTGTGTA	40,4	-6,78
09	CTGTCCCAGTGATTTTTTCATAAGCATTAAACAG	60,6	CCAGGTGTTTTATTG	40,3	-14,59
			CCAGGTGTTTTATTC	40,0	-8,26
11	GGGCTGATTCAGCTAGTTTGCACC	61,5	GTTAGCTATCAACG	38,0	-8,26
			TGTTAGCTATCAACA	39,7	-6,78

1) Reporter für Wildtyp und Mutant untereinander dargestellt, die Base am 3' Ende ist der SNP.

Tabelle 23: Capture-Sonden und die Diskriminatoren für eine Capture down-Typisierung von 48 SNPs bzw. für eine APEX-Typisierung von acht SNPs.

SNP	Capture 5' --- 3' + 6 dNTP + Biotin	Tm °C	Diskriminator Univ. Reporter <sup>1)</sup> + 5' --- 3'	Tm °C	BSE
01	TGCAAAGAACCCACATCAGCCC	61,7	TT + GCCAGCCAGTTT	42,5	-5,37
			TT + GCCAGCCAGTTC	43,2	-6,78
02	TACTTATACTTCAGTGCAGTGGTATAAAAAATGGAG	62,2	GTGTGACAAAAATG	39,5	-10,5
			AGTGTGACAAAAATA	39,7	-6,57
03	CCCCAAATTCAGCTTGGGAAGTCAAATTAAG	59,9	CATGACCAAGCA	38,8	-10,51
			CATGACCAAGCT	38,3	-9,81
04	TAATTTAATTTTAAAATGATCGTTCATAGACAATAGATAC	56	CC + CACTTCACCTTCTAA	40,5	-6,57
			CC + CACTTCACCTTCTAT	40,2	-5,37
05	GTGAACGTCTGTCATTAGGTTTAGCTCC	60,8	CC + CGTATTAAGTGTG	39,8	-8,26
			CC + TCGTATTAAGTGTGTA	40,4	-6,78
07	GCTCGTTGACCTCAGTCATCTCTTAC	58,6	TT + TCAACATTTTACAAGAT	39,6	-9,69
			TT + CAACATTTTACAAGAC	38,8	-6,57
08	TTGAGTTAGAAATAATAAACAACATCCTCATATACTTG	56,2	TT + TCAAATGTAATGCAG	39,3	-6,57
			TT + CTCATAATGTAATGCAA	40,6	-10,51
10	CTTCCCGTTGAATTCGGTGACACTC	59,7	TT + GAGACCTGCTGT	41,3	-9,81
			TT + AGACCTGCTGC	41,2	-8,26
11	GGGCTGATTCAGCTAGTTTGCACC	61,5	CC + GTTAGCTATCAACG	38,0	-8,26
			CC + TGTTAGCTATCAACA	39,7	-6,78
12	CTTTTCTCCAAAACAGTTGAATAATGACTCA	56,1	TCTAAATCATATTGTCTG	40,2	-14,59
			TTTCTAAATCATATTGTCTA	40,7	-10,51
13	AATATTGCTTATTCTTTAAGAATTAATAAAATTTACCCTCTGG	58	ACAACACTTGCG	40,3	-9,81
			AACAACACTTGCA	39,6	-5,37
14 <sup>3)</sup>	TGTTGTTTCTAAATGGATACTGAAAAGTGTCTTACTG	58,6	TT + CCTGAGCTGCT	40,7	-5,37
			TT + CCTGAGCTGCA	41,2	-6,57
15 <sup>2)</sup>	GCACATTTTCTCTCTCAGCCACAC	59,7	AA + CACACATTCAGCTA	40,4	-6,78
			AA + ACACATTCAGCTG	40,1	-8,26
16	GTCTGCAGAGGAGCCTTGGT	59,5	AA + CGTTACATACATTCT	41,8	-6,57
			AA + GGTTACATACATTCTG	39,5	-8,26
17	TTTGTAAGGCAAAACAAAATAAGCCAAAACCTCGAAC	59,6	TT + TGAGTGGTATAATCATATA	41,2	-6,57
			AA + GAGTGGTATAATCATATG	40,8	-10,51
18 <sup>2)</sup>	CAAAAATAACAGGCCCTCACATAAACAGTTTG	57,4	AA + GTGCTAGGCCA	40,6	-10,51
			AA + GTGCTAGGCC	42,7	-8,26
19	CTAGAATGGGAGGAGGAAAAGCAGG	58,9	TT + AGAGGAGTAGCTG	40,6	-14,59
			TT + CAGAGGAGTAGCTA	40,9	-10,51
20 <sup>3)</sup>	CCCCGAGCATACTTGAAAGC	56,1	ATGAGTATTTATTTATCTAAC	41,2	-10,51
			TT + TGAGTATTTATTTATCTAAA	39,4	-8,26
21 <sup>3)</sup>	TAAGGATTTAAATTTGGATGTGCTGAATATTTCTTTAC	57,1	GTAATGAAAATCCTTGG	41,9	-10,51
			TT + GTAATGAAAATCCTTGA	40,5	-6,57
22 <sup>3)</sup>	TTTCCAAGTGGGATGGCCTGTTG	60	TT + TGTCAGGCAAAC	39,1	-6,78
			TT + ATGTCAGGCAAAT	38,4	-5,37
23	TGAGAGATAGGTGACTAATAACAGAGACAAAATTC	57,4	CC + CTGCTATCCTGG	38,3	-10,51
			CC + TCTGCTATCCTGA	39,0	-6,57
24 <sup>3)4)</sup>	AGTTAGAAGGCCAAAGACAGATAATTGTGG	58,6	TA + TTGCCGGAGATG	42,4	-9,81
			TT + TATTGCCGGAGATA	40,9	-5,37

Tabelle 23: Fortsetzung.

SNP	Capture 5' --- 3' + 6 dNTP + Biotin	Tm °C	Diskriminator		Tm °C	BSE
			Univ. Reporter <sup>1)</sup> + 5'	--- 3'		
25 <sup>4)</sup>	GGGCAGCACTGGGATTTGAATG	59,1	TT + ATGTTACATTACAAA		38,8	-6,78
			TT + ATGTTACATTACAAG		39,8	-8,26
26	TCCCTGCAGTGGGCAGATG	60,1	TT + ACCTGAACTTTCAAT		40,3	-5,37
			TT + CCTGAACTTTCAAG		38,7	-10,51
27 <sup>4)</sup>	GAAGGATGCTGCCAAGATGCT	57,9	TT + GGCTGGAGCTT		41,1	-6,57
			TT + GGCTGGAGCTC		42,6	-9,69
28 <sup>4)</sup>	TCTCATTCATGCCTGTCTCCCTAACAC	59,9	AA + CTTGAACTCCCG		39,6	-10,51
			AA + TCTTGAACCTCCA		40,1	-6,57
29 <sup>4)</sup>	CACTTAATAAAGTATGCCTTTGTATTACTTTTGCTCAC	59,2	TT + AAAGATTGCTCATA		39,7	-10,51
			TT + AGGATTGCTCATG		39,3	-14,59
30	AATCATCTGCATCTCCCAGCAATCTC	60	TT + ACTTCAACGACTTAA		40,0	-5,37
			TT + CTTCACGACTTAC		38,6	-6,57
32 <sup>3)</sup>	CCCATGCTCCACCACTGTGAC	60,3	CC + TTCATATAAGTGAAATGT		39,7	-9,81
			CC + TCATATAAGTGAAATGC		40,3	-8,26
33 <sup>4)</sup>	CGGTCACATCTAACGTAAGGGGAA	59,3	CC + TCCTATCTCCAAATG		39,9	-14,59
			CC + GCCTATCTCCAAATA		40,2	-10,51
34	ATCCATAGGAGAAGAAGTTCGTGGAAATTCA	59,3	CC + GGAACCTTGCTG		41,2	-9,81
			CC + GGAACCTTGCTC		40,8	-6,57
35 <sup>4)</sup>	TGCAGCCATTCCCACCCAAAAG	60,7	TT + ACTCATAGCCACA		40,3	-6,57
			TT + CTCATAGCCACG		39,5	-10,51
36	GCCCTGATTTCTGGTGGTGAGG	60,2	AA + TCAGGTCCATCAT		39,8	-6,57
			AA + TCAGGTCCATCAA		40,2	-6,78
37	CATCCTCGAGGCCTCCAGTG	59,8	AA + TTGGCAAGTCCT		40,3	-9,81
			AA + TGGCAAGTCCC		41,0	-8,26
38	ATGCCAGGGCATCATGAATTGCC	60,7	AA + GGAGCCTTCTTTG		41,2	-9,81
			AA + GGAGCCTTCTTTC		40,7	-6,57
40 <sup>4)</sup>	TATTTTCATGAGAACTGGTTTTTCATAGGCGTTCTG	60	CTTACTCTAAGTGCA		40,3	-6,57
			TTACTCTAAGTGCG		39,7	-10,51
41	GGTGGAGATGGCCTCACATGT	59,4	AA + CCCAGGGATGAG		42,1	-9,81
			AA + CCCAGGGATGAA		40,7	-6,78
42	CTTTAGACAATAAAATGGTGTGTCTATGTCTATAGAAG	58,1	CC + AGGACTTCTTTGCT		42,2	-9,81
			CC + GGACTTCTTTGCC		41,6	-8,26
43	GGGACAGCCAGATATCTACTACTTCATCTG	59,1	GGCTTTTAAAGTAGAC		39,3	-9,69
			AGGCTTTTAAAGTAGAT		40,0	-6,57
44	GCCCTCTTTGGTGCCCATCA	60,5	AA + CCTGAGGGTGTT		41,1	-6,57
			AA + CCTGAGGGTGTC		42,5	-9,69
45 <sup>3)</sup>	TTTCAATGAGAATCCAATGCAACCAAGAGG	59,4	TT + CAGAATGGAATAGGAT		41,0	-5,37
			TT + AGAATGGAATAGGAC		39,9	-6,78
46	TCTTCCCAGGTCTCCTCTACTCC	59	TT + GCAGCCAGTG		38,5	-10,51
			TT + TGCAGCCAGTA		38,6	-6,57
48	GGGCATTTGGGGCCACAG	60	TT + GAACGTCGACAC		40,8	-9,69
			TT + GAACGTCGACAT		38,7	-6,57
49	AGTCGCCGCTGTTTCAGGG	60,6	TT + AGCCGTGGAATTA		41,3	-5,37
			TT + GCCGTGGAATTG		41,7	-9,81
50	TAACTCTTACATTATTACTATGGAGGTGTAGCAAAGTGA	59,5	CTTGCGCTACG		40,4	-10,51
			TCTTGCGCTACA		40,7	-6,57
52	CCACGGCCAACAATACATACCTGC	60,3	TA + GCTATGTAAGTACTGATT		40,7	-9,81
			TA + GCTATGTAAGTACTGATA		41,7	-10,51
53 <sup>2)</sup>	CTG TGT GCA GAT CCG CGG AG	60,8	CCTTACTCGACATCA		43,5	-10,51
			CTTACTCGACATCG		41,3	-14,59
54	TCTTAAATATGTTAAGGCTAAATATTGATCTCCAGGA	57,5	ACTGTAACATTTTTCAG		41,0	-10,51
			TACTGTAACATTTTTCAA		40,5	-6,57

1) Universal Reporter Wildtyp: Alexa Fluor 532 + CTGAGTCCGAACATTGAG, Mutant: Alexa Fluor 647 + GCAGTATATCGCTTGACA.

2) Beim Redesign Richtung gewechselt.

3) Beim Redesign ein oder beide Reporter verändert.

4) Capture-Sonden für APEX-Assay am 5' Ende biotinyliert.

### 2.2.6.5 Auswertung der elektronischen Microarray Daten

Unmittelbar nach einem Versuch auf der MBW-Plattform konnten Hinweise auf einen Typisierungserfolg durch die Onboard Software,  $n$ LAB™ NanoChip® Data Viewer and Analysis ermittelt werden. Ist ein bestimmtes Cartridge ausgewählt (schwarzer Pfeil in Abbildung 19), können die Aktivierungsdaten während des Loaderprotokolls überprüft werden (blauer Pfeil

in Abbildung 19). Dabei werden Informationen angegeben, wie die Anzahl der aktivierten Pads (+pad count: 2), gewählte und erreichte Amplitude (Amplitude: 2.000, 2.007V), durchschnittliche, höchste und niedrigste Aktivierungsstärke in nA pro Pad und für beide Pads (z.B. Mean: 467.77nA/pad und Mean: 935.54nA). Anschließend werden die Einstellungen für die Auswertung gewählt (violette Pfeile in Abbildung 19). Als Erstes kann der Mikrochip bei Bedarf konfiguriert werden, indem die Beladungen der einzelnen Pads definiert werden (NanoChip™ Config in Abbildung 19). Dabei können auch mehrere sog. Submaps eines Loaderprotokolls zusammengeführt werden. Zweitens werden die entsprechenden Scans ausgewählt (Analysis Tools in Abbildung 19). Schließlich werden die Analyseeinstellungen und Auswertungsgrenzen in drei Schritten festgelegt (Abbildung 20).

**Cartridge Inhaltsverzeichnis, Protokolle und Kommentare**

**New Submap (capture)**

	1	2	3	4	5	6	7	8	9	10	<b>Histidine</b> (5,1) (5,2) (6,1)
1	●	●	●	●	●	●	●	●	●	●	<b>sample</b> (1,1) (1,2) (1,3) (1,4) (1,5) (1,6) (1,7) (1,8) (1,9) (1,10) (2,1) (2,2) (2,3) (2,4) (2,5) (2,6) (2,7) (2,8) (2,9) (2,10)
2	●	●	●	●	●	●	●	●	●	●	
3	○	○	○	○	○	○	○	○	○	○	
4	○	○	○	○	○	○	○	○	○	○	
5	●	○	○	○	○	○	○	○	○	○	
6	●	○	○	○	○	○	○	○	○	○	
7	○	○	○	○	○	○	○	○	○	○	
8	○	○	○	○	○	○	○	○	○	○	
9	○	○	○	○	○	○	○	○	○	○	
10	○	○	○	○	○	○	○	○	○	○	

**Fluorescence Data Chooser**

- naflexecute/Kal Factor V 256.naf
  - Scan Low 256
    - \*fluorescence (1 frames)
      - Frame 1
  - Scan Low 1024
    - \*fluorescence (1 frames)
      - Frame 1

<b>Waveform:</b> Auto	<b>Source:</b> target	<b>Amplitude:</b> 2.000, 2.007 V
<b>Bias:</b> 0.00	<b>+ pad count:</b> 2	<b>- pad count:</b> 0
<b>Mean:</b> 467.77 nA/pad	<b>High:</b> 480.46 nA/pad	<b>Low:</b> 456.19 nA/pad
<b>Mean:</b> 935.54 nA	<b>High:</b> 960.92 nA	<b>Low:</b> 912.39 nA

Abbildung 19: Automatische Analyse mittels nLAB™ NanoChip® Data Viewer and Analysis Software. Die Maske ermöglicht alle Protokolle und Kommentare eines Cartridges (schwarzer Pfeil) und die Aktivierungspotentiale während des Loaderprotokolls (blauer Pfeil) anzusehen. Vor einer Analyse kann der Mikrochip konfiguriert werden, wobei die Komponenten der Pads bestimmt werden (NanoChip™ Config). Unter Analysis Tools müssen die Scans ausgewählt werden.

Im ersten Schritt werden die Bezeichnungen der roten und grünen Fluoreszenzsignale bestimmt, gewöhnlich wurden die Bezeichnungen M für Mutante und W für Wildtyp gewählt (Part 1 in Abbildung 20). An dieser Stelle besteht aber auch die Möglichkeit bei einer Capture down-Analyse, wo die Bezeichnungen für jeden SNP einzeln bestimmt werden, die entsprechenden Basen für Mutante und Wildtyp einzutragen. Bei diesem Schritt kann auch eine Heterozygotenkontrolle festgelegt werden. Weiterhin werden für die Entscheidungsfindung Grenzwerte der genotypbestimmenden Verhältnisse zwischen Grün und Rot definiert (Standardeinstellung: Heterozygote < 1:2 < Unbestimmt < 1:5 Homozygote, Abbildung 20), die minimale Signal-to-noise ratio for Pad Exclusion or No Designation zwischen Probe und Background

(Standardeinstellung: 5:1, d.h. nur das Signal, das 5mal höher ist als der Hintergrund wird gewertet) und die minimale Fluoreszenzstärke zur Berücksichtigung eines Pads vorgegeben (Standardeinstellung: >50RFU).

**PART 1: Define Probe Designations** Define *PROBES* for each *LOCUS*. Select *HETEROZYGOTE CONTROL* for normalizing Red/Green response.

Locus	Probe 1	Probe 2	Heterozygous Normalizer
Kd12	M	W	▼
Kd15			▼

Set Probe Designation Thresholds:  
 Heterozygote <  $\sqrt{1:2}$  < No Designation <  $\sqrt{1:5}$  < Homozygote  
 Signal-to-noise ratio for Pad Exclusion or No Designation:  $\sqrt{5:1}$   
 Minimum signal for Pad Exclusion or No Designation:  $\sqrt{50}$

**PART 2: Background Subtraction** *BACKGROUND PADS* are highlighted in blue. Click on a *LOCUS*. Select background pads for that *LOCUS*. Click *SET PADS* button to save selection. Repeat for each *LOCUS*. **NOTE:** All background pads are selected by default.

**Loci:** sample

*Pads with Histidine used for background.*

	1	2	3	4	5	6	7	8	9	10
1	●	●	●	●	●	●	●	●	●	●
2	●	●	●	●	●	●	●	●	●	●
3	●	●	●	●	●	●	●	●	●	●
4	●	●	●	●	●	●	●	●	●	●
5	●	●	●	●	●	●	●	●	●	●
6	●	●	●	●	●	●	●	●	●	●
7	●	●	●	●	●	●	●	●	●	●
8	●	●	●	●	●	●	●	●	●	●
9	●	●	●	●	●	●	●	●	●	●
10	●	●	●	●	●	●	●	●	●	●

**PART 3: Data Selection** Select the *FRAME* of data you wish to analyze. Use a *FILTER* to concentrate on specific *LOCI* or *SAMPLES*. Click *GENERATE SUMMARY* button.

Graph Height:  Graph Width:   Rotate graph

Graph Title:  (optional)

Frame:

Show warnings

Plot sample sample

Sample A ▼ for these loci:

Plot locus sample ▼

for these samples:

Sample A  
 Sample B  
 Sample C  
 sample D  
 Sample E

Abbildung 20: Analyseeinstellungen in drei Schritten (Part 1-3). Schritt 1 die Bezeichnungen und Auswertungsgrenzen festlegen, Schritt 2 Background definieren, Schritt 3 Scan und Plot Darstellungen wählen.

In einem zweiten Schritt werden die Pads bestimmt, die als Background gelten, um deren rote und grüne RFUs während der automatischen Analyse aus den Signalen der Proben subtrahieren zu können (Part 2 in Abbildung 20). Im dritten Einstellungsschritt wird ein Scan (z.B. Scan Low 256, Part 3 in Abbildung 20) oder alle Scans ausgewählt und je nach Fragestellung ein Plot-Format festgelegt. Dabei können alle Komponenten beliebig kombiniert werden, z.B. bei einer Capture down-Assay können entweder ein SNP für mehrere PCR-Produkte oder mehrere SNPs für ein PCR-Produkt dargestellt werden.

Nachdem die drei Schritte der Einstellungen festgelegt sind, erscheinen zeitgleich die Ergebnisse der Analyse. Je nach Vorgabe können die Ergebnisse in Trendgrafiken (Abbildung 21) oder in einem Histogramm mit Rohdatentabelle (Abbildung 22) abgelesen werden.

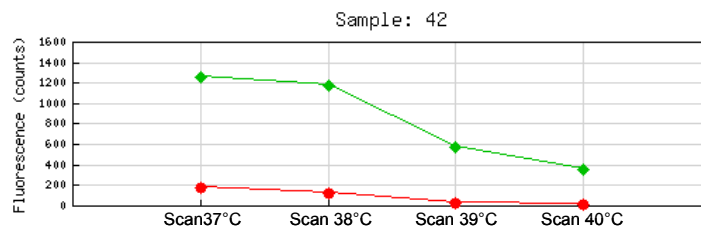


Abbildung 21: Typische Trendgrafik für einen homozygoten Wildtyp (grün). Die Y-Achse bestimmt die RFU Intensitäten, auf der X-Achse sind vier Scans bei unterschiedlichen Diskriminierungstemperaturen gelistet.

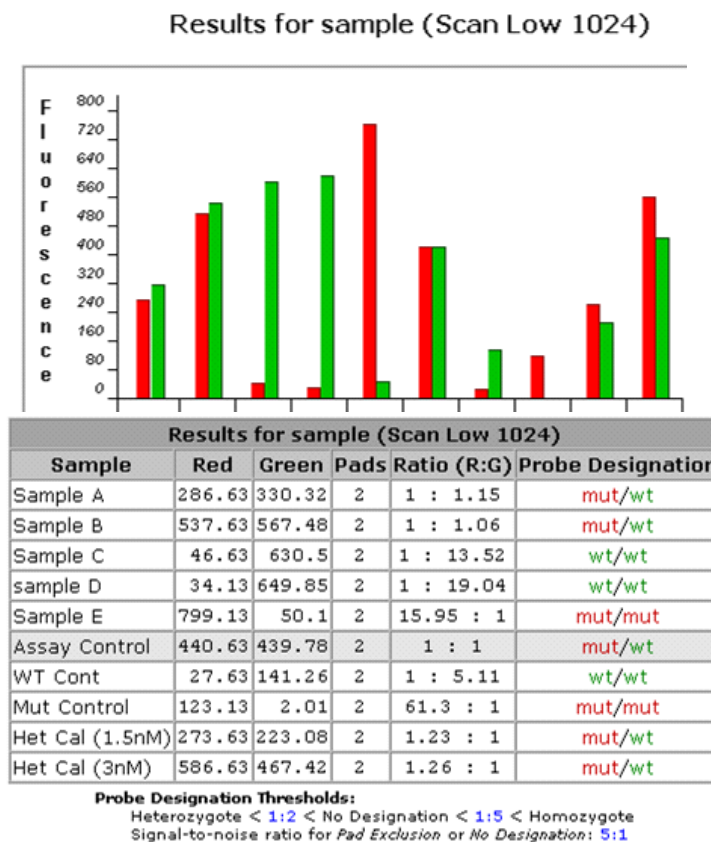


Abbildung 22: Typisches Histogramm und die dazu gehörende Rohdatentabelle. Einem Signalpaar aus Rot und Grün entspricht eine PCR-Probe (beim Amplicon down-Assay) oder SNP (beim Capture down-Assay).

Eine Trendgrafik entsteht dann, wenn alle Scans zur Analyse ausgewählt werden, damit sind mehrere Scans Voraussetzung für eine solche Darstellung (z.B. Scans bei verschiedenen Diskriminierungstemperaturen in Abbildung 21). Die Darstellung von nur einem Scan zeigt die Abbildung 22. Die Deutung eines Histogramms ist vom Assay abhängig. So stellt das Histogramm in Abbildung 22 im Falle eines Amplicon down-Assays die Ergebnisse eines SNPs für mehrere PCR-Produkte (X-Achse) dar. Während bei einem Capture down-Assay das Histogramm mehrere SNPs (X-Achse) für nur ein PCR-Produkt abbildet. Die Rohdatentabelle führt die Signalintensitäten für Rot und Grün, die Anzahl der gescannten Pads, das Verhältnis zwischen Rot und Grün und dessen Bedeutung auf.

Die automatische Auswertung erlaubt hohe Flexibilität bezüglich Einstellung und Analyse, jedoch lässt sie die Kombination der Rohdaten verschiedener Cartridges nicht zu und kann die genotypbestimmenden Verhältnisse zwischen den Fluoreszenzfarben nicht für jeden SNP einzeln festlegen. Eine Individualfestlegung kann jedoch die Auswertung zuverlässiger machen und eine Art Normalisierung für jeden SNP einer Multiplex-Analyse bedeuten. Um dieses zu erreichen, wurden die Intensitäten der Signale als Rohdaten exportiert und mit eigens programmierten Excelmakro ausgewertet.

Zunächst wurden die Ergebnisse der verschiedenen Assayetablierungen und die darauf folgende Capture down-Assayoptimierung für die Multiplex-Genotypisierung anhand der automatischen Datenanalyse bewertet. Die Ergebnisse der 40 Proben zur Validierung des Capture down-Assays wurden aber zusätzlich mit Hilfe von Excel ausgewertet, um ein besseres Verständnis und Transparenz über die einzelnen SNPs zu erlangen.

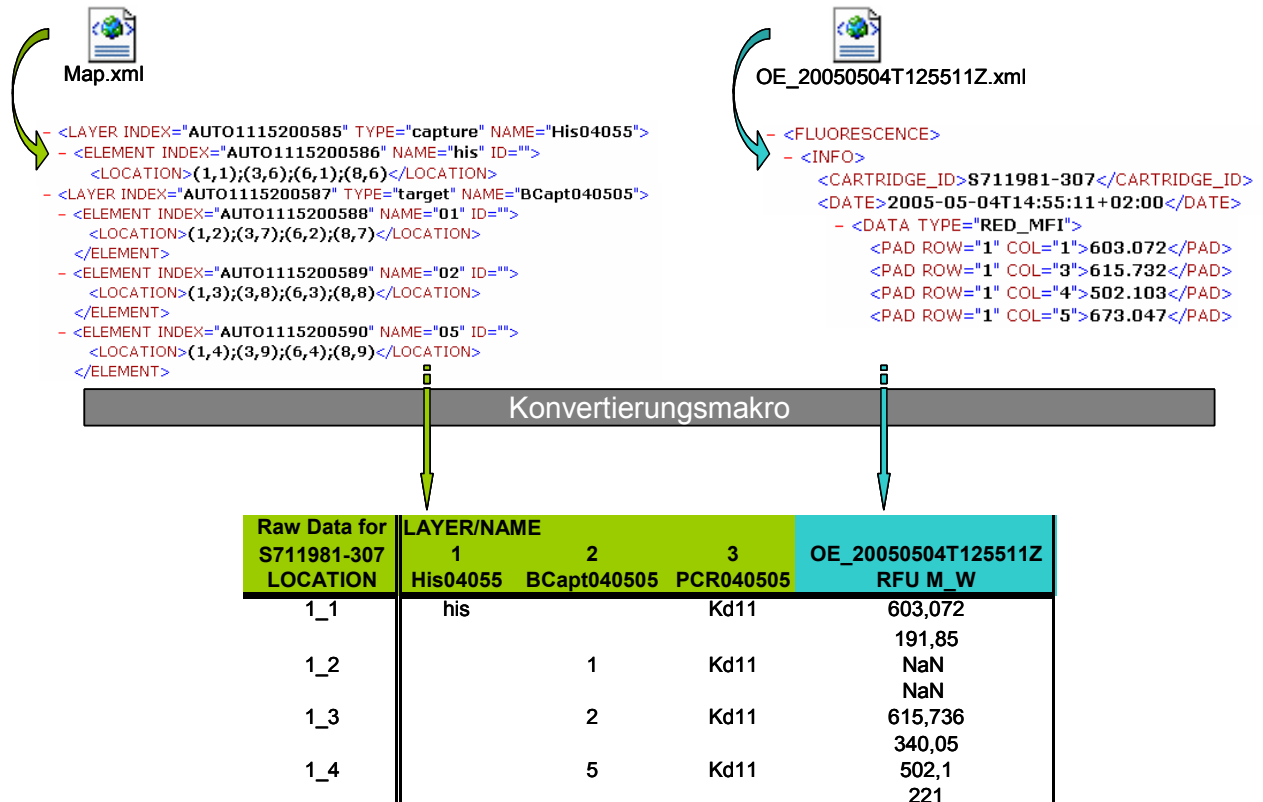


Abbildung 23: Der Konvertierungsmakro fügt alle Informationen in eine Tabelle zusammen und sortiert die Daten nach Pads.

Jeder einzelne automatische Schritt wird vom Loader oder Reader ausgelöst und alle Rohdaten sind pro Cartridge sortiert und in einer sog. HISTORY-Datei gespeichert. Darin befinden sich XML-Dokumente, Textdokumente, JPG- und DAT-Dateien, die den Gesamtprozess dokumentieren. Um die zur Auswertung benötigten Informationen aus der Datenflut zu filtern bedarf es mehrerer manueller Schritte zum Export, der Umwandlung oder dem Kopieren von Daten. Diese zeitaufwendige Prozedur, die auch als Fehlerquelle angesehen werden muss, konnte durch ein selbstprogrammiertes, komplexes Excelmakro vereinfacht werden. Das sog. Konvertierungsmakro nutzt exportierte Tabellen, die die Angaben des Loaders über die Pad-Anordnung und Adressierungs- und Hybridisierungsanzahl der Capture-Sonden und der Proben enthalten. Diese Daten werden durch das Makro nach Pads sortiert (Abbildung 23). Demnach stehen in der ersten Spalte („LOCATION“) die Pads von 1\_1 bis 10\_10 (d.h. Zeile1\_Spalte1 eines Mikrochips). Es folgen so viele Spalten wie der fortlaufend nummerierte „LAYER“ vorgibt (in Abbildung 23 drei Spalten, His04055, BCapt040505 und PCR040505). Diese ergeben die Informationen pro Pad sukzessive für alle Loader Maps, die jemals für diesen Cartridge definiert und analysiert wurden. Anschließend werden in die Exceltabelle alle Scans pro Pad eingefügt, d.h. jede Spalte (z.B. OE\_20050504T125511z, RFU M\_W) bedeutet einen Scan und die Fluoreszenzintensitäten werden für die Mutante und den Wildtyp untereinander angegeben. Am Namen der Spalte kann der Zeitpunkt des Scans abgelesen werden.

Diese systematische Anordnung der Daten mit Hilfe des Konvertierungsmakros war die wesentliche Voraussetzung für eine statistische Auswertung unter Verwendung der Excel Software. Mit Hilfe der Daten einer größeren Anzahl bekannter Genotypen wurden die Grenzbereiche in Form geometrischer Winkel für alle drei Genotypen (Homozygot Mutante, Heterozygot und Homozygot Wildtyp) eines SNPs bestimmt. Diese visualisierten Informationen konnten dann zur Genotypisierung einer Person mit unbekanntem Genotyp benutzt werden.

Als Ausgangspunkt diente eine Tabelle mit den Ergebnissen der 40 Proben, die nach SNPs sortiert wurde und alle Proben, ihre roten und grünen Signal-Werte und die SOLL-Genotypen nach SNaPshot-Typisierung beinhaltet. Für jeden SNP wurden die Genotyp-Grenzbereiche durch ein Makro aus aufeinander aufbauenden Schritten berechnet und im Koordinatensystem als Winkelkanäle dargestellt.

Als erstes wurden die Winkel, die die Verbindungsstrecken vom Nullpunkt (0/0) zum Messergebnispunkt mit der Abszisse einschließen, für jede Probe (Abbildung 24) mittels folgender Formel ausgerechnet:

$$\alpha_{1..n} = \arctan\left(\frac{W_{1..n}}{M_{1..n}}\right)$$

$\alpha$  = Winkel zwischen Nullpunkt-Messergebnispunkt und Abszisse

$n$  = Anzahl der Proben bzw. der Messergebnispunkte

$W$  = Messwert Wildtyp bzw. grüne Fluoreszenz

$M$  = Messwert Mutante bzw. rote Fluoreszenz

Zweitens wurden die Winkel ( $\alpha_{1..n}$ ) nach den SOLL-Genotypen Homozygot Mutante ( $\alpha_M$ ), Homozygot Wildtyp ( $\alpha_W$ ) oder Heterozygot ( $\alpha_{MW}$ ) sortiert.

Drittens wurde für die sortierten Winkel der jeweilige Median pro Genotyp ausgerechnet. Der Median ist der mittlere Winkel in den Beobachtungen.

Für eine gerade Datenanzahl gilt:

$$\alpha_{Median} = \left[ \frac{\frac{\alpha_{(1..n)}}{2} + \frac{\alpha_{(1..n)}}{2+1}}{2} \right]$$

Während die Formel für eine ungerade Datenanzahl lautet:

$$\alpha_{Median} = \frac{\alpha_{[(1..n)+1]}}{2}$$

Die Entscheidung für eine Medianberechnung statt des Mittelwerts ist darin begründet, dass der Median unempfindlich ist gegenüber sog. Ausreißern, im Gegensatz zum Mittelwert, der die Ergebnisse verfälschen könnte, indem einzelne große Messwertabweichungen den Mittelwert in Richtung Ausreißer verschieben würden. Praktisch würde das bedeuten, dass eine falsch typisierte Probe eine zu hohe Gewichtung bekäme und so die Bestimmung der Genotypen-Grenzbereiche zu stark aufweicht. Demnach wurden im dritten Rechenschritt Winkel-Mediane für die drei Genotypen ( $\alpha_{MedianM}$ ;  $\alpha_{MedianW}$  und  $\alpha_{MedianMW}$ ) ermittelt und in einem Diagramm, mit der X-Achse für die roten Fluoreszenzwerte und der Y-Achse für die grünen Fluoreszenzwerte, als Linien aufgetragen (Abbildung 24).

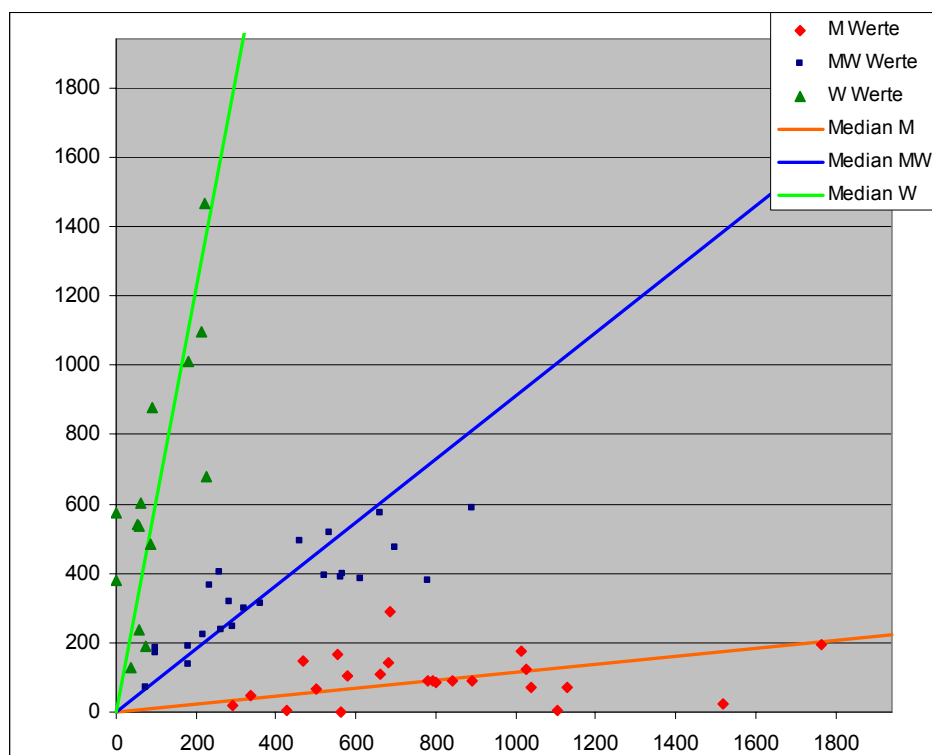


Abbildung 24: Die Messergebnispunkte der Proben sind entsprechend ihren Genotypen dargestellt. Die drei Linien geben den Median für die Genotypen Homozygot-Mutante, Heterozygot und Homozygot-Wildtyp wieder.



In einem vierten Schritt galt es, die Streuungsgrenzen gegenüber den Medianen für alle drei Genotypen zu bestimmen. Dabei wurden erst die Übergangsgrenzen zwischen zwei benachbarten Genotypen, wie Mutante zu Heterozygot oder Wildtyp zu Heterozygot berechnet, und zweitens wurden die unsicheren Grauzonen eines Genotyps zu den benachbarten Genotypen bestimmt. Ziel sollte für jeden Genotyp ein betrachterfreundlich visualisierter Winkelkorridor im Koordinatensystem (Rot/Grün) sein, in dem mit hoher Wahrscheinlichkeit die Soll-Werte zu finden sind.

Um die Übergangsgrenzen zwischen zwei benachbarten Genotypen zu bestimmen, wurden die Standardabweichungen zum Median pro Genotyp ausgerechnet. Die Standardabweichung gibt an, wie stark die Winkel der Messergebnispunkte um den Winkel-Median eines Genotyps schwanken. Die Berechnung der Standardabweichung ( $\sigma$ =Sigma) erfolgte pro Genotyp nach der Formel:

$$\sigma = \sqrt{\frac{(\alpha_1 - \alpha_{Median})^2 + (\alpha_2 - \alpha_{Median})^2 + \dots + (\alpha_n - \alpha_{Median})^2}{n-1}}$$

Anschließend wurden die Verteilungen der Winkel der Messergebnispunkte ermittelt. Es handelt sich um stetige Verteilungen, da die Werte der Winkel alle reellen Zahlen von  $-\infty$  bis  $+\infty$  annehmen können. Der Median wird aus Zufallswerten gebildet und kann daher nach dem zentralen Grenzwertsatz auch als normalverteilt angesehen werden. Aus diesem Grund wurden die Gauß- bzw. Normalverteilungen für alle drei Genotypen nach folgender Formel errechnet:

$$f_N(\alpha) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left[\frac{(\alpha - \alpha_{MedianM})^2}{2\sigma^2}\right]}$$

$f_N(\alpha)$  = Normalfunktion der Winkel, gerechnet pro Genotyp

$\pi$  = 3,14

$\sigma$  = Standardabweichung, gerechnet pro Genotyp

$e$  = Eulersche Zahl

$\alpha_{Median}$  = Median Winkel, gerechnet pro Genotyp

Die ausgerechneten Normalfunktionen der Winkel wurden genutzt, um die Gaußkurven für die drei Genotypen darzustellen, indem auf der X-Achse fiktive Winkel von 1-90 gewählt wurden und auf der Y-Achse die Normalfunktionen für diese Punkte errechnet wurden (Abbildung 25).

Erst jetzt konnten die Übergangsgrenzen zwischen zwei benachbarten Genotypen bestimmt werden. Die Übergangsgrenzen entsprechen den Schnittpunkten der Normalfunktionen, d.h. der Schnittpunkt liegt bei  $f_N(\alpha_M) = f_N(\alpha_{MW})$  und  $f_N(\alpha_W) = f_N(\alpha_{MW})$  (Abbildung 25).

Dies bedeutet gleichwohl, dass die Übergangsgrenze, die im Schnittpunkt zwischen zwei benachbarten Genotypen liegt, den maximalen Winkel des Messergebnispunktes für einen Genotyp bestimmt.

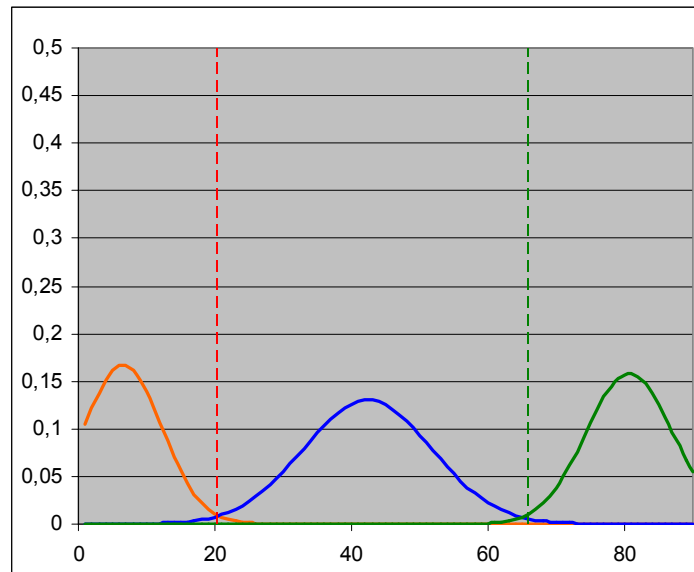


Abbildung 25: Normalverteilung der drei Genotypen (rote Linie = Mutante, blaue Linie = Heterozygot, grüne Linie = Wildtyp). Die gestrichelten Linien geben die Schnittpunkte zwischen zwei benachbarten Genotypen wieder.

Als nächstes wurden die sog. Grauzonen der drei Genotypen ermittelt. Die Grauzonen entsprechen den Konfidenzintervallen (Vertrauensbereichen). Dieser Schritt ist von Bedeutung, denn die einzelnen Ergebnisse weichen immer vom Median ab und über die Konfidenzintervalle können aus den Experimenten Bereiche abgeschätzt werden, die den Median mit einer vorgegebenen Wahrscheinlichkeit enthalten.

Die Konfidenzintervalle werden korridorartig durch die Konfidenzbänder links und rechts des Medians jedes Genotyps definiert. Dabei kann das minimalste bzw. maximalste Konfidenzband auf den Schnittpunkten zwischen zwei benachbarten Gaußkurven liegen. Die Winkel der Konfidenzbänder wurden mittels folgender Formel ausgerechnet:

$$\alpha_{\text{Konfidenzband}} = \frac{t^* \sigma}{\sqrt{n}}$$

$t$  = Quantil der t-Verteilung

$\sigma$  = Standardabweichung gerechnet pro Genotyp

$n$  = Anzahl Proben bzw. Messergebnispunkte

Dabei wird  $t$  aus einer Statistiktabelle entnommen und ist abhängig von der Anzahl der Proben minus 1 und von der Irrtumswahrscheinlichkeit, welche manuell und individuell eingegeben werden kann.

Mit Hilfe dieser Informationen werden im nächsten Schritt die Bereiche der einzelnen Genotypen bestimmt. Aus den Konfidenzintervallen ergeben sich die grauzonenbestimmenden Winkel und damit die minimalen Winkel für Wildtyp, maximalen und minimalen Winkel für Heterozygote und maximalen Winkel für Mutante. Diese können als die Grauzonen begrenzenden Linien in das Diagramm eingezeichnet werden. Alle Messergebnispunkte, die in die Grauzone fallen und damit außerhalb der drei visualisierten Bestimmungskorridore liegen, werden keinem Genotyp zugeordnet (Abbildung 26).

Der Vorteil der Konfidenzintervalle ist, dass die Konfidenzbänder ober- und unterhalb des Medians nicht die gleiche Breite haben müssen und damit auch die Grauzonen

bestimmenden Winkel links und rechts des Gaußschnitts nicht identisch sein müssen (Abbildung 26). Daraus ergibt sich eine zuverlässige Abgrenzung der Genotypen.

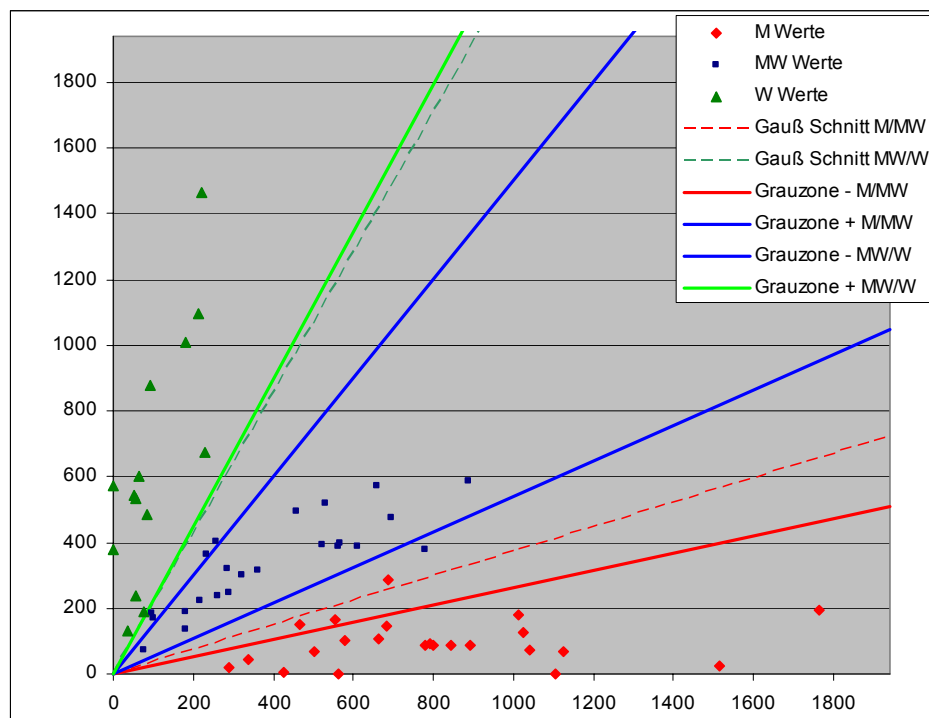


Abbildung 26: Die Linien definieren die Genotypkorridore bzw. die Grauzonen zwischen zwei Genotypen.

Mit Hilfe dieses sog. Berechnungsmakros, das die oben genannten Schritte sukzessiv mit hoher Geschwindigkeit durchführt, wurden die Grenzbestimmungen der drei Genotypen für alle SNPs der validierten Multiplex ermittelt.

Nachdem die Grenzbestimmungen für die drei Genotypen individuell für jeden SNP abgeschlossen waren, konnten diese Informationen genutzt werden, um unbekannte Messwerte einem Genotyp zuzuordnen. Die Genotypidentifizierung unbekannter Messwerte einer Probe für multiple SNPs erfolgte nach einem neuen Makro, dem Genotypisierungsmakro. Hierfür werden die Winkel von 0 zum Messergebnispunkt für jeden SNP einer Probe nach der gleichen Formel wie oben ausgerechnet:

$$\alpha = \arctan\left(\frac{W}{M}\right)$$

Anschließend wird die Zugehörigkeit des Winkels in einen Genotypbereich, dem sog. Bestimmungskorridor, gesucht und damit der Genotyp bestimmt. Gleichzeitig werden die genotypbestimmenden Allele oder „Grauzonen“ für jeden SNP angegeben.

Die drei Excel-basierten Makros, das Konvertierungsmakro, das Berechnungsmakro und das Genotypisierungsmakro eignen sich für die Verwendung weiterer Bestimmungen und sind flexibel gegenüber Datensatzveränderungen.

## 3 Ergebnisse

Gemäß der in Kapitel 1.3 vorgestellten SNPforID-Projektbeschreibung werden die Ergebnisse in folgende Themenkreise unterteilt und beschrieben: Die Y-chromosomale Validierungsstudie (Kapitel 3.1), die forensische Validierungsstudie (Kapitel 3.2) und die elektronische Microarray-Technologie (Kapitel 3.3). Während die Y-chromosomale und die forensische Validierungsstudie als Teilgebiete des SNPforID-Gesamtprojektes behandelt werden, basieren die Erkenntnisse über die Microarray-Technologie auf Resultaten, die ausschließlich im Rahmen dieser Arbeit generiert wurden.

### 3.1 Ergebnisse der Y-chromosomalen Validierungsstudie

Für die Y-chromosomale Validierungsstudie wurden zunächst 29 Y-chromosomale SNPs ausgewählt, die in ihrer Zusammensetzung die weltweiten Haupthaplogruppen bestimmen. Anschließend wurde für 1126 Probanden aus 12 verschiedenen Populationen mittels SNaPshot™-Kit die jeweilige Haplogruppe ermittelt. Die zentrale Koordination der Studie erfolgte durch die Arbeitsgruppe in Santiago de Compostela, wo alle Ergebnisse zusammenliefen und ausgewertet wurden (Brion et al. 2005). Als eines der drei beteiligten Labore lieferte Mainz im Rahmen dieser Arbeit die Ergebnisse zu den Populationen aus Deutschland, China, Thailand und Japan.

Aus diesem Grund werden die Ergebnisse dieser Populationen unter dem Aspekt eigener methodischer Erkenntnisse und populationsspezifischer Betrachtungen im Folgenden ausführlich beschrieben. Ergänzend sind die Gesamtdaten unter Berücksichtigung der Ergebnisse aller beteiligten Laboratorien mit detaillierten Angaben zu Frequenzen der einzelnen Haplogruppen, Haplogruppendiversitäten (HgD, Haplogroup diversity) innerhalb einer Population und interkontinentalen Distanzen im Anhang (Kapitel 7.3) dieser Arbeit als Originalveröffentlichung nachzulesen.

#### 3.1.1 Methodische Ergebnisse

Sowohl die Standardprotokolle für die PCR-Amplifikation als auch die enzymatische Aufreinigung mittels Exo I/SAP erfolgten routinemäßig in größeren Ansätzen mit sicherer Zuverlässigkeit, wie die regelmäßig angefertigten Gelbilder zeigen. Das Agarosegel in Abbildung 27a zeigt beispielhaft die Ergebnisse von 29plex-PCR-Amplifikationen für fünf Proben. Die Banden befinden sich entsprechend der PCR-Produktlängen (79-186bp) von unterhalb der 100bp-Leiterbande bis knapp unterhalb der 200bp-Bande. In Abbildung 27b sind die gleichen fünf Proben nach einer enzymatischen Aufreinigung dargestellt. Für die H<sub>2</sub>O-Kontrolle zeigen beide Gelbilder keine Produktbanden, damit bestätigen sie die

kontaminationsfreien Amplifikationen. Gleichmaßen konnten die Minisequenzierungsreaktionen mittels SNaPshot™-Kit nach Standardprotokoll in großen Ansätzen durchgeführt werden. Insgesamt konnten für 285 männliche Personen die Haplogruppen relativ zeitnah bestimmt werden (Tabellen 39, 40, 41 und 42, Kapitel 7.2).

Die Homogenität der Proben bei der Kapillarelektrophorese bestätigt die einwandfreie Validierung der Protokolle ebenso wie die gelungene Austarierung der Primerkonzentrationen (Abbildung 28). Die meisten Proben konnten mit großer Zuverlässigkeit automatisch analysiert werden, nur ganz selten war eine Editierung der Ergebnisse notwendig.

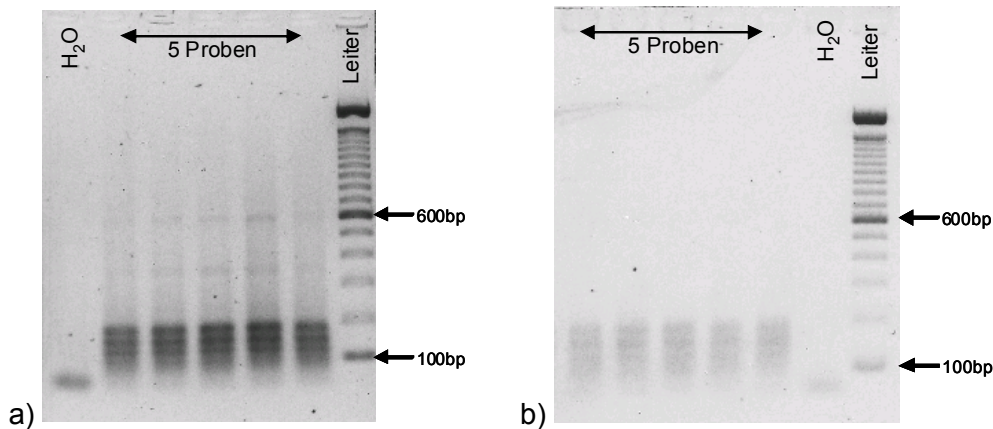


Abbildung 27: Agarosegelelektrophorese von fünf Proben und eine H<sub>2</sub>O-Negativkontrolle a) nach Multiplex-PCR-Amplifikation von 29 SNPs und b) nach anschließender enzymatischer Aufreinigung.

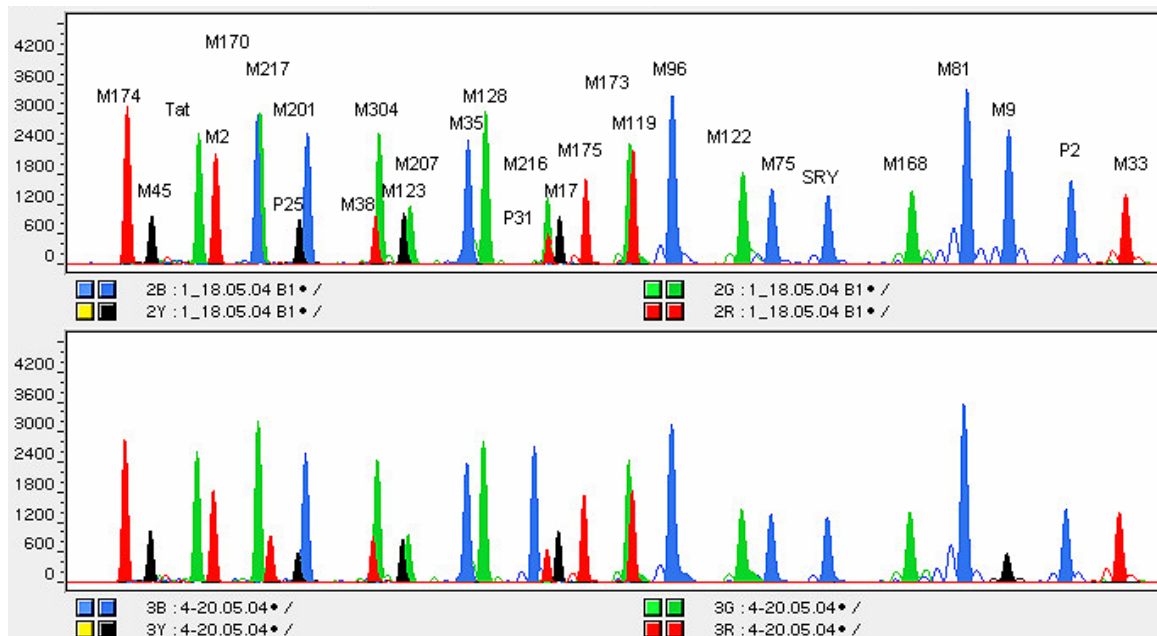


Abbildung 28: Zwei typische Elektropherogramme für die Genotypisierung von 29 Y-chromosomalen SNPs. Die Namen der SNPs sind über den entsprechenden Signalen aufgeführt.

Bei der Zuordnung der Signale zu einem Marker war es möglich, sich an den theoretischen Längen der Analyseprimer in der SNaPshot™ Reaktion zu orientieren (Tabelle 11). Obwohl ein gewisser Unterschied zwischen den erwarteten und den beobachteten Längen festzustellen war, hat sich das Designkriterium von vier Basen Abstand zwischen zwei SNPs mit gleichen Substitutionsmöglichkeiten bewährt (Kapitel 2.2.5.1). Die Abweichungen zwischen erwarteten und beobachteten Längen nehmen tendenziell mit der Größe der Analyseprimer ab. Darüber hinaus hängt die Größe dieser Abweichungen von der eingebauten Base ab, d.h. die durchschnittliche Abweichung der A- und T-Allele liegt bei 2,5nt, während die G- und C-Allele durchschnittlich nur um 0,9nt abweichen. Die allelabhängige Längenabweichung wird auch deutlich durch das Laufverhalten von zwei unterschiedlichen Allelen mit gleichen Analyseprimerlängen, wie das Allel G des SNPs M35 und das Allel A des SNPs M128. Die theoretischen Längen beider SNPs liegen bei 47nt, beobachtet wurden 48nt für G und 49nt für das Allel A (Abbildung 28). Die Allelabhängigkeit bei der Abweichung wird auch bei einem Doppelpeak des SNPs M174 deutlich, d.h. wenn beide Allele vorhanden sind, liegt das C-Allel immer 1nt vor dem A-Allel (Abbildung 29). Damit zeigt sich, dass die Längenabweichungen nicht nur eine Folge der Komposition oder Länge der Analyseprimer sind, sondern erheblich von dem eingebauten ddNTP und dessen Fluoreszenzmarkierung abhängen.

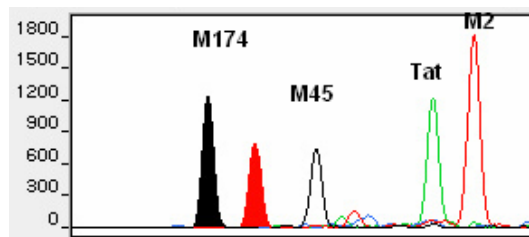


Abbildung 29: Doppelpeak C und T für SNP M174. Die theoretische Länge des Analyseprimers für M174 ist 18nt, das C-Allel liegt bei 20nt, T-Allel bei 22nt. Die Analyseprimer für die SNPs Tat und M2 besitzen eine Länge von 27nt, dennoch liegt das A-Allel 1nt vor dem T-Allel im Elektropherogramm.

Die Bestätigung der Ergebnisse durch Experimentwiederholung, Sichtung des Elektropherogramms oder manueller Korrektur war nur bei einzelnen Fällen notwendig, so zum Beispiel beim Auftreten von SNP spezifischen Doppelpeaks oder bei DNA-Proben von kritischer Qualität. Insgesamt wurden fünf PCR-Amplifikationen von 95 Proben aus China und Thailand wiederholt, um sichere Haplogruppen für alle 29 SNPs zu bekommen (Tabellen 39 und 40, Kapitel 7.2). Keine der 150 Proben aus Deutschland wurde reamplifiziert, wobei für drei Proben ein, bzw. zwei SNPs ausgefallen sind, was aber keine Auswirkung auf die Haplogruppenbestimmung hatte. So ist zum Beispiel bei Probe G16 der Locus M9 ausgefallen, es existiert aber eine Mutation an Locus P25 und daraus ergibt sich eine Haplogruppe R1b, welche auch bestätigt wird durch Mutationen an den Loci M173, M207 und M45, die auf dem phylogenetischen Baum höher angeordnet sind (Tabelle 42, Kapitel 7.2). Im Gegensatz zu den qualitativ guten Proben aus Deutschland, China und Thailand mussten 23 von 40 Experimenten an Proben aus Japan wiederholt werden (Kapitel 3.1.2), um eine zuverlässige Haplogruppenzuordnung vornehmen zu können (Tabelle 41, Kapitel

7.2). Die Präsenz eines Doppelpeaks nach einer automatischen Auswertung wurde stets durch eine Sichtung des Elektropherogramms verifiziert. So zeigte sich für den Locus P25 ausschließlich bei 64 Proben aus Deutschland ein Doppelpeak. Dabei war neben dem mutanten A-Allel auch immer das C-Allel vorhanden (Tabelle 42, Kapitel 7.2). Ebenfalls wurden zwei Peaks für den SNP M174 bei einer chinesischen und 12 japanischen Proben beobachtet. Hier zeigte sich das mutante C-Allel immer nur in Begleitung des Stammallels T (Abbildung 29).

### 3.1.2 Ergebnisse der WGA-Proben

Aufgrund quantitativer Mängel wurden die Proben aus Japan mit Hilfe der Whole Genome Amplification vervielfältigt, verdünnt und anschließend mittels Standardprotokoll amplifiziert (Kapitel 2.2.2). Die WGA-Reaktionen für alle 40 Proben lieferten erfolgreiche Ergebnisse. Nach einer WGA von 1ng/μl DNA lag die Probe hochmolekular und in einer Konzentration von 237-474ng/μl vor (Abbildung 30).

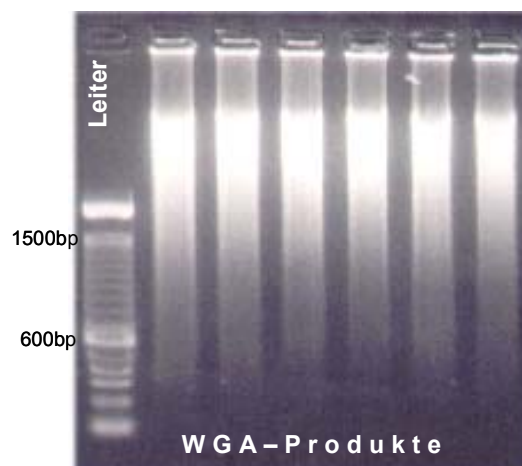


Abbildung 30: Gelelektrophoretische Auftrennung von sechs Proben nach WGA.

Damit konnte die PCR mit 20ng/μl WGA-DNA nach Standardprotokoll angesetzt werden. Dennoch mussten über 50% der Experimente mit Einsatz von originaler DNA wiederholt werden. War die WGA-DNA nur unvollständig und unzuverlässig genotypisiert, so reichte in den meisten Fällen eine Experimentwiederholung mit der gleichen DNA nicht aus. Aus diesem Grund erfolgte bei 23 Proben die dritte PCR-Amplifikation unter Einsatz originaler DNA ohne vorherige WGA-Vervielfältigung.

Prinzipiell lassen sich die Probleme, die durch die Verwendung von WGA-DNA aufgetreten sind, in drei Gruppen einteilen. Erstens ergaben die WGA-Proben bei der Detektion generell für alle 29 Loci erheblich niedrigere Fluoreszenzsignale als die nicht amplifizierten (Abbildung 31).



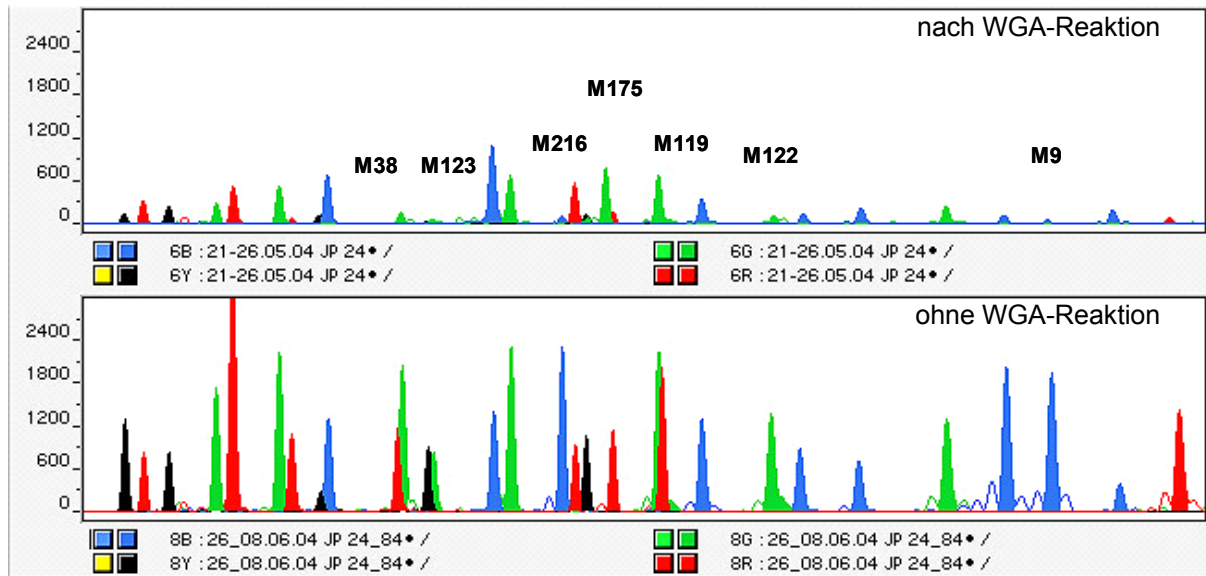


Abbildung 31: Niedrige Fluoreszenzsignale bei der Minisequenzierung der Probe JP24 nach WGA im Vergleich ohne WGA-Reaktion. Ausgefallene SNPs sind mit ihren Namen gekennzeichnet.

Darüber hinaus zeigen die WGA-Proben eine sehr starke Heterogenität, d.h. während einige SNPs hohe Signale liefern, sind andere sehr niedrig oder fallen ganz aus (z.B. SNPs M38, M123 und M119 in der Abbildung 31). Zusätzlich ist zu beobachten, dass tendenziell immer die gleichen Loci kritisch sind (z.B. M38, M123, M216, M175, M119, M122 und M9) und ihre Fluoreszenzstärken nicht mit den Signalen der nicht WGA-Probe korrelieren (z.B. das A-Allel des M175 und T-Allel des M119 in Abbildung 32).

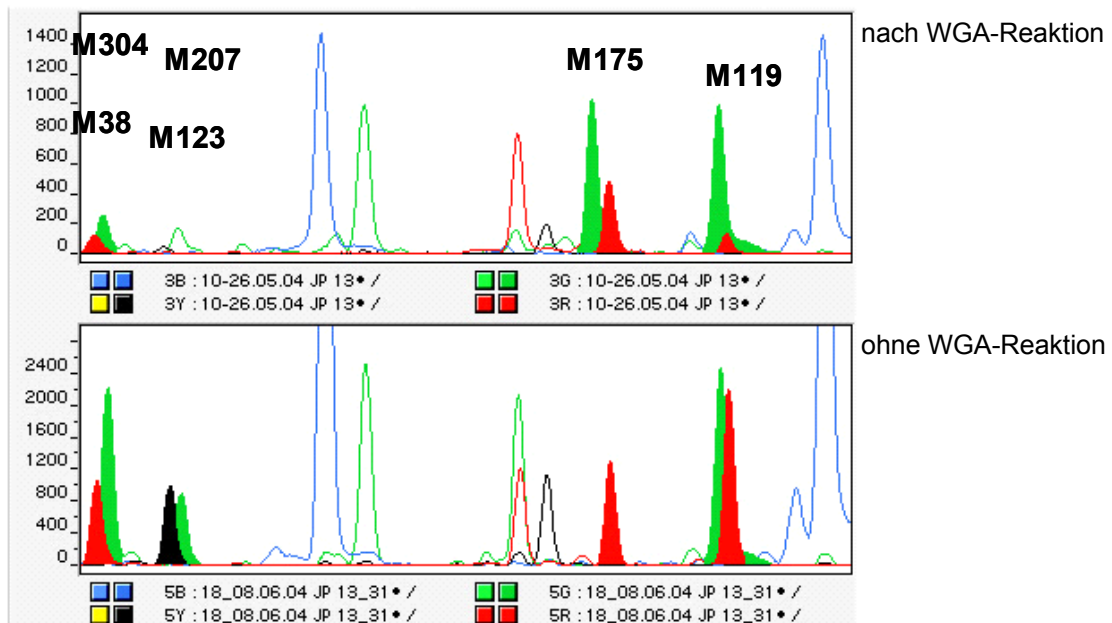


Abbildung 32: Vergleich der Minisequenzierungsergebnisse der Probe JP13 nach WGA-Reaktion und ohne WGA-Reaktion.



Als dritte und schwerwiegendste Problematik ist die Fehldetektion in Form von Doppelpeaks oder Artefakten bei Locus M175 zu deuten. Diese Beobachtung wurde bei 17 ausschließlich nur WGA-Proben gemacht. Niemals wurde ein Doppelpeak für nicht WGA-Proben detektiert (Abbildung 33).

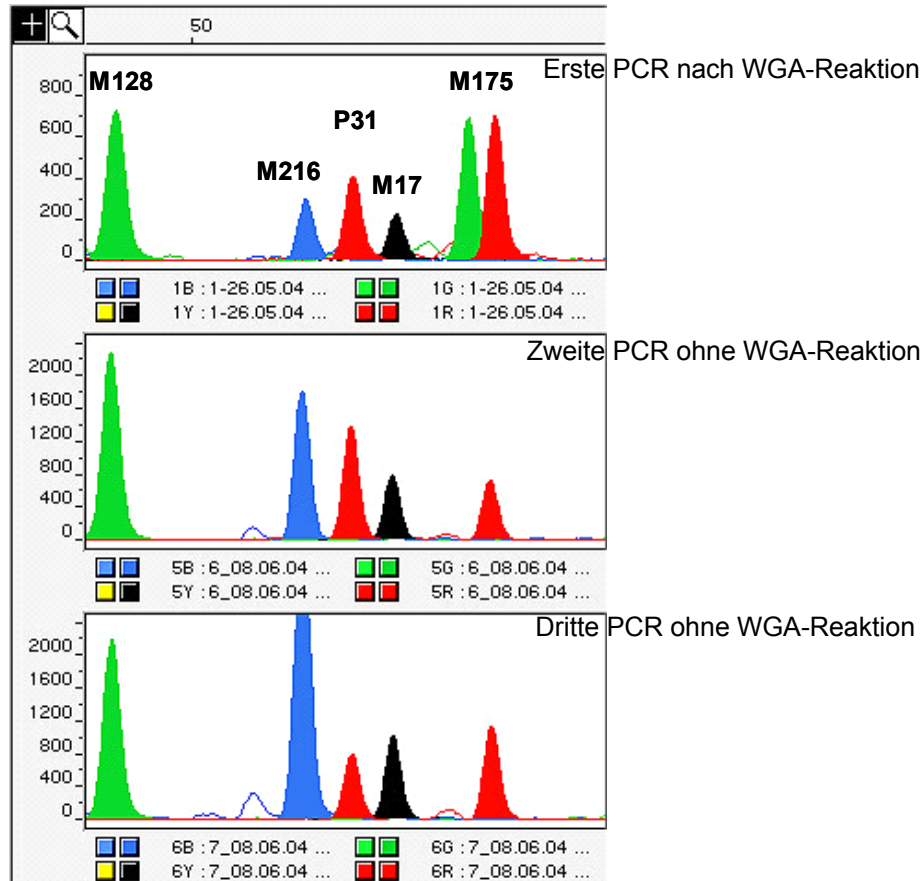


Abbildung 33: Gleichzeitiges Auftreten eines A- und T-Allels für SNP M175 wurde ausschließlich bei der Minisequenzierung der WGA-Probe beobachtet. Beide PCR-Reamplifikationen unter Einsatz von originaler DNA zeigen für die gleiche Probe nur Allel T.

Um bei abweichenden Typisierungen im Falle des SNPs M175 die Frage nach der Richtigkeit zu klären, konnte immer mit Hilfe des phylogenetischen Aufbaus der Haplogruppen bestätigt werden, dass das Experiment mit nicht WGA-DNA auch den richtigen Genotyp ergab. So auch für die Probe JP20, hier wurde mit der Verwendung von WGA-DNA das mutante Allel A detektiert. Im Gegensatz dazu ergab die Minisequenzierung von originaler DNA ein eindeutiges Allel T. Dieses konnte bestätigt werden durch die fehlende Mutation des phylogenetisch übergeordneten SNPs M9 sowie durch die vorhandene Mutation von SNP M217 (Tabelle 41, Kapitel Anhang 7.2).

### 3.1.3 Populationsspezifische Erkenntnisse

Bereits bei der Betrachtung der Allelverteilungen in den vier untersuchten Populationen zeichnet sich ab, dass die 29 Y-chromosomalen SNPs die Haupthaplogruppen charakterisieren. Die Allelfrequenzen in der Tabelle 24 weisen einen deutlichen Unterschied zwischen den Proben aus Deutschland und denen aus dem asiatischen Raum auf. Die Mutationen der Proben aus China, Thailand und Japan sind meist über die gleichen Haplogruppen verteilt, wenn auch mit etwas unterschiedlicher Häufigkeit.

Tabelle 24: Die Verteilung der Allelfrequenzen der 29 SNPs für vier Populationen.

Hg <sup>1)</sup>	SNP	Allel	Allelfrequenzen in %			
			China	Thailand	Japan	Deutschland
BR	SRY	A/G	0/100	0/100	0/100	13,5/86,5
CR	M168	G/A	0/100	0/100	0/100	0/100
C	M216	G/A	94,5/5,5	90/10	92,5/7,5	100/0
C2	M38	T/G	100/0	100/0	100/0	100/0
C3	M217	T/G	94,5/5,5	95/5	92,5/7,5	100/0
D	M174	T/C	98,2/1,8	100/0	70/30	100/0
E	M96	G/C	100/0	100/0	100/0	89,8/10,2
E1	M33	T/G	100/0	100/0	100/0	100/0
E2	M75	G/A	100/0	100/0	100/0	100/0
E3	P2	G/A	100/0	100/0	100/0	87,8/12,2
E3a	M2	T/C	100/0	100/0	100/0	98/2
E3b	M35	G/C	100/0	100/0	100/0	93,1/6,9
E3b2	M81	G/A	100/0	100/0	100/0	100/0
E3b3	M123	C/T	100/0	100/0	100/0	98,7/1,3
G	M201	G/T	100/0	100/0	100/0	98/2
I	M170	A/C	100/0	100/0	100/0	80,7/19,3
J	M304	A/C	100/0	97,5/2,5	100/0	93,9/6,7
K	M9	G/C	7,3/92,7	20/80	42,5/57,5	40,5/59,5
N	M128	A/T	100/0	100/0	100/0	100/0
N3	Tat	A/G	100/0	100/0	100/0	98,7/1,3
O	M175	T/A	12,7/87,3	25/75	42,5/57,5	100/0
O1	M119	T/G	83,6/16,4	90/10	97,5/2,5	100/0
O2	P31	T/C	83,6/16,4	77,5/22,5	65/35	100/0
O3	M122	A/G	45,5/54,5	60/40	82,5/17,5	100/0
P	M45	C/T	100/0	100/0	100/0	42/58
R	M207	A/G	100/0	100/0	100/0	44/56
R1	M173	A/C	100/0	100/0	100/0	44/56
R1a1	M17	C/G	100/0	100/0	100/0	86,7/13,3
R1b	P25	C/A	100/0	100/0	100/0	70,1/29,9

- 1) Haplogruppenzuordnung  
 MAF (Minor Allel Frequency) >20%  
 MAF (Minor Allel Frequency) >0 und <20%

Tabelle 25: Verteilung der Haplogruppen für vier Populationen.

Hg <sup>1)</sup>	Haplogruppenfrequenzen			
	China	Thailand	Japan	Deutschland
CR		7,50%	2,50%	
C		5,00%	2,50%	
C3	5,45%	5,00%	7,50%	
D	1,82%		30,00%	
E3				2,00%
E3a				2,00%
E3b				6,67%
E3b3				1,33%
G				2,00%
I				19,33%
J		2,50%		6,67%
K	5,45%	5,00%	2,50%	0,67%
N3				1,33%
O		2,50%		
O1	16,36%	10,00%	2,50%	
O2	16,36%	22,50%	35,00%	
O3	54,55%	40,00%	17,50%	
P				2,00%
R1a1				13,33%
R1b				42,67%
HgD <sup>2)</sup>	0,64	0,76	0,75	0,75

1) Haplogruppen die durch die vier Populationen vertreten sind.

2) Haplogruppendiversität,  $HgD=1-SUMME(Hg_{1-n}^2)$ .

Die 150 Proben aus Deutschland charakterisieren 12 Haplogruppen, davon ist die Haplogruppe R1b mit 42,7% mit Abstand am häufigsten vertreten und die Haplogruppendiversität liegt bei 75% (Tabelle 25). Die drei Populationen aus Asien repräsentieren 10 Haplogruppen mit etwas unterschiedlicher Verteilung (Tabelle 25). Während die Proben aus China und Thailand meist in der Haplogruppe O3 (Haplogruppenfrequenz 54,6% bzw. 40%) vertreten sind, liegen die Proben aus Japan mit einer Frequenz größer 30% bei Haplogruppe O2 und D. Die Haplogruppenverteilung der Proben aus Thailand ergibt die höchste Diversität von 76%, gefolgt von den Proben aus Japan mit 75%. Die Haplogruppendiversität der Proben aus China liegt nur bei 64%. Die farblich dargestellten Haplogruppen und ihre Verteilungen für die vier Populationen zeigen das Potenzial der 29 SNPs noch deutlicher (Abbildung 34).

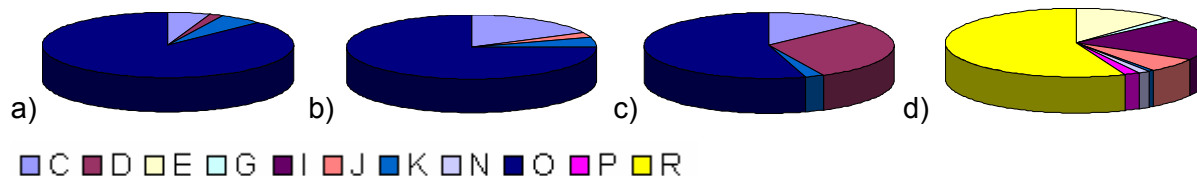


Abbildung 34: Die Kreisdiagramme geben die Haplogruppenverteilungen für die vier Populationen wieder; a) China (n=55), b) Thailand (n=40), c) Japan (n=40) und d) Deutschland (n=150). Die Untergruppen aus der Tabelle 25 wurden in Haupthaplogruppen zusammengefasst aufgeführt, so enthält z.B. die Haplogruppe R die Untergruppen R1a1 und R1b.

## 3.2 Ergebnisse der forensischen Validierungsstudie

Die folgende Ergebnisbeschreibung zur forensischen Validierungsstudie ist unterteilt in zwei Bereiche, als erstes werden in Kapitel 3.2.1 die eigenen Teilergebnisse unter Berücksichtigung des forensisch relevanten Ausgangsmaterials geschildert. Anschließend werden in Kapitel 3.2.2 die Ergebnisse der gesamten Studie, die die Resultate aller beteiligten Laboratorien vereint, beschrieben. Insbesondere wurden die Ergebnisse dabei unter methodischen Aspekten betrachtet und die SNaPshot-Genotypisierungserfolge im Vergleich zu der standardisierten STR-Typisierung dargestellt.

### 3.2.1 Teilergebnisse der forensischen Validierungsstudie

Die Teilergebnisse der Validierungsstudie beruhen auf methodischen Erkenntnissen, die bei der SNaPshot-Genotypisierung von 52 autosomalen SNPs der 20 natürlich degradierten und niedrig konzentrierten Proben, der 14 Verdünnungsproben und der 12 Mischproben gesammelt wurden (Tabelle 4 Kapitel 2.1.3). Die Genotypisierung aller dieser Proben erfolgte dreifach, anschließend wurde für jede Probe, nach den in der Tabelle 26 angegebenen Richtlinien, aus den drei beobachteten Genotypen der Konsensgenotyp bestimmt. Die Anzahl der Ausfälle bzw. Artefakte diente zur Ermittlung des Genotypisierungserfolgs.

Tabelle 26: Richtlinien für die Bestimmung eines Genotyps. Die beobachteten Genotypen aus drei Analysen ergeben den Konsensgenotyp. In Vergleich mit dem Referenzgenotyp kann die Anzahl der Ausfälle und Artefakte bestimmt werden (Musgrave-Brown 2007, modifiziert).

Referenz	Beobachtete Genotypen				Anzahl Ausfälle	Anzahl Artefakte
	Genotyp 1	Genotyp 2	Genotyp 3	Konsens		
C	C	C	C	C	0	0
C	C	C	C/T	C	0	0
C	C/T	C/T	C	C/T	0	1
C	T	C/T	T	T	2	2
C/T	C	C/T	C/T	C/T	0	0
C/T	C	T	N	N	2	0

Zuletzt werden die Ergebnisse der künstlich degradierten Proben beschrieben, wobei größere Akzente gesetzt werden auf die Herstellung dieser Proben und ihre Eignung für die forensische Validierungsstudie, während die Ergebnisse der 52plex SNaPshot-Genotypisierung in Kapitel 3.2.2 behandelt werden.

### 3.2.1.1 Ergebnisse der natürlich degradierten Proben

Die natürlich degradierten Proben, die über Monate der Einwirkung von Mikroorganismen ausgesetzt waren, lagen in sehr geringer Konzentration vor. Die Probenquantifizierung mittels PicoGreen-Fluoreszenzmessung ergab für die meisten Proben eine DNA-Konzentration unter 0,1ng/µl und nur vier Proben lagen über 0,2ng/µl, davon die höchstkonzentrierte Probe Ld10 bei 0,35ng/µl (Tabelle 27).

Tabelle 27: SNP-Genotypisierung der natürlich degradierten Proben, zweigeteilt in die 23plex und die 29plex autosomale Minisequenzierung. Die Ergebnisse sind sortiert nach abnehmender Konzentration der Ausgangs-DNA und nach zunehmender Länge der Minisequenzierungsprodukte. Die Ausfallquote ergibt sich, indem die Anzahl der ausgefallenen SNPs (N) durch die der Experimente geteilt wurde.

		Natürlich degradierte Proben <sup>2)</sup>																				
SBE nt	SNP <sup>1)</sup>	0,35	0,34	0,24	0,21	0,15	0,14	0,14	0,12	0,08	0,06	0,05	0,03	0,02	0,02	0,02	0,02	0,01	0,01	0,01	0	
		Ld10	Ld13	Ld8	Ld7	Ld6	Ld12	Ld2	Ld14	Ld1	Ld9	Ld5	Ld16	Ld11	Ld15	Ld17	Ld20	Ld18	Ld19	Ld4	Ld3	
23plex autosomale Minisequenzierung	18 01	N	N	T,T	C,C	N	N	T,T	N	C,T	N	T,T	T,T	N	T,T	T,T	N	N	T,T	T,T	T,T	
	18 07	N	N	N	N	N	N	N	N	N	N	N	N	AA	N	N	N	N	N	N	N	N
	24 02	T,T	N	N	C,T	N	N	N	N	T,T	N	N	N	N	N	N	N	N	N	N	N	N
	25 13	AA	AA	N	N	N	N	N	N	AA	AA	N	N	N	AA	N	N	N	N	N	N	AA
	29 03	N	N	N	N	N	N	N	N	AA	N	N	N	N	N	N	N	N	N	N	N	N
	29 09	N	N	N	N	N	N	N	N	C,C	N	N	N	N	N	N	N	N	N	N	N	N
	34 10	T,T	T,T	T,T	T,T	T,T	N	N	N	C,T	T,T	N	N	N	N	N	N	N	C,C	N	N	N
	34 22	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	N	N	N	N	N	N	N	N	N	N	N
	38 21	AG	AG	AG	AG	AG	AG	AG	AA	AG	AG	N	AA	AG	AG	AA	N	AG	AA	AG	AA	AA
	42 08	T,T	T,T	N	N	N	N	N	N	T,T	N	N	N	N	N	N	N	N	N	N	N	N
	42 17	AA	N	N	N	N	N	N	N	AA	N	N	N	N	N	N	N	N	N	N	N	N
	46 11	C,C	C,T	C,T	C,T	T,T	C,T	N	C,T	T,T	C,T	N	N	N	N	N	N	N	N	N	N	N
	46 23	AA	AA	AG	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	G,G	AA	AA	AA	AA	AA	AA
	50 20	N	N	N	G,G	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	54 06	N	N	N	N	AA	N	N	N	AA	N	N	N	N	N	N	AA	N	N	N	N	N
	58 12	N	N	N	N	N	N	N	N	G,G	AA	N	N	N	N	N	N	N	N	N	N	N
	58 19	N	N	N	T,T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	62 14	N	A,T	A,T	A,T	A,T	A,T	N	A,T	A,T	A,T	N	N	N	A,T	N	N	N	N	N	N	N
	66 18	T,T	T,T	T,T	T,T	N	T,T	T,T	N	T,T	N	N	N	N	N	N	N	N	N	N	N	N
	70 16	G,T	G,T	G,T	G,T	G,G	G,T	G,G	G,T	G,G	G,T	N	G,G	G,G	N	G,G	N	G,G	G,G	G,G	N	N
74 05	AG	AG	AG	AG	G,G	G,G	N	G,G	AA	AG	N	N	G,G	N	N	N	G,G	N	N	N	N	
74 15	T,T	T,T	T,T	T,T	T,T	T,T	N	T,T	C,T	T,T	N	N	T,T	N	N	N	N	N	N	N	N	
78 04	N	N	N	N	N	N	N	N	T,T	N	N	N	N	N	N	N	N	N	N	N	N	
<b>23plex Ausfall</b>		<b>0,43</b>	<b>0,48</b>	<b>0,48</b>	<b>0,39</b>	<b>0,57</b>	<b>0,61</b>	<b>0,74</b>	<b>0,65</b>	<b>0,13</b>	<b>0,52</b>	<b>0,96</b>	<b>0,83</b>	<b>0,70</b>	<b>0,83</b>	<b>0,83</b>	<b>0,91</b>	<b>0,78</b>	<b>0,83</b>	<b>0,87</b>	<b>0,83</b>	
29plex autosomale Minisequenzierung	16 41	C,C	T,T	T,T	C,C	C,T	T,T	N	C,T	C,T	C,T	T,T	T,T	C,C	C,C	C,C	C,C	C,T	T,T	C,T	T,T	
	16 46	AG	AA	AG	G,G	AG	G,G	AA	AG	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
	22 33	T,T	C,T	T,T	N	C,T	C,T	C,T	C,T	T,T	T,T	T,T	T,T	C,T	C,T	C,C	C,C	C,T	C,T	C,T	C,T	
	23 27	AA	G,G	AA	AG	AA	AG	N	G,G	AA	AA	G,G	AG	AA	AA	AA	G,G	G,G	G,G	G,G	G,G	
	27 36	AA	T,T	T,T	T,T	T,T	AA	T,T	A,T	T,T	N	AA	T,T	AA	N	N	N	N	AA	T,T	T,T	
	27 38	C,C	C,G	C,C	C,C	C,C	C,C	N	G,G	C,C	C,G	N	C,G	G,G	G,G	G,G	C,C	C,G	C,G	C,G	G,G	
	32 44	T,T	N	T,T	C,T	N	N	N	N	C,T	N	N	N	N	N	N	N	N	N	N	N	N
	32 51	AA	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	36 48	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	N	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	N	T,T	T,T
	36 49	AG	AG	AA	AG	AG	AA	AA	AA	AG	AA	AA	AA	N	N	AA	AA	AA	N	AA	AA	AA
	40 34	C,C	C,G	C,G	G,G	C,G	N	N	G,G	C,G	C,G	N	N	C,C	N	N	C,C	C,C	N	C,C	N	N
	40 52	AA	AA	AA	AA	AA	A,T	N	AA	AA	AA	N	N	N	N	N	N	N	N	N	N	N
	44 32	C,C	C,C	C,C	C,C	C,C	C,T	C,C	C,C	C,C	T,T	C,C	N	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C
	44 42	G,G	AA	AA	N	AA	N	AA	AA	AA	AA	G,G	AA	G,G	AA	N	AA	AA	AA	AA	AA	G,G
	48 37	C,C	N	T,T	C,T	T,T	N	N	N	T,T	N	N	N	N	N	N	N	N	N	N	N	N
	48 53	AG	AA	AA	AA	N	N	N	N	AA	N	AA	AA	N	N	N	N	N	N	AA	N	AA
	52 25	G,G	AG	AG	G,G	N	G,G	G,G	G,G	AG	G,G	N	N	N	N	N	N	N	G,G	N	N	N
	56 24	C,C	C,C	C,C	C,C	C,T	C,C	C,T	C,C	C,T	C,T	N	N	N	N	N	N	N	C,T	N	N	N
	56 29	G,G	N	N	AG	AA	AA	N	N	G,T	AA	N	AA	N	AA	N	N	N	N	N	N	AA
	60 54	AA	AA	N	AA	N	N	N	N	AA	N	N	N	N	N	N	N	N	N	N	N	N
64 26	T,T	T,T	T,T	T,T	I,G	T,T	T,T	I,G	T,T	T,T	N	N	N	N	N	N	N	T,T	N	N	N	
68 40	AG	AG	AG	AA	N	N	G,G	G,G	AG	N	N	N	N	N	N	N	G,G	N	N	N	N	
68 45	T,T	T,T	T,T	T,T	N	N	N	N	C,T	N	N	N	N	N	N	N	N	N	N	N	N	
72 30	C,C	C,C	C,C	C,C	AA	AC	N	AC	AA	AA	N	N	AA	N	N	N	AA	N	AA	AA	AA	
76 50	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
80 39	N	T,T	A,T	A,T	N	N	N	N	A,T	N	N	N	N	N	N	N	N	N	N	N	N	
84 35	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	N	N	N	N	N	N	N	AA	N	N	N	
88 43	AG	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
92 28	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	N	N	N	N	N	N	N	G,G	N	N	N	
<b>29plex Ausfall</b>		<b>0,09</b>	<b>0,22</b>	<b>0,17</b>	<b>0,22</b>	<b>0,39</b>	<b>0,52</b>	<b>0,61</b>	<b>0,39</b>	<b>0,13</b>	<b>0,43</b>	<b>0,70</b>	<b>0,70</b>	<b>0,57</b>	<b>0,65</b>	<b>0,70</b>	<b>0,61</b>	<b>0,48</b>	<b>0,70</b>	<b>0,57</b>	<b>0,52</b>	

1) SNPs für 23plex und 29plex sortiert nach der Länge der Analyseprimer.  
 2) Natürlich degradierte Proben mit Angaben sortiert nach ihrer DNA-Konzentration [ng/µl].  
 N Locus Ausfall.

Die DNA-Konzentration der Probe Ld3 lag unterhalb der Bestimmungsgrenze. Diese DNAs stellten damit für die forensische Routine typische Proben von stark degradiertem Zustand und geringer Konzentration dar. Tabelle 27 fasst die Resultate für die 20 natürlich degradierten Proben für beide SNaPshot-Multiplexe zusammen. Gleichzeitig muss beachtet werden, dass an dieser Stelle nur eine methodische Beurteilung möglich ist, weil die zu erwartenden Genotypen für diese Proben nicht bekannt waren. Aus diesem Grund kann an dieser Stelle nur die Präsenz oder der Ausfall eines SNPs beurteilt, nicht aber die Richtigkeit bestimmt werden.

Insgesamt gesehen ist bei der Genotypisierung aller natürlich degradiert Proben eine sehr hohe Ausfallrate zu beobachten, wobei diese bei der 23plex-Minisequenzierung deutlich höher liegt (durchschnittlich 67%) als bei der 29plex-Reaktion (durchschnittlich 47%, Tabelle 27 und Tabelle 28). Tabelle 27, die die Ergebnisse, sortiert nach DNA-Gehalt der Proben und nach Länge der Analyseprimer darstellt, zeigt bei der 29plex-Minisequenzierung eine deutliche Ausfalltendenz in Richtung geringer DNA-Konzentration und zunehmender Länge der Primer. Dies ist bei der 23plex-Minisequenzierung nicht zu beobachten, wo die kurzen Fragmente gleich häufig ausfallen. Eine mögliche Erklärung dafür ist, dass es bei der elektrophoretischen Auftrennung zu einer Linksverschiebung der Peaks kam (Abbildung 35) und dadurch die Auswertung dieser Signale nur durch manuelle Editierung möglich war. Die Verschiebung betraf immer nur den Anfang eines Laufs.

Tabelle 28: Ausfälle der natürlich degradierten Proben für die 23plex- und 29plex-Minisequenzierung und als Mittelwert der beiden für die gesamten 52 SNPs. Die Proben sind nach ihren DNA-Ausgangskonzentrationen sortiert und die Spalte Input führt die DNA-Mengen auf, die für die 52plex-PCR-Amplifikation eingesetzt wurden.

Probe	DNA [ng/µl]	Input [ng]	23plex Ausfälle	29plex Ausfälle	52plex Ausfälle
Ld3	k.A.	k.A.	0,83	0,52	0,67
Ld18	0,01	0,06	0,78	0,48	0,63
Ld19	0,01	0,06	0,83	0,70	0,76
Ld4	0,01	0,06	0,87	0,57	0,72
Ld11	0,02	0,11	0,70	0,57	0,63
Ld15	0,02	0,11	0,83	0,65	0,74
Ld17	0,02	0,11	0,83	0,70	0,76
Ld20	0,02	0,11	0,91	0,61	0,76
Ld16	0,03	0,17	0,83	0,70	0,76
Ld5	0,05	0,29	0,96	0,70	0,83
Ld9	0,06	0,34	0,52	0,43	0,48
Ld1	0,08	2	0,13	0,13	0,13
Ld14	0,12	0,68	0,65	0,39	0,52
Ld12	0,14	0,8	0,61	0,52	0,57
Ld2	0,14	0,8	0,74	0,61	0,67
Ld6	0,15	0,86	0,57	0,39	0,48
Ld7	0,21	1,2	0,39	0,22	0,30
Ld8	0,24	1,37	0,48	0,17	0,33
Ld13	0,34	1,94	0,48	0,22	0,35
Ld10	0,35	2	0,43	0,09	0,26
<b>Durchschnitt</b>			<b>0,67</b>	<b>0,47</b>	<b>0,57</b>

k.A. DNA-Konzentration unterhalb der Detektionsgrenze.

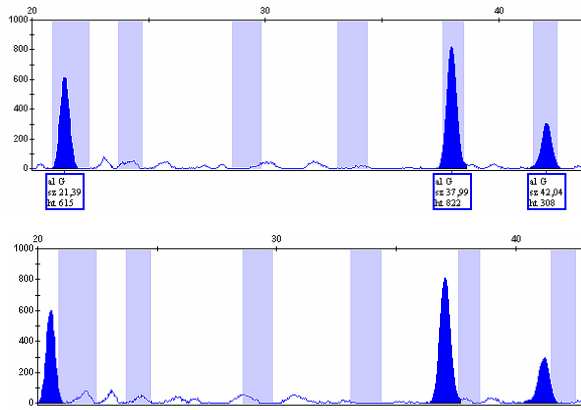


Abbildung 35: Zwei Elektropherogramme gegenübergestellt, das obere zeigt einen automatisch auswertbaren Lauf, während das untere eine typische Linksverschiebung darstellt.

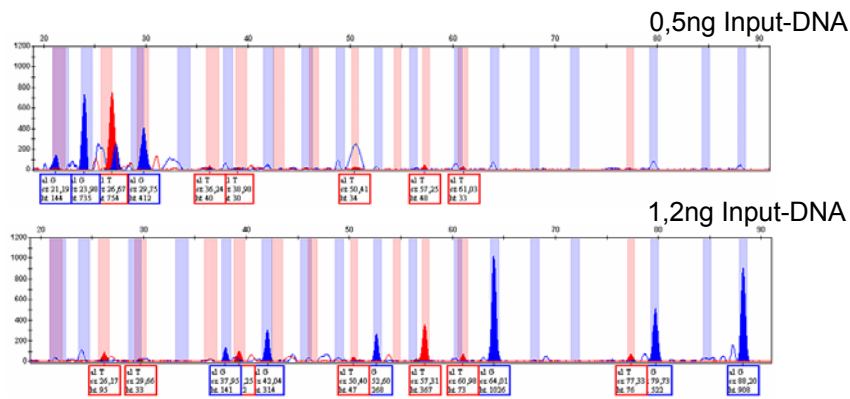


Abbildung 36: Das obere Elektropherogramm zeigt ein typisches Bild für eine natürlich degradierte Probe, beispielhaft sind die G (blau) und T (rot) Substitutionen dargestellt. Unten wurde die gleiche DNA mittels Vakuums konzentriert, sodass 1,2ng DNA in die PCR-Amplifikation eingesetzt werden konnten.

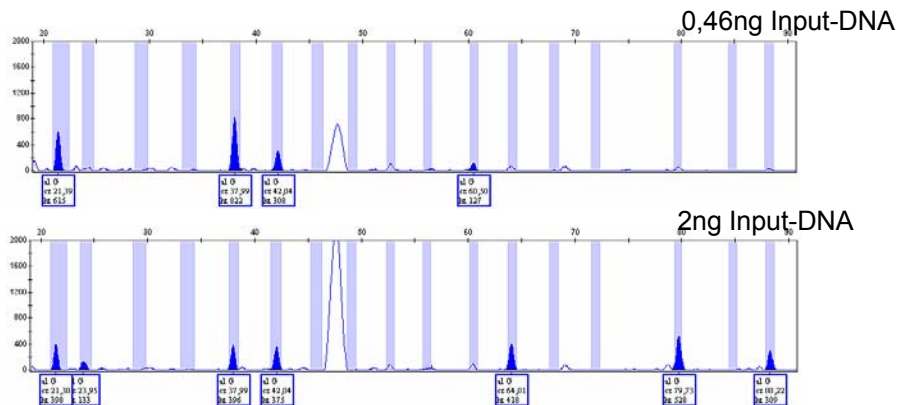


Abbildung 37: Zwei Elektropherogramme der natürlich degradierten Probe Ld1 gegenübergestellt, das obere zeigt die Minisequenzierung von 0,46ng Input-DNA. Für das untere wurde die DNA mittels Vakuums konzentriert, sodass 2ng DNA in die PCR-Amplifikation eingesetzt werden konnten.

Das obere Elektropherogramm in Abbildung 36 zeigt ein typisches Ergebnis für die 29plex-Minisequenzierung, wo die Signale am Anfang agglomerieren und mit wachsendem Analyseprimer rapide abnehmen und dann ganz ausfallen. Das untere Elektropherogramm

stellt das Ergebnis der gleichen Probe dar, mit dem Unterschied, dass die Konzentration der Ausgangs-DNA für diese Analyse mittels Vakuum erhöht wurde (Abbildung 36).

Tabelle 28 liefert Beweise dafür, dass für Proben mit Ausgangskonzentrationen unter 0,2ng/µl die SNP-Ausfälle dramatisch zunehmen, so sind schon 48% der 52 SNPs für die Probe Ld6 mit 0,15ng/µl DNA-Gehalt ausgefallen. Die einzige Ausnahme zeigt die Probe Ld1, welche, obwohl die DNA-Konzentration nur 0,08ng/µl beträgt, trotzdem für beide Minisequenzierungen eine 87%-ige Genotypisierung ermöglichte. Dies lässt sich dadurch erklären, dass die Input-DNA-Konzentration durch Einsatz eines Vakuums erhöht wurde. Abbildung 37 veranschaulicht die Vorteile, die durch Einsatz von DNA-Konzentrat erreicht wurden. Die Input-DNA für ein 25µl PCR-Volumen betrug für diese Probe 2ng.

### **3.2.1.2 Ergebnisse der Verdünnungsproben**

Die Verdünnungsproben basieren auf zwei Standard-DNAs von jeweils 10ng/µl Ausgangskonzentration, die in je sechs Verdünnungen bis in den LCN-Bereich eingesetzt wurden (Kapitel 2.1.3) und dadurch einen Hinweis auf die Sensitivität der SNaPshot-Genotypisierungsmethode geben konnten. Die Genotypen für die zwei 10ng/µl Ausgangsproben (St7 und St8) galten als Referenz und die Konsensgenotypen der Verdünnungsproben wurden nach den oben beschriebenen Richtlinien (Kapitel 3.2.1) ermittelt. Die Ergebnisse der Verdünnungen für die 23plex- und die 29plex-Minisequenzierung sind in Tabelle 29 zusammengefasst. Für jede Probe wurden die Ausfälle und Artefakte getrennt für die 23plex- bzw. 29plex-Minisequenzierung bestimmt und aus ihrem Mittelwert die Ausfälle und Artefakte für die gesamten 52 autosomalen SNPs errechnet (Tabelle 29). Gleichzeitig wurden die Ausfälle und Artefakte auch für jeden einzelnen SNP und alle Proben bestimmt (Tabelle 29).

Aus Tabelle 29 wird deutlich, wie stark der Genotypisierungserfolg mit Verringerung der Konzentration der Ausgangs-DNA abnimmt. So konnten in über 80% der Fälle die Genotypen für die Proben mit 1ng/µl bzw. 0,5ng/µl DNA-Konzentrationen erlangt werden, während die Proben mit 0,2 und 0,1ng/µl DNA nur mit 68-80%-igem Erfolg typisiert wurden. Darunter liegende DNA-Konzentrationen lieferten die Genotypen nur noch in 57%-61% der Fälle. Die Artefakte mit durchschnittlich 4%, verteilten sich relativ ausgewogen über die Ergebnisse aller Verdünnungen jedoch ausschließlich bei der 23plex-Minisequenzierung (Tabelle 29). Die Ergebnisse lassen außerdem erkennen, dass die erfolgreiche SNaPshot-Genotypisierung von LCN-Proben in adäquater Qualität nicht von der Länge der Analyseprimer abhängt. Vielmehr ist das Gelingen SNP-bedingt, da einzelne Loci, wie 20, 51 oder 43, mit großem Abstand häufiger ausfallen (100%, 83% und 92%) als andere (Tabelle 29). Auch eine Wiederholung mit erhöhter Primerkonzentration brachte keine Verbesserung. Darüber hinaus wurden diese Ausfälle auch bei frischen, hochkonzentrierten Speichelproben beobachtet (Abbildung 38).



Tabelle 29: SNP-Genotypisierung der Verdünnungsproben, zweigeteilt in die 23plex und die 29plex autosomale Minisequenzierung. Die Ergebnisse sind sortiert nach zunehmender Länge der Minisequenzierungsprodukte. Die Ausfälle und Artefakte ergeben sich, indem die Ergebnisse mit der Referenzprobe [10ng/µl] verglichen wurden.

SBE nt	SNP <sup>1)</sup>	Verdünnungsproben <sup>2)</sup> (10-0,02ng/µl)						Verdünnungsproben <sup>3)</sup> (10-0,02ng/µl)						Ausfälle	Artefakte		
		10	1	0,5	0,2	0,1	0,05	0,02	10	1	0,5	0,2	0,1			0,05	0,02
		St7	St7_1	St7_2	St7_3	St7_4	St7_5	St7_6	St8	St8_1	St8_2	St8_3	St8_4			St8_5	St8_6
18	01	T,T	T,T	C,T	T,T	T,T	C,T	T,T	C,C	C,C	C,C	C,C	C,C	C,C	C,T	0,00	0,13
18	07	G,G	G,G	N	N	N	N	A,A	A,G	A,A	A,A	A,A	A,A	A,A	N	0,71	0,08
24	02	C,C	C,T	C,C	C,T	T,T	C,C	C,T	C,T	C,T	C,T	T,T	N	T,T		0,25	0,21
25	13	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,G	A,G	A,A	A,A	A,G	A,A	A,A	0,17	0,00
29	03	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	0,00	0,00
29	09	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	N	N		0,17	0,00
34	10	C,C	C,C	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	0,00	0,21
34	22	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	N	N	0,17	0,00
38	21	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,00	0,00
42	08	C,T	C,T	C,T	C,T	T,T	T,T	T,T	C,T	C,T	C,T	T,T	T,T	T,T	T,T	0,29	0,00
42	17	G,G	G,G	G,G	N	N	N	N	A,G	N	N	N	N	N	N	0,83	0,00
46	11	C,T	C,T	C,T	C,T	C,T	C,T	T,T	C,C	C,T	C,T	C,T	C,T	C,T	C,T	0,04	0,25
46	23	A,A	A,A	A,G	A,A	A,A	A,A	A,A	A,G	A,G	A,G	A,A	A,G	A,A	A,A	0,13	0,04
50	20	G,G	N	N	N	N	N	N	G,T	N	N	N	N	N	N	1,00	0,00
54	06	A,A	A,A	A,A	A,A	N	N	N	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,25	0,00
58	12	G,G	G,G	G,G	G,G	G,G	N	N	A,A	A,A	A,A	A,A	N	N	N	0,42	0,00
58	19	C,T	T,T	N	N	N	T,T	T,T	C,T	C,T	C,T	N	N	T,T	T,T	0,63	0,00
62	14	A,A	A,A	N	A,A	A,A	A,A	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	0,08	0,04
66	18	G,T	G,G	N	N	N	N	N	T,T	T,T	T,T	N	N	N	N	0,79	0,00
70	16	G,T	G,T	G,T	G,T	G,G	G,G	G,T	T,T	G,T	G,T	G,T	G,T	G,T	G,T	0,08	0,25
74	05	G,G	G,G	G,G	A,G	G,G	G,G	G,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	0,00	0,04
74	15	T,T	T,T	N	C,T	T,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	0,08	0,13
78	04	A,A	A,A	A,T	A,A	A,T	A,T	A,T	A,A	A,T	A,A	A,A	A,T	A,A	A,T	0,00	0,29
23plex Ausfälle			0,09	0,26	0,22	0,35	0,33	0,33		0,11	0,13	0,26	0,28	0,41	0,41	0,26	
23plex Artefakte			0,02	0,09	0,09	0,09	0,09	0,15		0,07	0,04	0,04	0,07	0,04	0,09		0,07
16	41	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	0,00	0,00
16	46	A,G	A,G	A,G	A,G	A,A	A,A	A,A	A,G	A,G	A,G	A,G	A,G	A,A	A,A	0,21	0,00
22	33	C,T	C,T	C,T	C,C	C,C	C,T	T,T	C,C	C,C	C,C	C,C	C,C	C,C	C,C	0,13	0,00
23	27	A,G	A,G	A,G	A,G	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,13	0,00
27	36	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,00	0,00
27	38	G,G	G,G	N	N	N	N	N	C,G	C,C	C,C	C,C	C,C	N	C,C	0,71	0,00
32	44	C,C	C,C	C,C	C,C	C,C	N	N	C,C	C,C	C,C	C,C	C,C	N	N	0,33	0,00
32	51	A,G	A,G	N	N	N	N	N	G,G	G,G	N	N	N	N	N	0,83	0,00
36	48	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	0,00	0,00
36	49	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,00	0,00
40	34	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	0,00	0,00
40	52	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,00	0,00
44	32	C,T	C,T	C,T	C,T	C,T	T,T	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	0,08	0,00
44	42	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,00	0,00
48	37	C,C	C,C	C,C	C,C	C,C	C,C	N	C,T	C,T	C,T	C,T	C,T	C,C	C,C	0,17	0,00
48	53	G,G	G,G	G,G	G,G	G,G	N	N	A,G	A,G	A,G	A,G	A,G	N	N	0,33	0,00
52	25	A,A	A,A	A,A	A,A	A,A	A,A	N	A,G	A,G	A,G	G,G	A,G	A,G	A,G	0,08	0,00
56	24	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	0,00	0,00
56	29	A,G	A,G	A,G	A,G	A,G	N	N	A,G	A,G	A,G	G,G	N	N	A,A	0,42	0,00
60	54	A,A	A,A	A,A	N	N	N	N	A,A	A,A	N	N	N	N	N	0,75	0,00
64	26	T,T	T,T	T,T	T,T	N	N	N	G,T	G,T	G,T	G,G	N	N	G,G	0,50	0,00
68	40	G,G	G,G	G,G	G,G	G,G	G,G	G,G	A,A	A,A	A,A	A,A	A,A	A,A	N	0,08	0,00
68	45	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C	C,T	0,04	0,00
72	30	C,C	C,C	C,C	C,C	C,C	N	N	A,A	A,A	A,A	A,A	A,A	A,A	A,A	0,17	0,00
76	50	A,G	A,G	N	N	N	N	N	A,A	A,A	A,A	N	N	N	N	0,75	0,00
80	39	A,T	A,T	A,T	A,T	A,T	T,T	T,T	A,T	A,T	A,T	A,T	T,T	T,T	T,T	0,21	0,00
84	35	A,A	A,A	A,A	A,A	A,A	A,A	A,A	G,G	G,G	G,G	G,G	G,G	G,G	N	0,08	0,00
88	43	A,A	A,A	N	N	N	N	N	A,G	N	N	N	N	N	N	0,92	0,00
92	28	A,G	A,G	A,G	A,G	N	N	N	A,A	A,A	A,A	N	N	N	N	0,58	0,00
29plex Ausfälle			0,00	0,14	0,19	0,29	0,45	0,53		0,05	0,12	0,22	0,28	0,41	0,41	0,26	
29plex Artefakte			0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00		0,00
52plex Ausfälle			0,04	0,20	0,20	0,32	0,39	0,43		0,08	0,13	0,24	0,28	0,41	0,41	0,26	
52plex Artefakte			0,01	0,04	0,04	0,04	0,04	0,08		0,03	0,02	0,02	0,03	0,02	0,04		0,04

1) SNPs für 23plex und 29plex sortiert nach der Länge der Analyseprimer.  
 2) Hergestellt aus Standard-DNA #9947a, St7 [10ng/µl] ist die Referenz, St7\_1 bis St7\_6 sind die Verdünnungen [1-0,02ng/µl].  
 3) Hergestellt aus Standard-DNA #9948, St8 [10ng/µl] ist die Referenz, St8\_1 bis St8\_6 sind die Verdünnungen [1-0,02ng/µl].

N Locus Ausfall.  
 Allel Ausfall.  
 Artefakt.

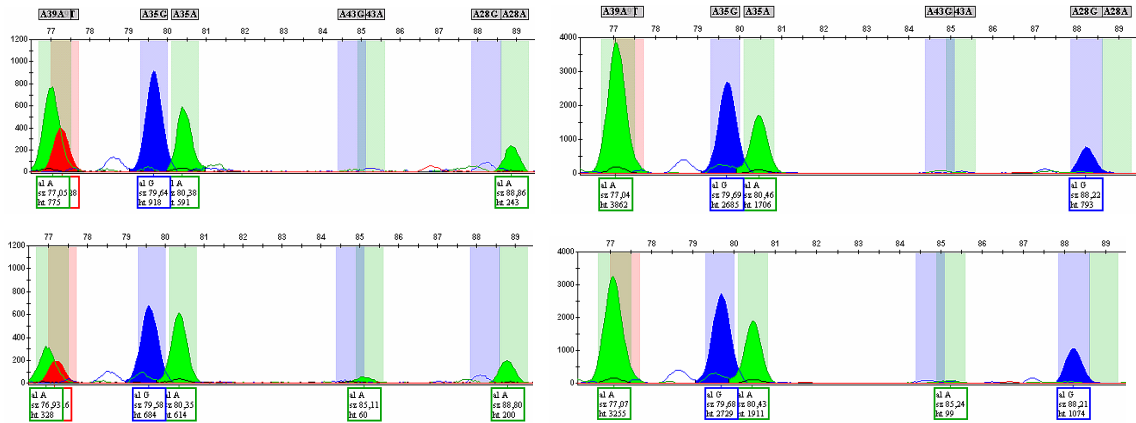


Abbildung 38 SNP bedingte Ausfälle des Locus SNP43. Die Elektropherogramme stellen die Ergebnisse nach Standard-PCR und -Minisequenzierung für zwei Speichelproben dar und darunter nach Reaktionswiederholung mit erhöhten Primern für den SNP43.

Nach Abschluss der forensischen Validierungsstudie wurde die Reagenzienabhängigkeit der 52 autosomalen Multiplex-PCR mittels der elektronischen Microarray-Methode überprüft. Diese Technologie nutzt direkt die PCR-Produkte ohne eine zwischengeschaltete zyklische Reaktion und erlaubt dadurch eine Aussage über die PCR-Primer (Kapitel 2.2.6.3.1). Es wurden Amplifikationen einerseits mit altem und andererseits mit neu bestelltem Primer angesetzt und sechs Loci (SNP17, SNP19, SNP20, SNP38, SNP43 und SNP51) der beiden Multiplexe analysiert (Abbildung 39). Für alle diese sechs Loci konnte mittels der neuen Primer ein viel höheres Signal detektiert werden, während die Loci unter Verwendung der alten Primer nicht auswertbare Signale lieferten. Diese Ergebnisse zeigten nachträglich, dass für die PCR-Amplifikationen der gesamten forensischen Validierungsstudie nicht die optimalen Reagenzien benutzt wurden und liefern damit eine mögliche Erklärung für die allgemein niedrige Genotypisierungsrate.

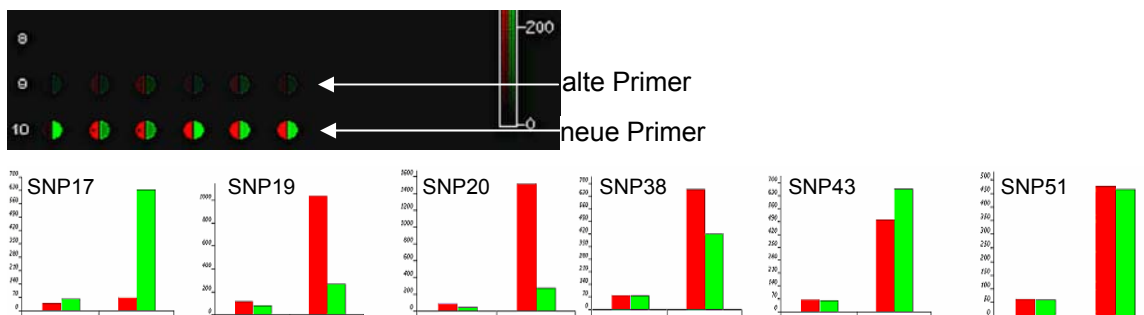


Abbildung 39: Primerkontrolle mittels elektronischer Microarray-Technologie. In Zeile 9 des Mikrochips sind die PCR-Produkte der alten Primer und in Zeile 10 die PCR-Produkte der neuen Primer aufgetragen. Die SNP Reihenfolge entspricht der der Histogramme darunter. Die Histogramme enthalten die Ergebnisse der alten und neuen Primer nebeneinander.

### 3.2.1.3 Ergebnisse der Mischproben

Alle Mischproben besitzen eine Ausgangskonzentration von 1ng/µl und setzen sich aus den zwei Standard-DNAs, die bereits die Basis für die Verdünnungsproben bildeten (Kapitel 2.1.3), in unterschiedlichen Verhältnissen zusammen. In Tabelle 30 sind die ermittelten Genotypen für alle diese Mischproben im Vergleich zu den Referenzgenotypen der zwei Standard-DNAs dargestellt.

Tabelle 30: SNP-Genotypisierung der Mischproben und der Blindproben, geteilt in die 23plex und die 29plex autosomale Minisequenzierung. Die Ergebnisse sind sortiert nach zunehmender Länge der Minisequenzierungsprodukte. Grün markierte Zeilen heben die SNPs hervor, die für die zwei Proben unterschiedliche homozygote Genotypen zeigen. N,N bedeutet SNP Ausfall und die Rote Markierung deutet ein unsicheres Allel an.

SBE nt	SNP1 <sup>1)</sup>	Referenz		Mischproben								Blindproben				
		St7	St8	Mix78 1:10	Mix78 1:5	Mix78 1:2	Mix78 1:1	Mix78 1:1	Mix78 2:1	Mix78 5:1	Mix78 10:1	M1_78 1:1	M4_78 2:1	M3_78 5:1	M2_78 10:1	
23plex autosomale Minisequenzierung	18	01	T,T C,C	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C	C,T	T,T	C,C
	18	07	G,G A,G	A,A	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	G,G	A,G
	24	02	C,C C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T
	25	13	A,A A,G	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A
	29	03	T,T T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T
	29	09	C,C C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C
	34	10	C,C C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C	C,C	C,C
	34	22	G,G G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G
	38	21	A,A A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A
	42	08	C,T C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C	C,T	C,T	C,T	C,T	T,T	C,T
	42	17	G,G A,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G
	46	11	C,T C,C	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C
	46	23	A,A A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G
	50	20	G,G G,T	N	T,T	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G
	54	06	A,A A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A
	58	12	G,G A,A	A,A	A,A	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,A	A,A
	58	19	C,T C,T	T,T	T,C	N	T,T	T,T	T,T	T,T	T,T	T,T	T,T	N	N	N
	62	14	A,A A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,A	A,A
66	18	G,T T,T	T,T	T,T	T,T	T,T	T,T	G,T	G,T	G,T	G,T	G,T	T,T	T,T	T,T	
70	16	G,T T,T	G,T	G,T	G,T	G,T	G,T	G,T	G,T	G,T	G,T	G,T	G,T	G,G	G,T	
74	05	G,G A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	G,G	G,G	A,G	
74	15	T,T C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	T,T	C,T	T,T	T,T	T,T	C,T	
78	04	A,A A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	
29plex autosomale Minisequenzierung	16	41	C,C C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C
	16	46	A,G A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G
	22	33	C,T C,C	C,C	C,C	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C	C,C	C,C
	23	27	A,G A,A	A,A	A,A	A,G	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A
	27	36	A,T A,A	A,A	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T
	27	38	G,G C,G	C,G	C,C	C,C	C,G	C,C	G,G	G,G	C,G	C,G	C,C	C,C	C,C	C,C
	32	44	C,C C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C	C,C
	32	51	A,G G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G
	36	48	T,T T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T
	36	49	A,G A,A	A,A	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,A
	40	34	G,G G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G	G,G
	40	52	A,A A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A
	44	32	C,T C,C	C,C	C,C	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C	C,C
	44	42	A,A A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A
	48	37	C,C C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,C	C,T	C,C	C,T
	48	53	G,G A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	G,G	A,G
	52	25	A,A A,G	A,G	A,G	A,G	A,G	A,G	A,A	A,A	A,A	A,A	A,A	A,G	A,G	A,A
	56	24	T,T T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T	T,T
	56	29	A,G A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G
	60	54	A,A A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	N	A,A	A,A
64	26	T,T G,T	G,T	G,T	G,T	G,T	G,T	T,T	T,T	T,T	T,T	T,T	G,T	T,T	G,G	
68	40	G,G A,A	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	
68	45	C,T C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	C,T	
72	30	C,C A,A	A,A	A,C	A,C	A,C	A,C	A,C	A,C	A,C	A,C	A,C	A,C	A,C	A,A	
76	50	A,G A,A	A,A	A,A	A,A	A,A	A,A	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,A	
80	39	A,T A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	A,T	
84	35	A,A G,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	A,G	G,G	
88	43	A,A A,G	N	N	N	N	N	N	N	N	N	A,A	A,A	N	N	
92	28	A,G A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,A	A,G	A,A	

1) SNPs für 23plex und 29plex, sortiert nach der Länge der Analyseprimer. Für die Grün markierten SNPs gibt es keine Allelüberschneidung für die zwei Ausgangsproben (St7 und St8).

? Unsichere Allelbestimmung.

N Locus Ausfall.

Auffällig ist der geringe Anteil ausgefallener Loci, bzw. die fast ausschließlich SNP-bedingten Ausfälle, wie die des SNPs 43 oder 19 für die sog. Blindproben. Um die Korrelation zwischen dem Mischungsverhältnis und dem detektierten Genotyp einer Probe zu analysieren wurden die SNPs, die durch die zwei Standard-DNAs mit unterschiedlichen Homozygoten repräsentiert sind, in grün hervorgehoben. So besitzt z.B. die Referenzprobe St7 für den SNP01 nur einen Genotyp TT, während die Probe St8 Homozygot C ist. Beim Betrachten des SNPs 12 erscheint eine gewisse Korrelation. Das heißt für die Proben mit einem Mischungsverhältnis von 1:10 und 1:5 (St7:St8) konnte nur das Allel A, welches die Probe St8 repräsentiert, detektiert werden. Dagegen wurde für die Proben umgekehrter Mischungsverhältnisse, d.h. mit St7 als dominierendem Anteil, nur das Allel G detektiert, welches die Probe St7 repräsentiert. Abgesehen von dieser Einzelbeobachtung am SNP12 ist jedoch generell keine Korrelation zwischen dem Mischungsverhältnis und den detektierten Allelen festzustellen. Selbst für die weiteren SNPs mit unterschiedlichen Homozygoten für die Referenz-DNA, wie den SNPs 01, 40 und 35, sind beide Allele, unabhängig von den Mischungsverhältnissen, zu beobachten. Die vier Blindproben wurden als forensisch relevante Proben behandelt ohne zu wissen, dass es sich um Mischproben handelt, hergestellt in gleichen Mischungsverhältnissen derselben Standard-DNAs. Bei diesen Blindproben wurden die Bearbeiter aufgefordert, alle Signale zu dokumentieren und die unzuverlässigen Allele mit einem Fragezeichen zu kennzeichnen. Die Detektion der Genotypen für alle vier Proben erfolgte so eindeutig und balanciert, dass insgesamt für alle Experimente nur zwei Allele (in Tabelle 30 rot markiert) für unzuverlässig deklariert wurden. Nachdem die Genotypisierung und Dokumentation abgeschlossen war, wurde die Herstellung dieser Proben bekannt gemacht. Damit konnte die Beobachtung, dass keine Korrelation zwischen dem Mischungsverhältnis und dem detektierten Genotyp besteht, anhand der Blindproben bestätigt werden. Dies wurde bereits bei der Genotypisierung der bekannten Mischproben festgestellt.

#### **3.2.1.4 Ergebnisse der künstlich degradierten Proben**

Anhand künstlich hergestellter Degradierungsproben mit bekannten DNA-Fragmentlängen sollten die strukturbedingten Vorteile einer SNP-Genotypisierung gegenüber der STR-Typisierung herausgestellt werden.

Dafür wurden Proben von vier unterschiedlichen Längen (200bp, 300bp, 600bp und 1000bp) durch den Einsatz von Ultraschall präpariert (Kapitel 2.2.3). Die Effektivität der Degradierung konnte intensiviert werden, indem sich die Proben während des Ultraschallvorgangs in -20°C kalter Flüssigkeit befanden. Der Sonifikationserfolg wurde im Agarosegel unter UV Licht visualisiert. Dabei wird die unterschiedlich degradierende Wirkung der verschiedenen Sonifikationsvorgänge (10 Sekunden bzw. 40 Sekunden in drei Intervallen mit je 40 Sekunden Ruhephase dazwischen) auf die DNA deutlich (Abbildung 40). Bei den ersten Kontrollversuchen wurde deutlich, dass sich die 10 Sekunden sonifizierten Proben nicht eignen, um 200 bzw. 300bp lange DNA-Fragmente zu extrahieren. Die STR-Typisierung dieser Proben ergab trotz längen abhängiger Extraktion richtige Profile auch für 500bp lange

Loci. Gleichfalls konnten die 40 Sekunden sonifizierte Proben nicht für die 600bp bzw. 1000bp Proben herangezogen werden, weil diese keine vollständigen STR-Profile ergaben. Folglich wurden die 200 bzw. 300bp DNAs aus den 40 Sekunden sonifizierten Proben und die 600 bzw. 1000bp DNAs aus den 10 Sekunden sonifizierten Proben extrahiert (Tabelle 31). Nach der längen abhängigen Extraktion aus dem Agarosegel mittels Dialyseschlauch lag die DNA-Konzentration der vier Proben zwischen 13 und 100ng/µl (Tabelle 31).

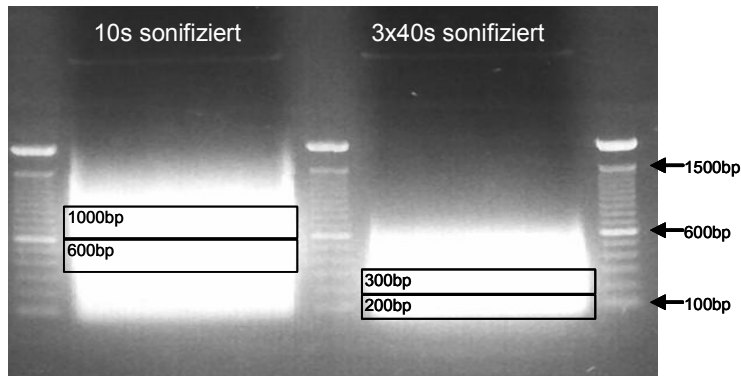


Abbildung 40: Vorgangsabhängige Wirkung der 10 Sekunden bzw. dreimal 40 Sekunden Sonifikation. Die Kästchen weisen auf die Bereiche hin, die aus dem Gel herausgetrennt und für die Elution benutzt wurden. Daraus ergeben sich vier Proben, jeweils nach ihrer Größe benannt 1000bp (600bp-1000bp), 600bp (300bp-600bp), 300bp (200bp-300bp) und 200bp (<200bp).

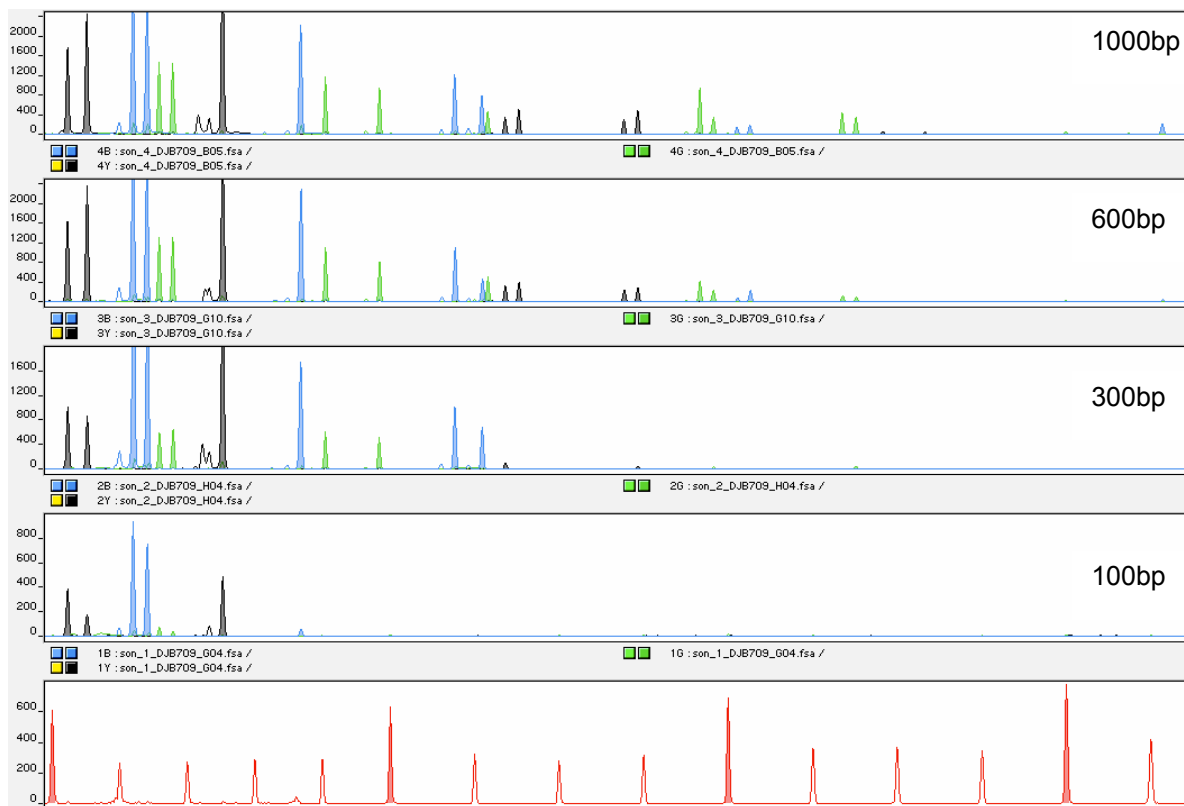


Abbildung 41: STR-Typisierung der natürlich degradierten Proben.

Tabelle 31: DNA-Konzentration der künstlich degradierten Proben.

Probe	Sonifikation	Fragmentlänge	DNA [ng/μl]
<b>200bp</b>	3x40s	< 200bp	13
<b>300bp</b>	3x40s	200-300bp	36
<b>600bp</b>	10s	300-600bp	77
<b>1000bp</b>	10s	600-1000bp	100

Die entsprechenden Fragmentlängen konnten anschließend durch STR-Typisierung verifiziert werden. Die Legitimation der Degradierungsstufe der natürlich hergestellten Proben wurde in dem Londoner Labor mittels des Powerplex® 16 Kits bestätigt (Abbildung 41). Die Auswirkung der künstlichen DNA-Degradierung auf die SNaPshot-Genotypisierung der 52 autosomalen SNPs im Vergleich zur standardisierten STR-Typisierung konnte in London hervorgehoben werden (Kapitel 3.2.2).

### 3.2.2 Gesamtergebnisse der forensischen Validierungsstudie

Die gesamte forensische Validierungsstudie umfasst alle Genotypen von 54 forensisch relevanten Proben, die in fünf beteiligten Laboratorien bearbeitet wurden. Die Probenverteilung, Koordination und Auswertung erfolgte zentral in London (Musgrave-Brown et al. 2007, Kapitel 7.3). Dafür wurden als erstes die Konsensgenotypen für jede Probe und jeden Bearbeiter nach den oben geschilderten Richtlinien (Tabelle 26) bestimmt. Anschließend wurde der Genotypisierungserfolg für jede Probe prozentual berechnet, indem die Anzahl der korrekt typisierten Allele und die sog. drop-in Allele (Artefakte) gezählt wurden. Zusätzlich wurden die Ergebnisse des Powerplex® 16 Kits für jede Probe auf gleiche Weise ausgewertet und den Ergebnissen des SNaPshot 52 Kits gegenübergestellt.

Abbildung 42(a) stellt die Ergebnisse der Verdünnungsproben mittels Powerplex® STR-Analyse und zweier 52plex-SNaPshot-Genotypisierungen verschiedener Labore dar. Bei allen Verdünnungsstufen lieferte die STR-Typisierung deutlich bessere Ergebnisse, selbst für die Probe mit 1ng/μl DNA-Konzentration konnte in keinem der Labore ein 100%-iger 52plex SNP-Genotyp erzielt werden. Zusätzlich ist hier eine starke Differenz zwischen den Laboratorien festzustellen, indem die Genotypisierung in einem Labor immer erfolgreicher war. Gleichzeitig nimmt die Anzahl der Artefakte zu und Labor 2 weist auf eine Korrelation zwischen den Ausfällen und Artefakten hin, indem diese parallel zunehmen (Abbildung 42b). Bei der Auswertung der 26 natürlich degradierten Proben wurden noch erheblich stärkere Diversitäten zwischen den beteiligten Laboratorien festgestellt. Da es sich hier um die generelle methodische Validierung des 52plex-Kits handelt, werden in Abbildung 43a nur die Ergebnisse des Labors mit dem höchsten Erfolg dargestellt und damit die laborspezifische Problematik an dieser Stelle außer Acht gelassen.

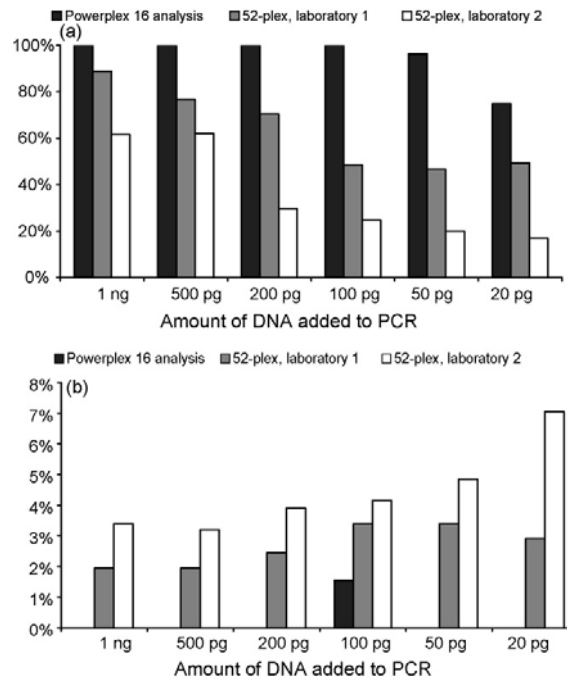


Abbildung 42: STR-Typisierung und zwei SNP-Genotypisierungen der Verdünnungsproben. Die Diagramme zeigen den Typisierungserfolg (a) und die Detektion von Artefakten (b) in Prozent (Musgrave-Brown et al. 2007).

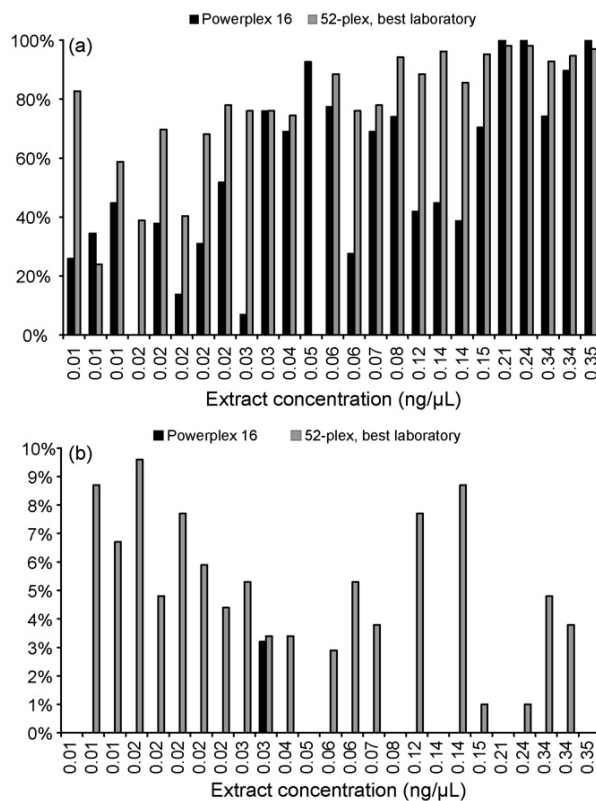


Abbildung 43: STR-Typisierung und SNP-Genotypisierung der natürlich degradierten Proben. Die Diagramme zeigen den Typisierungserfolg (a) und die Detektion von Artefakten (drop in) (b) in Prozent (Musgrave-Brown et al. 2007).

Die natürlich degradierten Proben, die gleichermaßen qualitative und quantitative Herausforderungen stellten, lieferten sehr variable Ergebnisse. Dennoch war die SNaPshot-Genotypisierung dieser Proben meist erfolgreicher als die STR-Typisierung (Abbildung 43a). Bei 21 von 26 Proben konnte mittels SNaPshot ein besseres Ergebnis erzielt werden, gleichzeitig war aber eine häufige Artefaktbildung zu beobachten (Abbildung 43b).

Die Auswertung weiterer forensisch relevanter Proben, wie künstlich degradierte Proben, Extrakte vom telogenen Haar und vom zersetztem Muskelgewebe, liefert eindeutige Beweise für die Vorteile der SNP-Genotypisierung gegenüber der in der Routine verwendeten STR-Analyse. Bei den natürlich degradierten Proben nehmen die STR-Typisierungserfolge rapide ab, während die SNP-Genotypisierungen konstant bleiben (Abbildung 44). Die SNaPshot-Genotypisierung bleibt bei der Analyse von zersetztem Muskelgewebe ebenfalls konstant, obwohl die STR-Typisierung versagt (Abbildung 45).

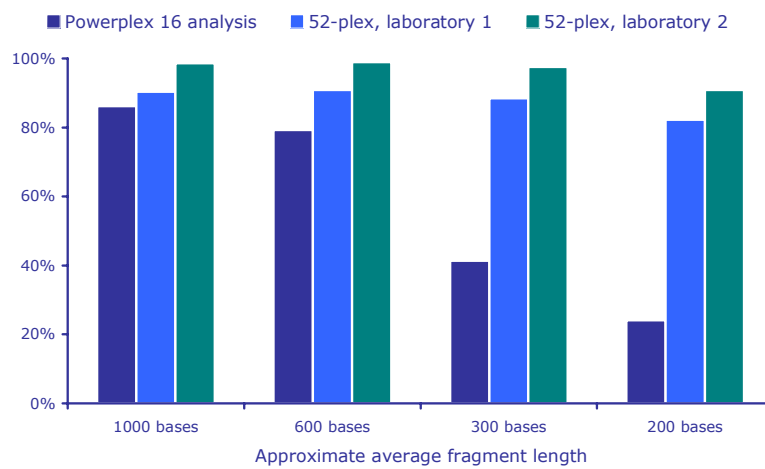


Abbildung 44: STR-Typisierung und zwei SNP-Genotypisierungen der künstlich degradierten Proben. Das Diagramm zeigt den Typisierungserfolg in Prozent.

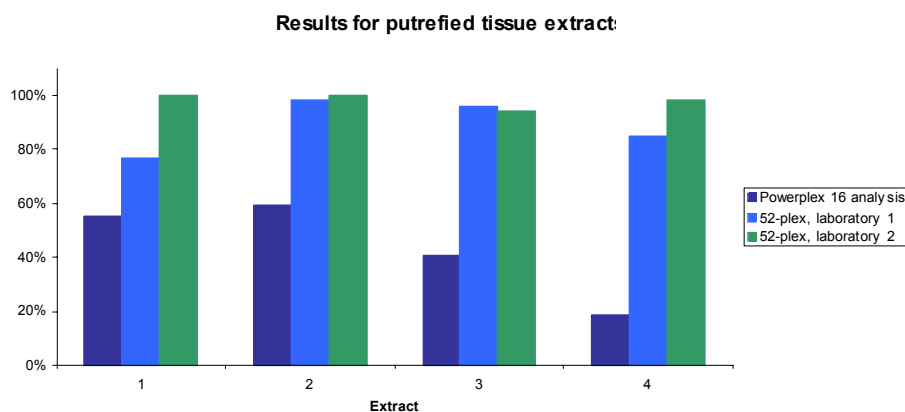


Abbildung 45: STR-Typisierung und zwei SNP-Genotypisierungen der Proben von zersetztem Muskelgewebe. Das Diagramm zeigt den Typisierungserfolg in Prozent.



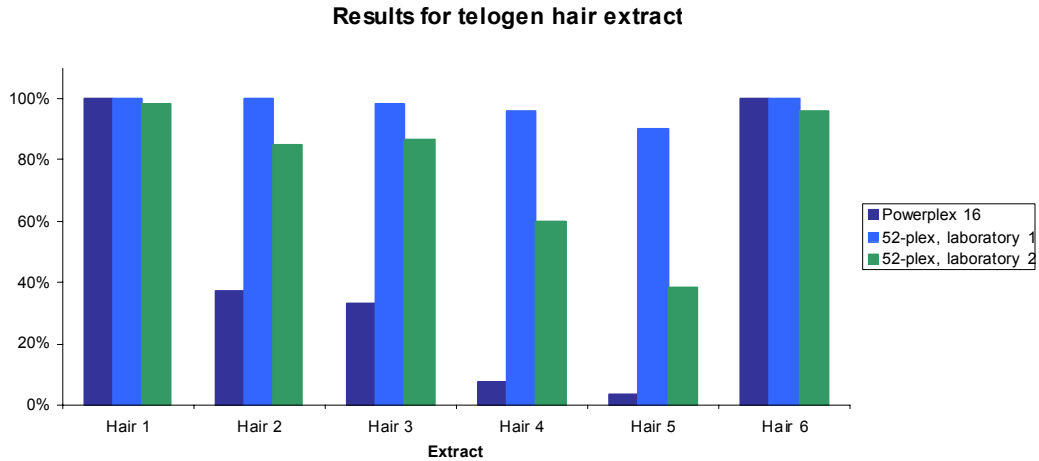


Abbildung 46: STR-Typisierung und zwei SNP-Genotypisierungen der Proben von telogenen Haaren. Das Diagramm zeigt den Typisierungserfolg in Prozent.

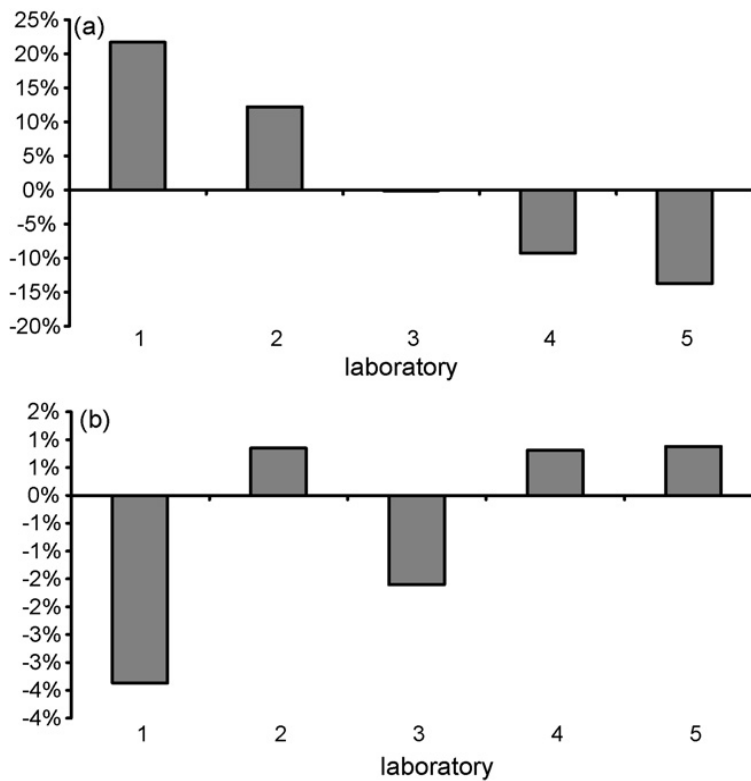


Abbildung 47: Genotypisierungsunterschiede der beteiligten Laboratorien. Die Diagramme zeigen die prozentuale Abweichung des Typisierungserfolgs (a) und der Detektion von Artefakten (b) vom Mittelwert (Musgrave-Brown et al. 2007).

Auch die Analysen der telogenen Haare konnten die Vorteile der SNP-Genotypisierung im Vergleich zu STR eindeutig beweisen (Abbildung 46). Selbst beim Haar 5, wo die STR-Typisierung gänzlich versagte, konnte in einem Labor ein Genotyp von 90% bestimmt werden, gleichwohl zeigt aber diese Probe erneut laborbedingte Unterschiede (Abbildung 46). Die große Diversität zwischen den Laboratorien, die sich insbesondere bei der Untersuchung der Verdünnungsproben und der natürlich degradierten Proben zeigte,

erforderte eine Gegenüberstellung der Beteiligten. Dafür wurde der durchschnittliche Prozentwert der erfolgreich genotypisierten Loci für jede Probe ausgerechnet und dessen Abweichung vom Mittelwert für jedes Labor einzeln errechnet. Nachdem diese Rechnung für alle Proben erfolgte, konnte die durchschnittliche Abweichung für jedes Labor dargestellt werden (Abbildung 47a). Nach dem gleichen Verfahren wurden die durchschnittlichen Abweichungen auch für die Artefakte berechnet und in Abbildung 47b dargestellt. Beide Diagramme weisen für die SNaPshot-Genotypisierung auf eine starke Laborabhängigkeit hin, nicht zuletzt zeigt sich das auch darin, dass das Labor1 mit der erfolgreichsten Typisierung am wenigsten Artefakte detektiert hat, während die Labor4 und Labor5 mit einer Genotypisierungsrate von knapp 10 bzw. 15 Prozent unterhalb des Durchschnitts auch die meisten Artefakte detektierten (Abbildung 47).

### 3.3 Ergebnisse der Microarray-Technologie

#### 3.3.1 Amplicon down-Assayetablierung

Die Etablierung des Amplicon down-Assays erfolgte zunächst mittels Singleplex- und Duplex-PCR-Amplifikationen, gefolgt von einer Pentaplex-PCR. Letztere wurde anschließend verwendet, um die Anwendbarkeit der verschiedenen Aufreinigungsmethoden zu bestimmen. Abbildung 48 zeigt den ersten Versuch, der ein Singleplex für SNP01 und ein Duplex für SNP09 und SNP11 mittels Amplicon down-Assay typisiert. Die Genotypisierung lieferte eindeutige und zuverlässige Ergebnisse für beide PCR-Produkte (Abbildung 48). Die automatisch ermittelten Fluoreszenzsignale sind abzüglich der Stärke des Backgrounds zu verstehen und zeigen keine Kontaminationen zwischen den einzelnen Reportern. Diese Ergebnisse bestätigten sich durch Reproduzierbarkeit.

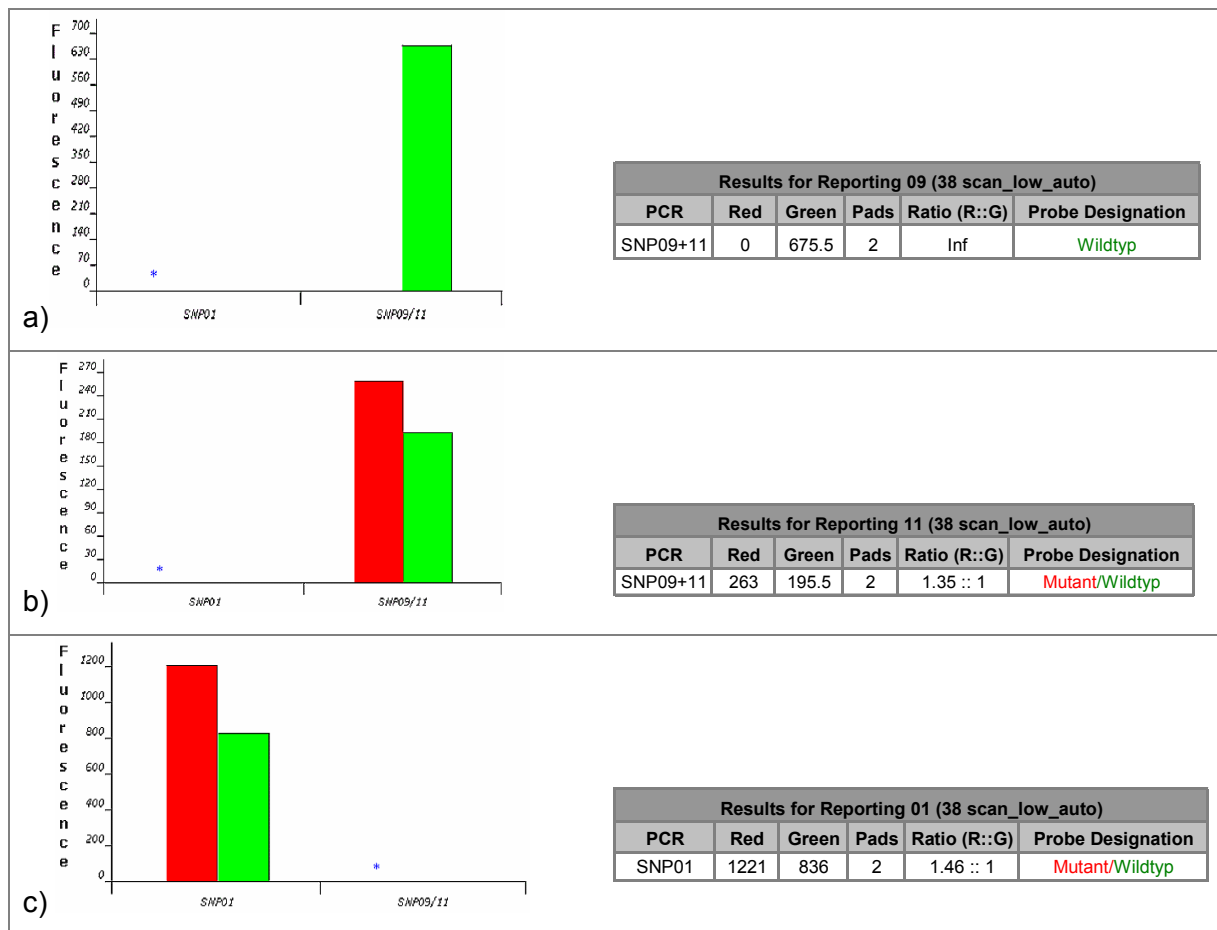


Abbildung 48: Nach der elektronischen Adressierung der Singleplex-PCR für SNP01 und Duplex-PCR für SNPs 09 und 11 erfolgte als erstes die Hybridisierung für SNP09 (a). Nach der thermischen Diskriminierung, Scannen und Stripping wurde der Vorgang für SNP11 (b) und schließlich für SNP01 (c) wiederholt. Rechts neben den Histogrammen sind jeweils die Rohdaten der automatisierten Auswertung abzulesen.

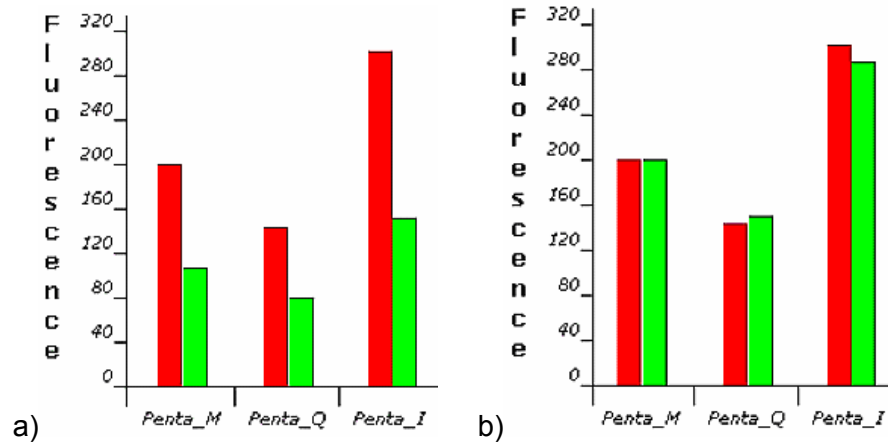


Abbildung 49: Beide Histogramme stellen den gleichen Scan vor mit dem Unterschied, dass für Histogramm a) keine Heterozygotenkontrolle zur Normalisierung gewählt wurde, während für b) die Probe *Penta\_M* als Heterozygot definiert wurde.

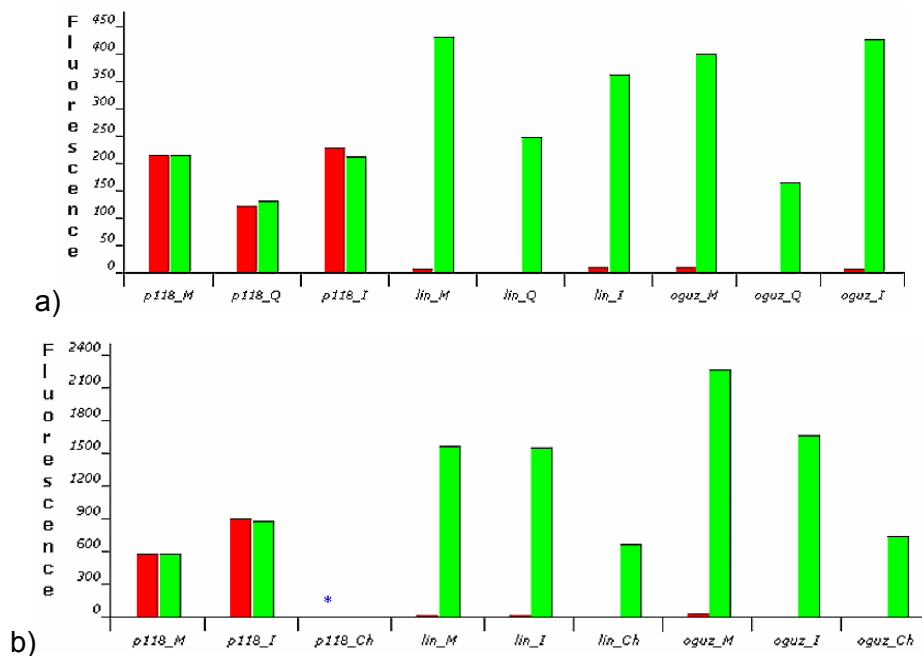


Abbildung 50: Vergleich verschiedener Aufreinigungsmethoden für SNP1. Dafür wurden drei Pentaplex-Proben (sog. *p118*, *lin* und *oguz*) mittels vier verschiedener Methoden aufgereinigt (M = Millipore, Q = Qiagen, I = Invitex und Ch = ChargeSwitch). Histogramm a) zeigt die Unterschiede zwischen den Methoden Millipore, Qiagen und Invitex, während Histogramm b) zwischen Millipore, Invitex und ChargeSwitch unterscheidet.

Die Amplicon down-Analysen von weiteren Singleplex- bzw. Duplex-PCR-Produkten lieferten stets auswertbare Resultate. Damit konnte das Assay rasch auf eine Pentaplex-PCR ausgeweitet werden. Bei der Genotypisierung von Pentaplex-PCR-Produkten, wo sukzessives Reporting und Stripping je fünf Mal erfolgte, hat es sich bewährt, eine Heterozygotenkontrolle zu definieren. Abbildung 49 verdeutlicht die Unterschiede bei der Auswertung ohne (a) und mit Heterozygotenkontrolle (b). Mit Hilfe einer Kontrolle wird bei der automatischen Auswertung das Verhältnis zwischen den zwei Fluoreszenzfarben normalisiert und führt zu eindeutigen Ergebnissen (Abbildung 49b).

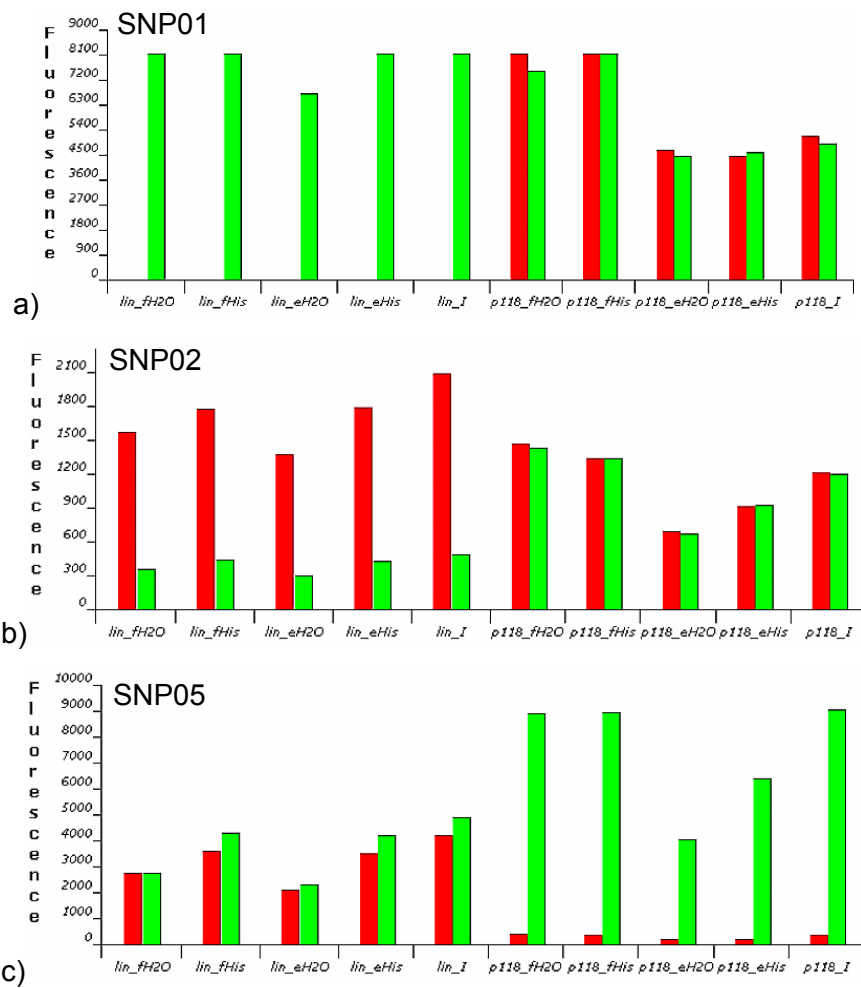


Abbildung 51: Schwankungen der Millipore-Aufreinigungsmethode beispielhaft für SNP01 a), SNP02 b) und SNP05 c) dargestellt. Zwei Pentplex-Proben (sog. lin und p118) wurden auf fünf verschiedene Weisen aufgereinigt, fH<sub>2</sub>O = Millipore, neu angebrochene Filter-Mikrotiterplatte und H<sub>2</sub>O eluiert, fHis = Millipore, neue Filter-Mikrotiterplatte und His eluiert, eH<sub>2</sub>O = Millipore benutzte Filter-Mikrotiterplatte und H<sub>2</sub>O eluiert, eHis = Millipore, benutzte Filter-Mikrotiterplatte und His eluiert, I = Invitex Aufreinigung.

Bei weiteren Analysen der Pentaplex-Proben zeigten sich erste Anzeichen für die entscheidende Bedeutung der Aufreinigungsmethoden, mit deren Hilfe die Produkte von Salzen befreit und gleichzeitig die Produktmengen konzentriert wurden. Abbildung 50 zeigt beispielhaft die Ergebnisse für SNP01 und drei verschiedene Pentaplex-PCR-Produkte (sog. p118, lin, oguz), die mittels vier unterschiedlicher Methoden (Millipore, Qiagen, Invitex und ChargeSwitch, Kapitel 2.2.4.5.2) gereinigt wurden. Es sind wiederkehrende Muster zu erkennen, indem die Fluoreszenzstärken Qiagen aufgereinigter Produkte immer niedrigere Signale lieferten als die Millipore- und Invitex-Produkte des ersten Experiments (Abbildung 50a). Bei dem darauf folgenden Experiment setzten sich die Millipore- und Invitex-Proben noch deutlicher von den ChargeSwitch gereinigten Proben ab (Abbildung 50b). Die Korrelation zwischen der Stärke der Fluoreszenzsignale und den Aufreinigungsmethoden konnte in mehreren Versuchen reproduziert werden. Damit stand fest, dass die Millipore- und

Invitex-Aufreinigungsmethoden für die elektronische Microarray-Typisierung mittels Amplicon down-Assay besonders geeignete PCR-Produkte lieferten.

Durch die zahlreichen Versuche zur vergleichenden Beurteilung der Aufreinigungsmethoden wurde eine gewisse Variabilität der Ergebnisse der Millipore gereinigten Produkten sichtbar. Die Schwankungen konnten schließlich darauf zurückgeführt werden, dass es einen Unterschied machte, ob eine neu angebrochene 96-er Mikrotiterplatte mit Filtermembranboden oder eine bereits mehrmals benutzte Platte zum Filtrieren verwendet wurde. Diese physikalische Abhängigkeit zeigt sich in der Abbildung 51. Dabei ergaben die Proben, die mittels einer frischen Mikrotiterplatte gereinigt wurden, stets höhere und stabilere Signale (fH<sub>2</sub>O und fHis). Gleichzeitig konnte das Ergebnis bzw. die Signalstärke verbessert werden, indem die PCR-Produkte mittels Histidin [100mM] statt HPLC-Wasser von der Filtermembran eluiert wurden (Abbildung 51).

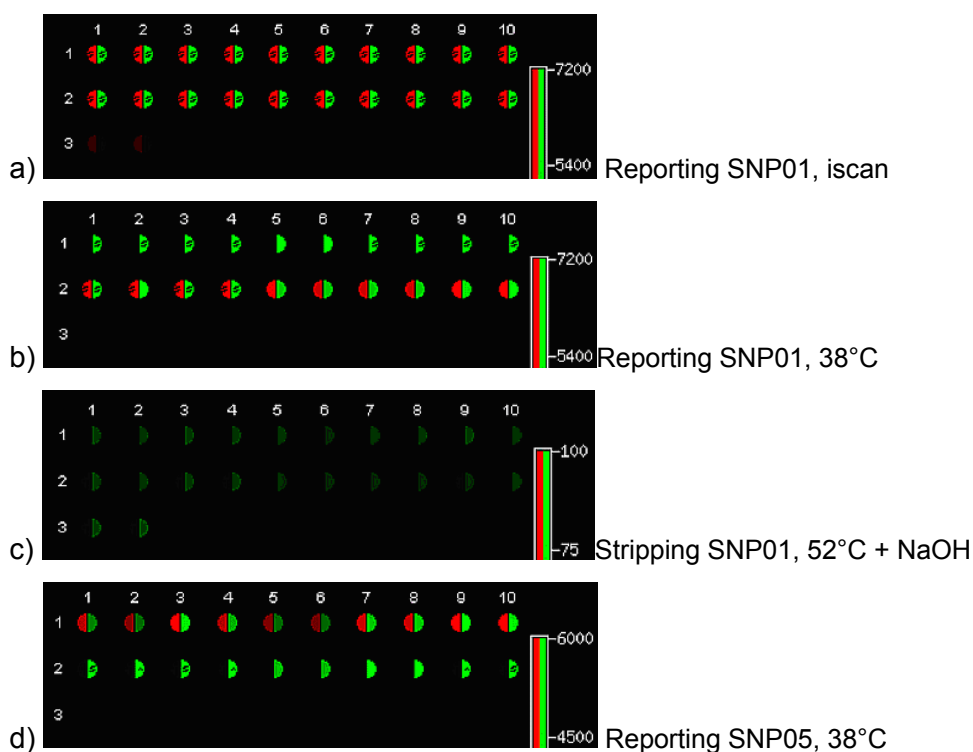


Abbildung 52: Scanbilder für zwei Pentaplex-Proben. Bild a) zeigt einen sogenannten Initialscan (iscan); hier hat die Hybridisierung der Reporter für SNP01, aber noch keine Diskriminierung stattgefunden. Der Scan nach thermischer Diskriminierung ist auf Bild b) zu sehen. Nach thermischem Stripping und Waschung mit NaOH wurden keine spezifischen Signale detektiert (c). Bild d) zeigt den Scan nach Hybridisierung und thermischer Diskriminierung der Reporter für SNP05. Zwischen Stripping SNP01 (c) und Reporting SNP05 (d) fanden zusätzlich die Hybridisierung, das Scannen und das Stripping von SNP02 statt (hier nicht gezeigt).

Darüber hinaus konnte bei diesem Experiment die Effektivität des Strippings gesteigert werden, indem ein passiver Denaturierungsschritt mittels NaOH vor jeder Hybridisierung eingeführt wurde (Abbildung 52c). Die Scanbilder in Abbildung 52 demonstrieren die Wirkung der Stringenz bei thermischer Diskriminierung (b), beim thermischen und chemischen Stripping (c) und bei thermischer Diskriminierung des darauf folgenden SNPs (d). Die

Zuverlässigkeit des etablierten Amplicon down-Assays ist in der zweiten Reihe des Mikrochips zu erkennen, hier wird nicht nur die einwandfreie Diskriminierung deutlich, sondern auch das saubere kontaminationsfreie Stripping (Abbildung 52c). In Abbildung 52d, zweite Reihe, zeigt die homozygote Wildtyp Probe keine Kontaminationen durch den roten Farbstoff, obwohl diese Probe für die zwei vorhergehend typisierten SNPs (SNP01 und SNP02) Heterozygot ist (Abbildung 51a und b).

Diese Ergebnisse liefern Beweise für einen erfolgreich etablierten Amplicon down-Assay für Pentaplex-PCR-Produkte. Die Resultate bestätigen nicht nur den Assayaufbau bzw. Ablauf und die geeignete Aufreinigungsmethode, sondern auch die erfolgreiche Definition der Oligodesignkriterien.

### 3.3.2 APEX-Assaytablierung

Die Ergebnisse zweier Experimente demonstrieren, dass der Aufbau und die Durchführung eines APEX-Assays direkt auf der NanoChip-Oberfläche möglich ist (Abbildung 53). Die zwei Histogramme zeigen die Ergebnisse einer APEX-Reaktion, welche parallel zwei Proben für acht verschiedene SNPs typisiert. Die Beispiele in Abbildung 54 zeigen, dass unterschiedliche Personen in einem Experiment auch unterschiedliche Genotypen ergeben können. Damit lieferte das APEX-Experiment, das ohne Assayoptimierung erfolgte, ausreichend Hinweise zur Verwendbarkeit der Nanogen-Plattform für eine Polymerase induzierte Extensionsmethode.

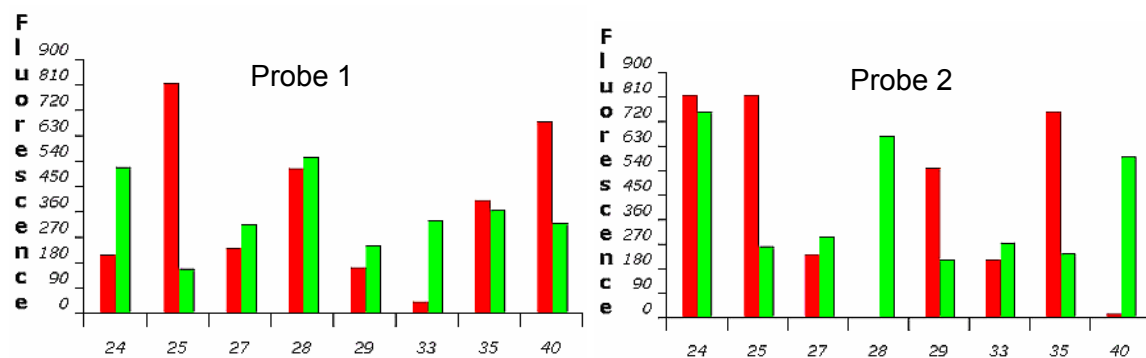


Abbildung 53: APEX-Experiment für zwei 8plex-Proben. Die X-Achse gibt die acht untersuchten SNPs wieder (SNP24, SNP25, SNP27, SNP28, SNP29, SNP33, SNP35 und SNP40).

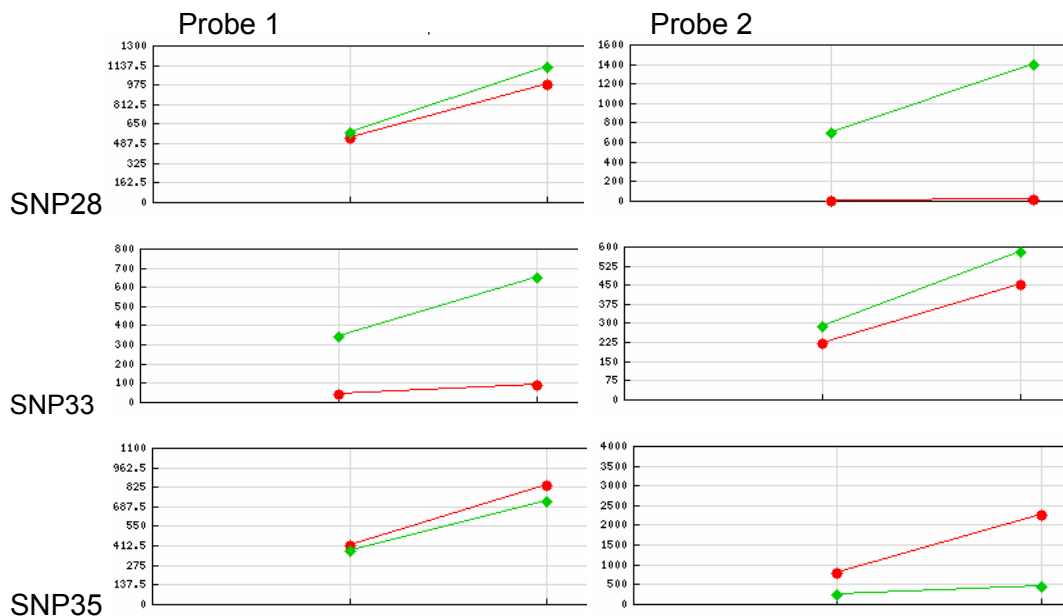


Abbildung 54: Trendgrafik für drei SNPs (SNP28, SNP33 und SNP35), die für die zwei Proben unterschiedliche Genotypen ergeben. Auf der Y-Achse sind die Fluoreszenzwerte angegeben und auf der X-Achse zwei Scanvorgänge, Akkumulationszeit 500 Millisekunden bzw. automatische Akkumulation bei mittlerer PMT-Stärke gescannt.



### 3.3.3 Capture down-Assay

Die Ergebnisse des Capture down-Assays sind in drei Abschnitte gegliedert. Kapitel 3.3.3.1 behandelt die einzelnen Arbeitsschritte, die den Capture down-Assayaufbau etablieren. Darauf folgend sind in Kapitel 3.3.3.2 die Maßnahmen beschrieben, die der Optimierung des angewandten Assays dienen. Schließlich werden in Kapitel 3.3.3.3 die Ergebnisse der Validierung des etablierten und optimierten Capture down-Assays dargestellt.

#### 3.3.3.1 Capture down-Assayetablierung

Die Etablierung des Capture down-Assays basierte ausschließlich auf theoretischen Vorgaben und auf eigenen Erfahrungen, die während der Etablierung des Amplicon down-Assays gesammelt wurden. Zunächst wurden die fünf SNPs, die bereits für Amplicon down etabliert wurden, herangezogen und im Singleplex-, Duplex- bzw. Pentaplexformat analysiert. Nach mehreren Optimierungsschritten konnte die Anzahl der SNPs, die mittels Capture down-Assay parallel analysiert werden, auf 13 erhöht werden.

Die ersten single- und duplexbasierten Versuche lieferten einwandfreie und zuverlässige Typisierungen (Abbildung 55). Dabei war auch kein Ergebnisverlust oder Unterschied zwischen einer Singleplex- und Duplex-Analyse zu beobachten (Abbildung 55).

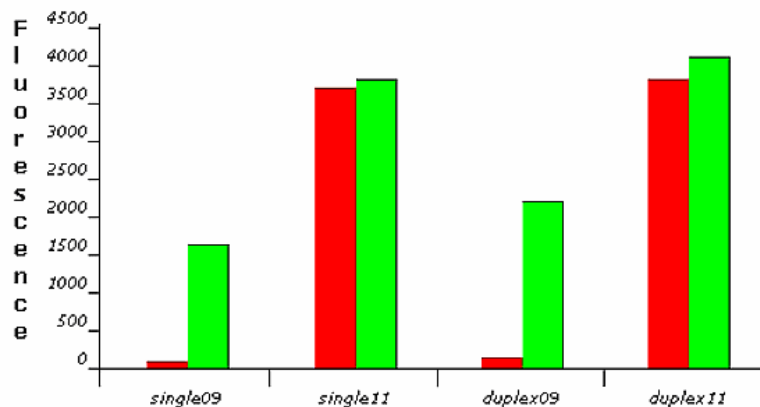


Abbildung 55: Genotypisierung einer Probe für zwei SNPs (SNP09 und SNP11) in Singleplex- und Duplexformat. Die ersten zwei Signalpaare stellen die Ergebnisse einer Singleplex und die letzten beiden einer Duplex dar.

Die vielversprechenden Resultate der Singleplex- und Duplex-Experimente erlaubten eine rasche Ausweitung der Analyse auf Pentaplex-Ansätze. Dabei wiesen bereits die ersten Experimente auf Schwierigkeiten hin und machten deutlich, dass die Problemlösung einer komplexen Analyse für mehrere SNPs umfassender Optimierungen bedarf.

Die Pentaplex-Proben lieferten allgemein niedrige Signale, sodass die Ergebnisse nur bei einer Einstellung der Signal-to-noise ratio for Pad Exclusion or No Designation: 2:1 und Minimum background-subtracted signal for Pad Inclusion: 10 darzustellen sind (Abbildung 56). Abgesehen von

unbalancierten Signalhöhen der verschiedenen SNPs (Abbildung 56), wurde auch eine Unhomogenität innerhalb zweier identisch behandelter Pads beobachtet. Damit waren die ersten Beweise erbracht, dass ein unzuverlässiges Ergebnis auch auf Mikrochip-Fehler oder Pad-Aktivierungsunterschiede zurückgeführt werden kann. So wurden zum Beispiel bei dem Experiment in Abbildung 57 die Pads 3,7 und 3,8 bzw. Pads 6,3 und 6,4 oder auch 6,5 und 6,6 jeweils identisch behandelt und dennoch zeigen die Ergebnisse starke Abweichungen (Abbildung 57).

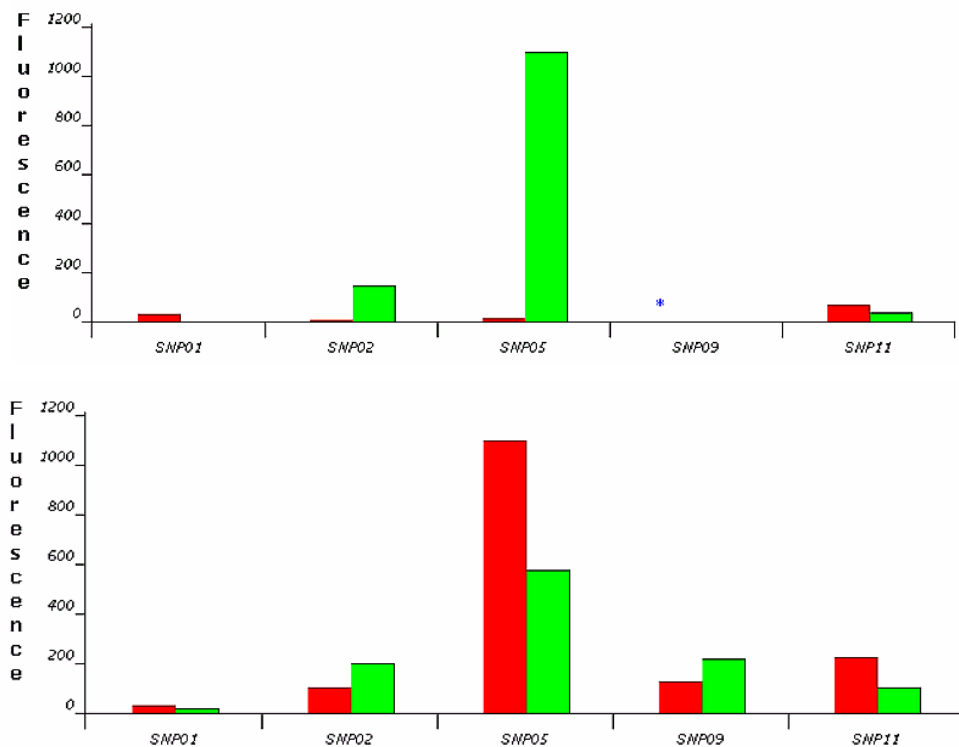


Abbildung 56: Zwei Histogramme stellen die Ergebnisse der ersten pentaplex Versuche für zwei Proben dar. Dabei wurden die SNPs 01, 02, 05, 09 und 11 typisiert.

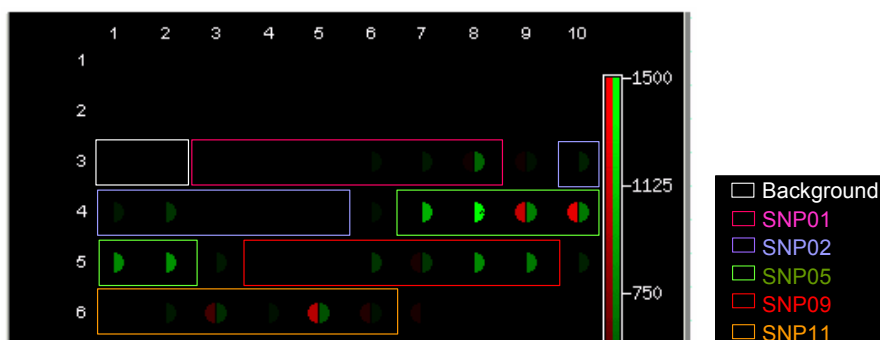


Abbildung 57: Scanbild der Genotypisierung von drei Proben für fünf SNPs. Die Anordnung der Capture-Adressierungen ist mittels Farbkästchen angedeutet. Zu jeder Capture-Oligo wurden drei Proben doppelt hybridisiert, z.B. auf die SNP01 Capture wurden auf Pads 3,3 und 3,4 die erste Probe, auf 3,5 und 3,6 die zweite und auf 3,7 und 3,8 schließlich die dritte Probe hybridisiert.

Als nächstes erfolgten mehrere Experimente zur Optimierung der PCR-Hybridisierung. Abbildung 58a zeigt die Ergebnisse einer Duplex-Probe, welche manuell auf das Cartridge gegeben wurde, damit die Hybridisierung passiv bei Raumtemperatur erfolgte. Im direkten Vergleich dazu sind in Abbildung 58b die Ergebnisse der elektronischen Hybridisierung der gleichen Probe dargestellt. Dabei sind die großen Unterschiede in der Signalstärke sehr auffällig. Die elektronisch hybridisierte Probe liefert zuverlässige Resultate unter Standardkonditionen (Abbildung 58b), während die passive Hybridisierung sehr niedrige Signale ergab, sodass die Auswertung nur mit einer Einstellung der Signal-to-noise ratio for Pad Exclusion or No Designation: 2:1 darstellbar ist (Abbildung 58a).

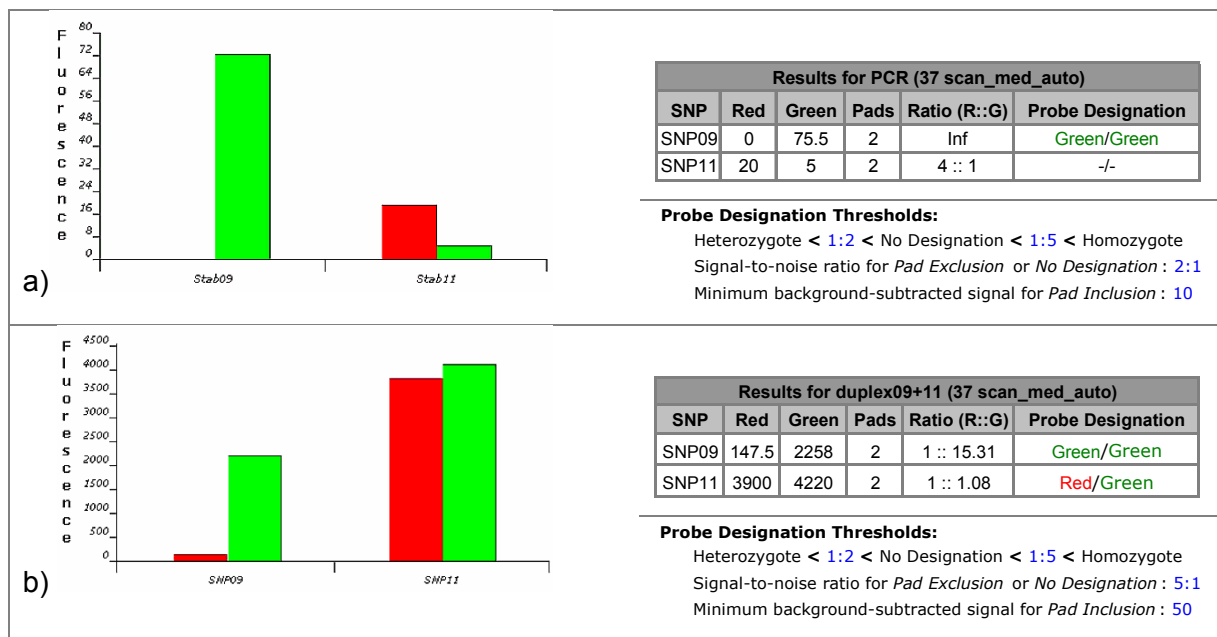


Abbildung 58: Vergleich passiver und elektronischer Hybridisierung einer Duplex-Probe. Die passive Hybridisierung (a) liefert niedrige und unklare Ergebnisse, während nach einer elektronischen Hybridisierung (b) die Signale einwandfrei auswertbar sind, auch unter Standardkonditionen der Auswertungseinstellungen. Die Unterschiede der Signalstärken können in den Rohdatentabellen (jeweils rechts von den Histogrammen) abgelesen werden.

Nach der Qualitätssicherung des Ablaufs der PCR-Hybridisierung war es wichtig, die geeigneten Puffer-Bedingungen für die elektronische Hybridisierung zu bestimmen. Aus diesem Grund wurde ein Versuch erstellt, um die Wirkungen von His [50mM] und HSB [50mM] bei der elektronischen Hybridisierung zu vergleichen. Beide Puffer lieferten Ergebnisse, die unter Standardkonditionen ausgewertet werden konnten. Jedoch waren bei Verwendung von His als Hybridisierungspuffer (Abbildung 59a) die Fluoreszenzsignale wesentlich höher im Vergleich zum HSB-Puffer (Abbildung 59b). Dieses Experiment wurde zu einem späteren Zeitpunkt ergänzt, indem Arginin [100mM] als Hybridisierungspuffer getestet und bestätigt werden konnte (Abbildung 59c). Weitere Schritte während der Assayetablierung erfolgten mit His [50mM] als PCR-Hybridisierungspuffer, erst bei der Validierung wurde Arg [100mM] standardisiert eingesetzt.

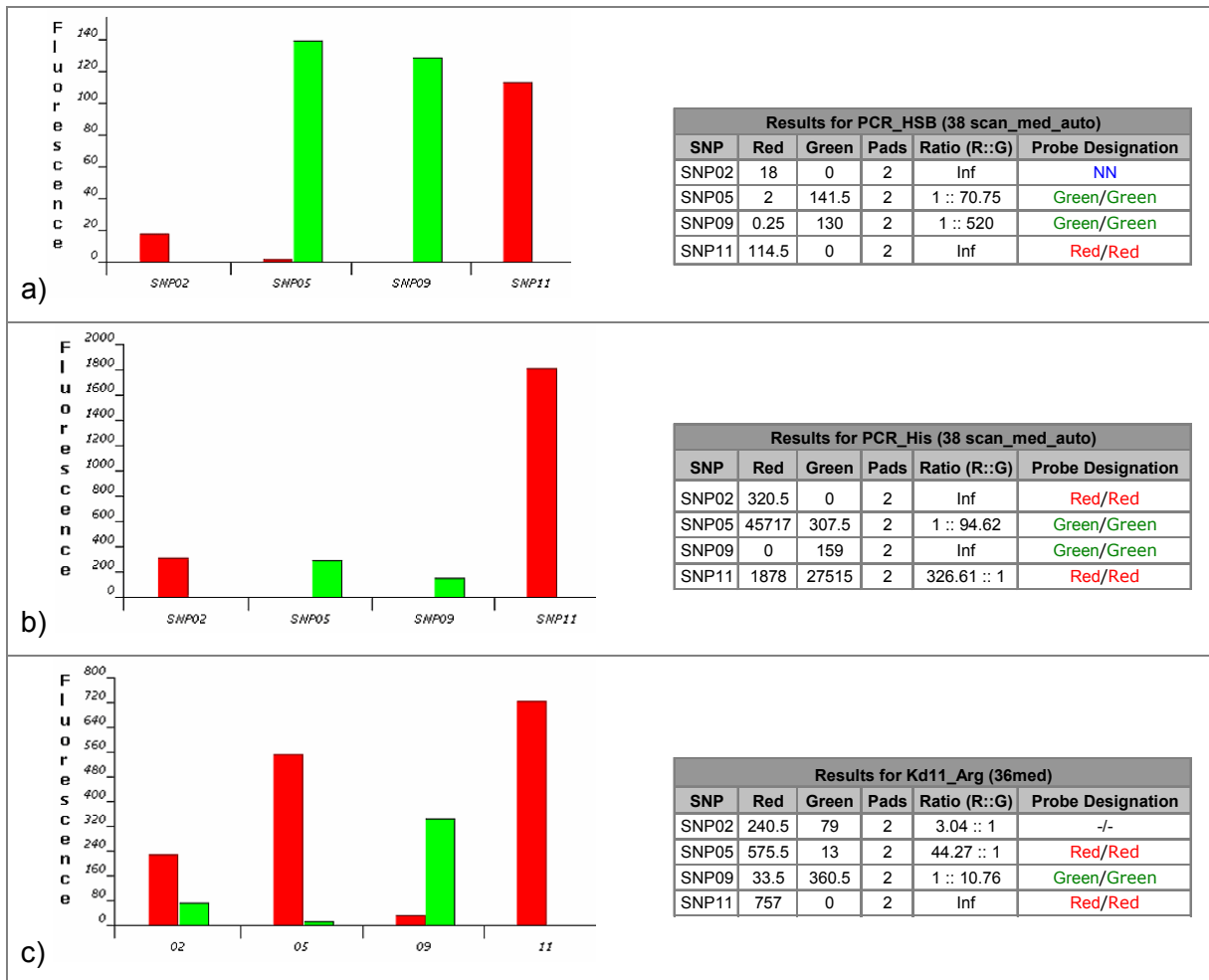


Abbildung 59: Vergleich unterschiedlicher Pufferbedingungen bei der elektronischen Hybridisierung. Die niedrigsten Signale liefert das Histogramm a), wo die Hybridisierung in HSB-Puffer erfolgte. An der Rohdatentabelle kann abgelesen werden, dass SNP02 unauswertbar ist, das liegt daran, dass die Signal-to-noise ratio for Pad Exclusion or No Designation nicht einmal 2:1 erreicht. Histogramm b) mit Hybridisierung in His-Puffer liefert sichere Signale, jedoch in sehr unhomogener Höhe. Zu einem späteren Zeitpunkt wurde die Hybridisierung in Arg-Puffer bestätigt (c) mit stets homogenen Signalen als bei der Hybridisierung mit His-Puffer.

Nachdem die PCR-Hybridisierung in Ablauf, Zusammensetzung und Bedingung etabliert war, wurde größeres Augenmerk auf die Diskriminatoren, die Universal reporter und das Readerprotokoll gelegt. Dazu wurden Pentaplex-Ansätze von verschiedenen Personen mit unterschiedlichen Genotypen gleichzeitig analysiert. Es sollten die Eigenschaften zur Hybridisierung und Diskriminierung einzelner SNPs erforscht werden. Parallel zu der Microarray-Analyse wurden die Proben auch mittels SNaPshot-Standardprotokoll als Referenz typisiert, damit konnten die Konkordanzen und die Diskriminierungspotentiale der SNPs bestimmt werden (Abbildung 60). Beide Proben in Abbildung 60 zeigen eine hohe Übereinstimmung mit den mittels SNaPshot ermittelten Genotypen. Bis auf SNP02 der ersten PCR-Probe stimmen alle Genotypen überein. Beide PCR-Produkte konnten mit großer Zuverlässigkeit typisiert werden. Auch zeigen die parallelen elektronischen Hybridisierungen der beiden Proben keine gegenseitigen Kontaminationen.

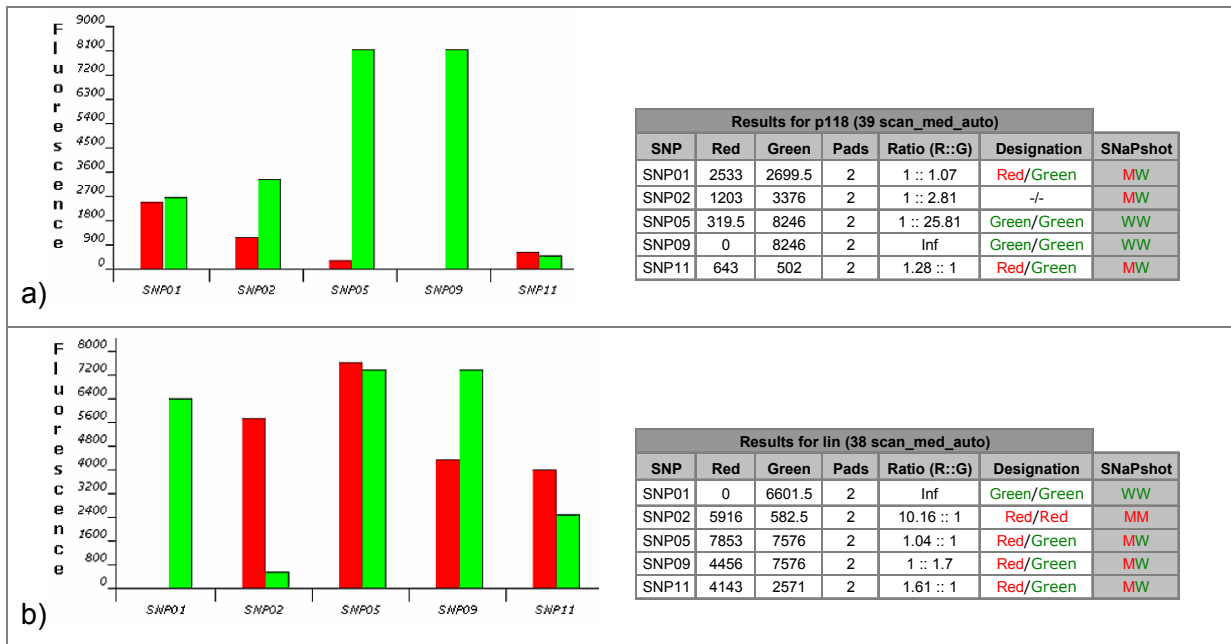


Abbildung 60: Ergebnisse von zwei Pentaplex-Proben, die parallel typisiert wurden. Die automatische Auswertung erfolgte nach Standardeinstellung. Die Rohdatentabellen sind mit den SNaPshot ermittelten Genotypen (Grau unterlegte Spalte) ergänzt. Damit kann die Konkordanz zwischen Nanogen- und SNaPshot-Genotypisierung bestimmt werden.

Abbildung 61 stellt die sogenannten Trendgrafiken der einzelnen SNPs bei differenzierter Diskriminierungstemperatur dar. Dabei ist das Diskriminierungspotenzial jedes einzelnen SNPs zu beobachten und anhand des zu erwartenden Genotyps die richtige Temperatur für eine sichere Diskriminierung zu bestimmen. So ist hier zum Beispiel abzulesen, dass die mismatch Mutante für SNP05 erst bei 39°C diskriminiert, das heißt ein Scan bei 37°C oder 38°C hätte einen falschen Heterozygoten ergeben. Gleichzeitig bedeutet die Temperaturerhöhung auf 39°C Verluste in der Signalstärke bei anderen SNPs (z.B. SNP01, SNP02 und SNP11 Abbildung 61). Die ersten erfolgreichen Ergebnisse weisen darauf hin, dass das Capture down-Assay für die Multiplex-Analyse umfassende Optimierungen verlangt und dabei jeder SNP einzeln behandelt werden muss.

Da es sich gezeigt hat, dass eine Assayetablierung mit nur fünf bestimmten SNPs nicht auf weitere SNPs übertragbar ist, sondern alle SNPs einzeln behandelt werden müssen, war eine Multiplexerweiterung sinnvoll. Zunächst wurden noch weitere acht SNPs in die Multiplex-PCR mit aufgenommen, sodass ein Multiplex aus 13 SNPs zur Verfügung stand. Die Ergebnisse in Abbildung 62 zeigen, dass die passive Hybridisierung der Diskriminatoren und der Universal reporter in zwei Schritten erfolgreich sein kann. Gleichzeitig ist aber zu bemerken, dass nicht alle SNPs einen eindeutigen Genotyp ergeben und das Verhältnis zwischen Signal und Background zu klein ist, sodass die Auswertung nur bei der 2:1 Einstellung erfolgen kann. Im Gegensatz dazu kann eine Probe nach elektronischer Hybridisierung der Diskriminatoren nicht ausgewertet werden (Abbildung 63a). Hierbei sind die Signale allgemein hoch und balanciert, aber es findet keine Diskriminierung während des Readerprotokolls statt. Dieser Zustand ändert sich auch dann nicht, wenn die Diskriminierungstemperatur im Reader auf 41°C erhöht wird (Abbildung 63b).

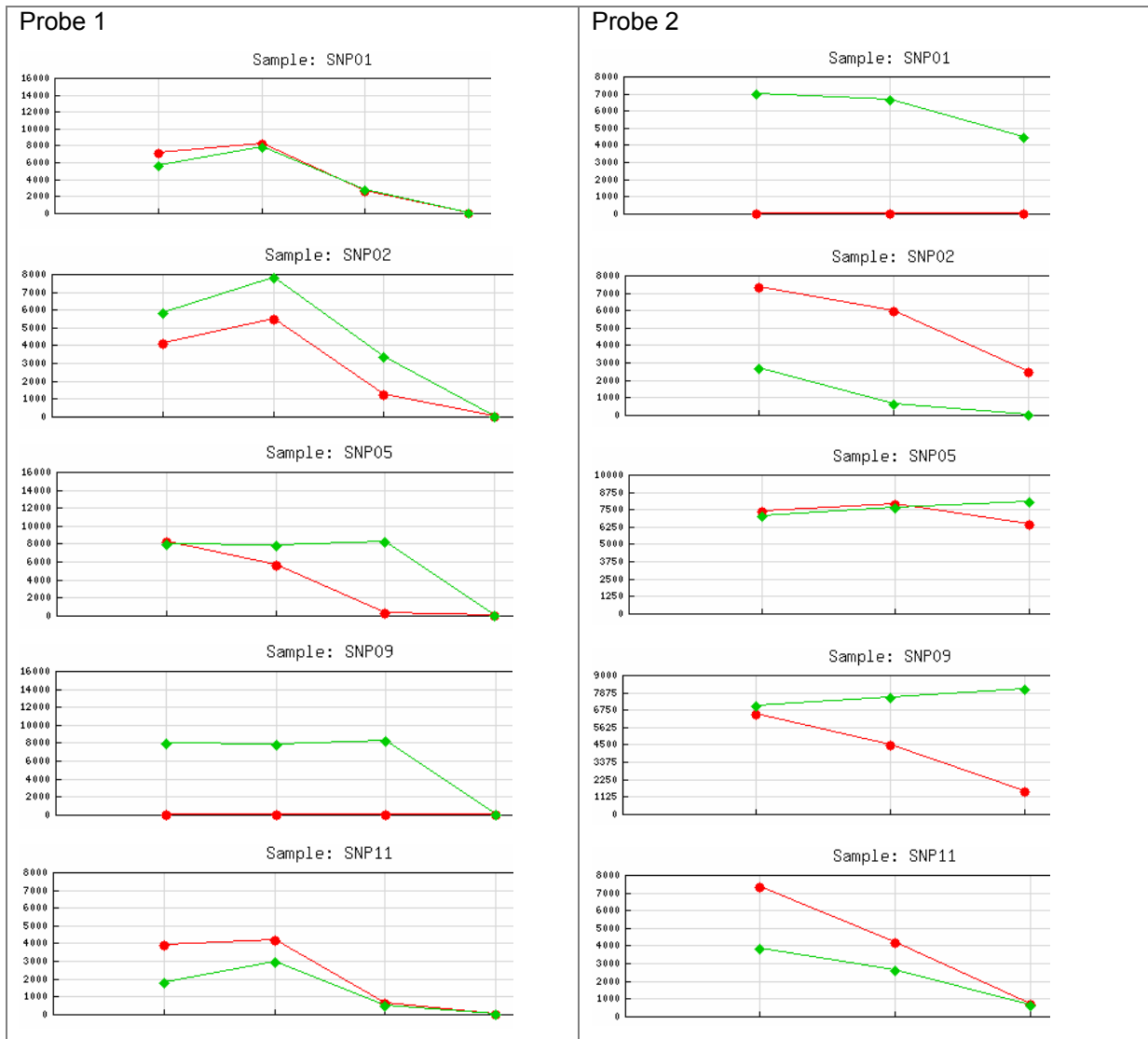


Abbildung 61: Die Trendgrafiken zeigen die Scanergebnisse der zwei Pentaplex-Proben nach drei thermischen Diskriminierungen (37°C, 38°C und 39°C). Der vierte Scan bei Probe 1 entstand nach einem Strippingschritt. Mittels dieser Grafiken können die Diskriminierungen der einzelnen Reporter bei Temperaturerhöhung verfolgt werden.

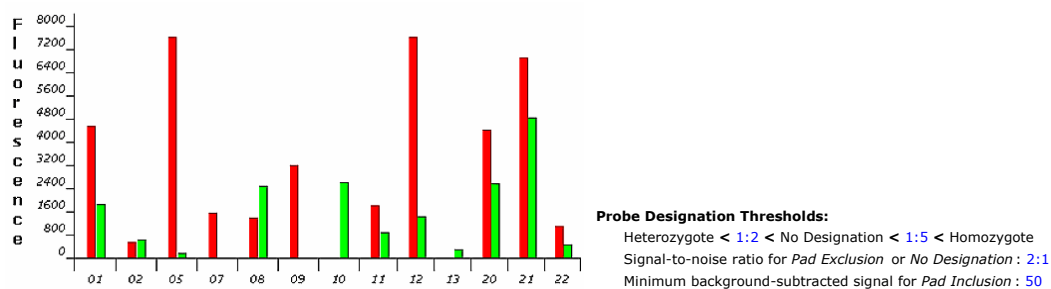


Abbildung 62: Simultane Genotypisierung von 13 SNPs. Die passive Hybridisierung erfolgte in zwei Schritten, erstens wurden alle Diskriminatoren für die 13 SNPs hybridisiert und nach einem HSB-Waschgang wurden die zwei Universal reporter passiv hybridisiert. Auf der X-Achse sind die Namen der SNPs angegeben. Daneben sind die Einstellungen bei der automatischen Auswertung aufgezeigt.

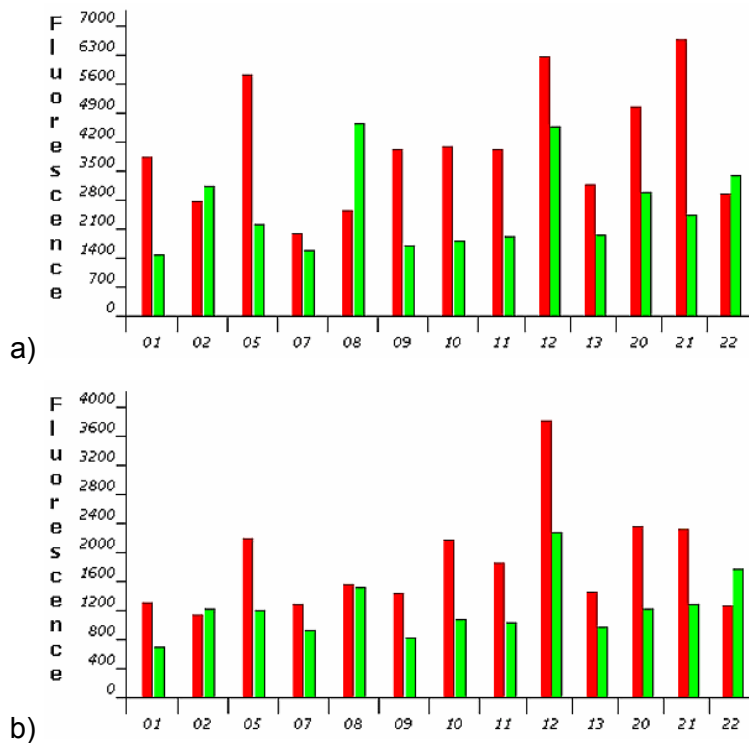


Abbildung 63: Simultane Genotypisierung von 13 SNPs. Die Hybridisierung der Diskriminatoren erfolgte elektronisch in 60 Sekunden mit Hilfe des Loaders. Anschließend wurden die zwei Universal reporter passiv hybridisiert. Das obere Histogramm a) gibt die Fluoreszenzsignale nach 38°C thermischer Diskriminierung wieder und das Histogramm b) nach 41°C.

Das beste Ergebnis konnte mit der dritten Alternative erzielt werden, indem die passive Hybridisierung in einem Schritt erfolgte (Abbildung 64). Dabei war es von großer Bedeutung, die Diskriminatoren und die Universal reporter in HSB vorher für 10 Minuten bei Raumtemperatur zu inkubieren. Bei der Inkubation machte es keinen Unterschied, ob alle Oligos insgesamt oder nur paarweise gemischt wurden. Zusätzlich zeigt Abbildung 64 die Konkordanz des Microarray-Ergebnisses und der SNaPshot-Typisierung dieser Probe.

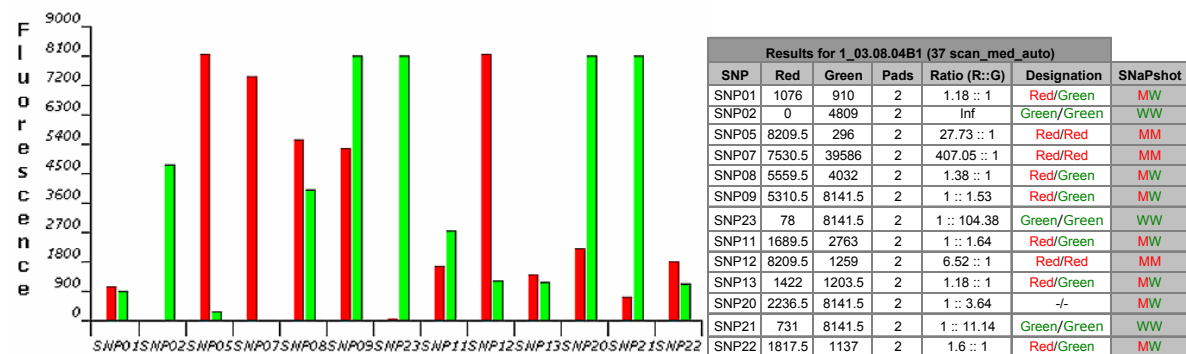


Abbildung 64: Simultane Genotypisierung von 13 SNPs. Die Hybridisierung der Diskriminatoren für alle SNPs und der Universal reporter erfolgte passiv in einem Schritt. Die Rohdatentabelle ist mit den SNaPshot ermittelten Genotypen (Grau unterlegte Spalte) ergänzt. Damit kann die Konkordanz zwischen Nanogen und SNaPshot-Genotypisierung bestimmt werden.

Abbildung 65 zeigt die Diskriminierungseigenschaften einiger SNPs für zwei verschiedene Proben. Diese beispielhaft aufgeführten Trendgrafiken bestätigen die Feststellung, dass die einzelnen SNPs bei verschiedenen Temperaturen diskriminieren und daher individuell zu behandeln und zu optimieren sind. Während einige SNPs nur bei hohen Temperaturen, wie 39°C diskriminieren (z.B. Probe 2 SNP08, SNP09, SNP21), fallen andere bei dieser Temperatur aus (z.B. Probe 1 SNP01, SNP07, SNP11) (Abbildung 65).

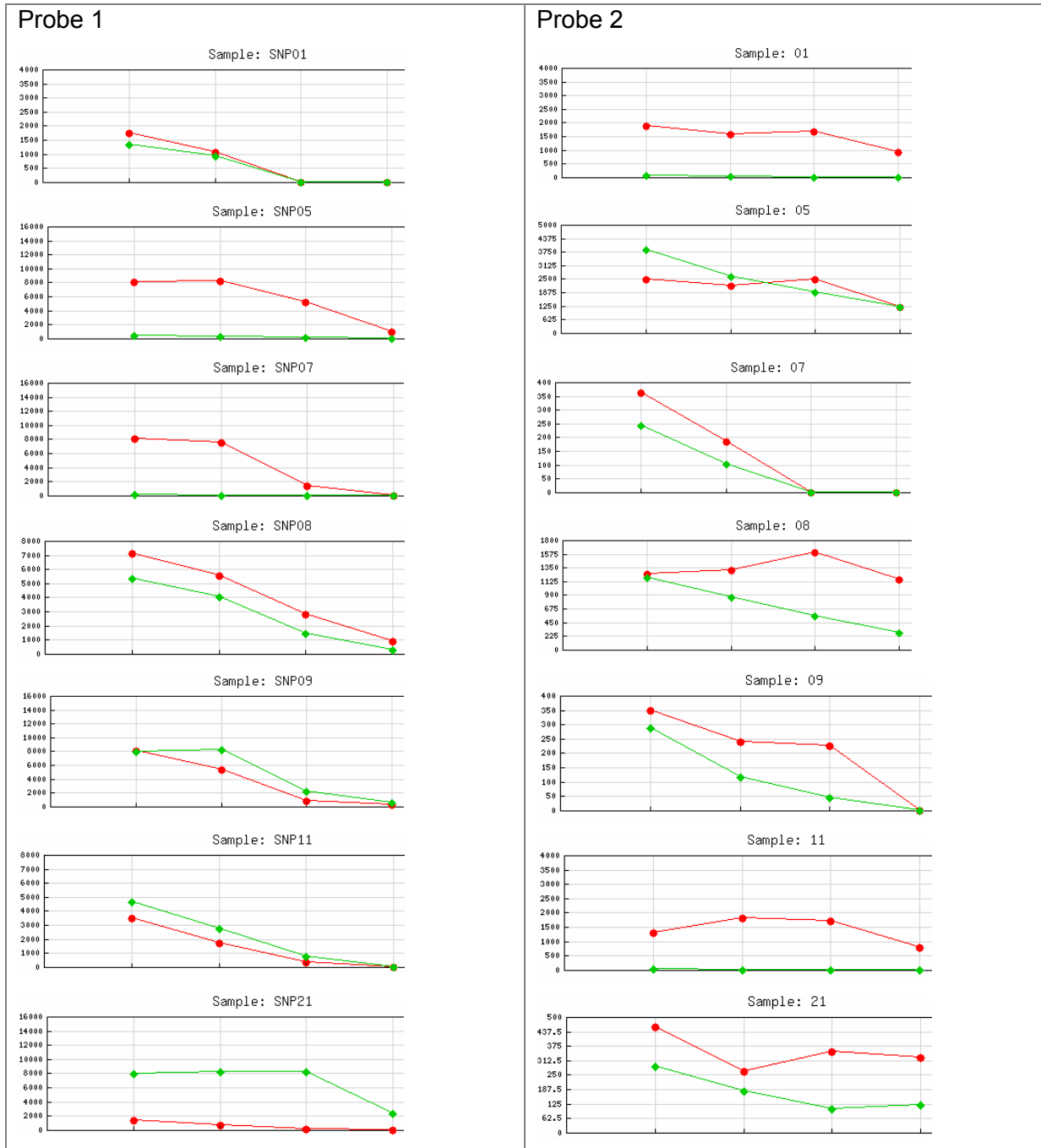


Abbildung 65: Die Trendgrafiken zeigen Scanergebnisse von zwei 13plex-Proben nach vier thermischen Diskriminierungen (Abszisse 37°C, 38°C, 39°C und 40°C). Mittels dieser Grafiken können die Diskriminierungen der einzelnen Reporter bei Temperaturerhöhung verfolgt werden.



Insbesondere SNP08 und SNP09 stellen die Anforderung, die Eigenschaften beider Diskriminatoren aufeinander abzustimmen. Eine homozygote Probe ist für SNP08 bei 40°C (Probe 2) bzw. für SNP09 bei 39°C (Probe 2), wo aber eine heterozygote Probe ausfallen würde (Probe 1), zuverlässig zu detektieren (Abbildung 65). Dagegen zeigt SNP05 ein positives Beispiel, denn beide Proben mit unterschiedlichen Genotypen (Mutante für Probe 1 und Heterozygot für Probe 2) können bei 38°C und 39°C mit großer Zuverlässigkeit erfasst werden (Abbildung 65). Nachdem die Auswirkung einer Normalisierung durch eine Heterozygotenkontrolle bei dem Amplicon down-Assay positiv bewertet wurde (Abbildung 49 in Kapitel 3.3.1), folgten mehrere Experimente zur Verbesserung der Ergebnisse des Capture down-Assays mittels Heterozygotenkontrolle. Für die Pentaplex-Probe in Abbildung 66b wurde bei der Auswertung SNP05 als Heterozygotenkontrolle definiert, wonach der Ausgleich der Farbintensitäten automatisch für alle SNPs erfolgte. Abbildung 66, welche die gleiche Probe ohne (a) bzw. mit Heterozygotenkontrolle (b) zeigt, macht die Wirkung einer Kontrolle deutlich.

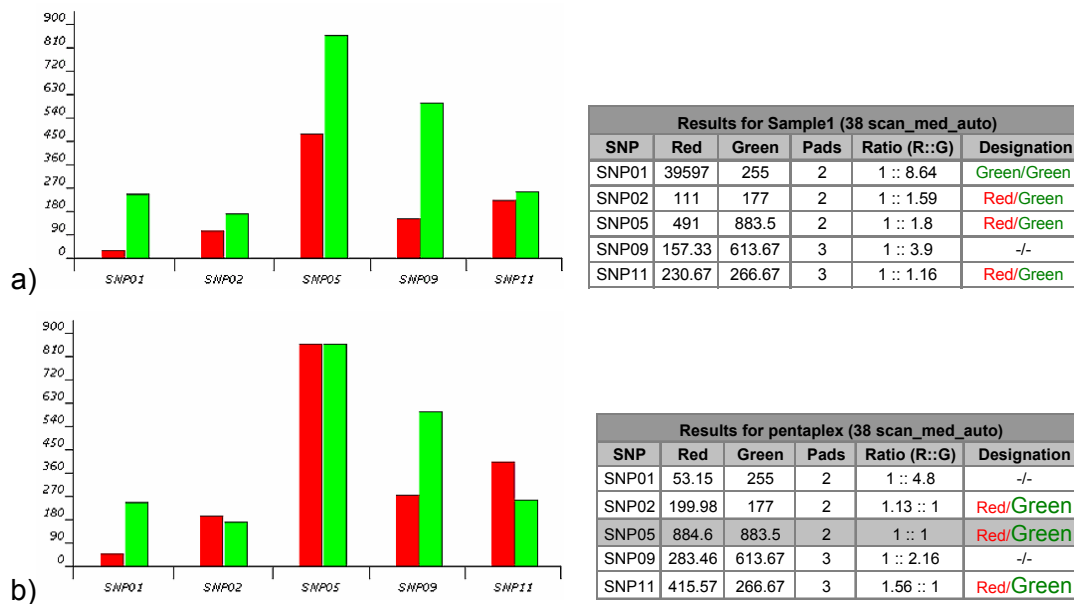


Abbildung 66: Beide Histogramme stellen den gleichen Scan vor, mit dem Unterschied, dass für Histogramm a) keine Heterozygotenkontrolle zur Normalisierung gewählt wurde, während für b) SNP05 als Heterozygot definiert wurde. Aus den Rohdatentabellen können die Auswirkungen einer Heterozygotenkontrolle abgelesen werden.

Die Abweichungen zwischen den Diskriminierungseigenschaften von zwei SNPs beeinflussen auch die Normalisierung, mit der Folge, dass SNP01, der eindeutig homozygot ist, nicht als solcher automatisch typisiert werden kann (Abbildung 66b). Aus diesem Grund wurde eine Heterozygotenkontrolle geschaffen, welche keinen diskriminatorischen Abläufen ausgesetzt ist. Die Komplementärsequenzen zu den Universal reporter wurden elektronisch auf einem Pad adressiert und waren damit bereit für eine direkte Hybridisierung der Universal Reporter während des Readerprotokolls. Abbildung 67 ist als Beispiel für mehrere Experimente mit so einer künstlich hergestellten Heterozygotenkontrolle zu betrachten. Während für SNP05 und SNP09 durch den Farbausgleich eine sichere Genotypbestimmung

erfolgen konnte, waren andere SNPs, wie SNP02, SNP10 und SNP21 nicht mehr typisierbar (Abbildung 67b). Obwohl hier eine Kontrolle benutzt wurde, welche keine Diskriminierung durchläuft, zeigt sich die Notwendigkeit einer spezifischen Heterozygotenkontrolle für jeden einzelnen SNP. Dieses ist bei einer komplexen Multiplex-Analyse aus praktischen Gründen nicht durchführbar.

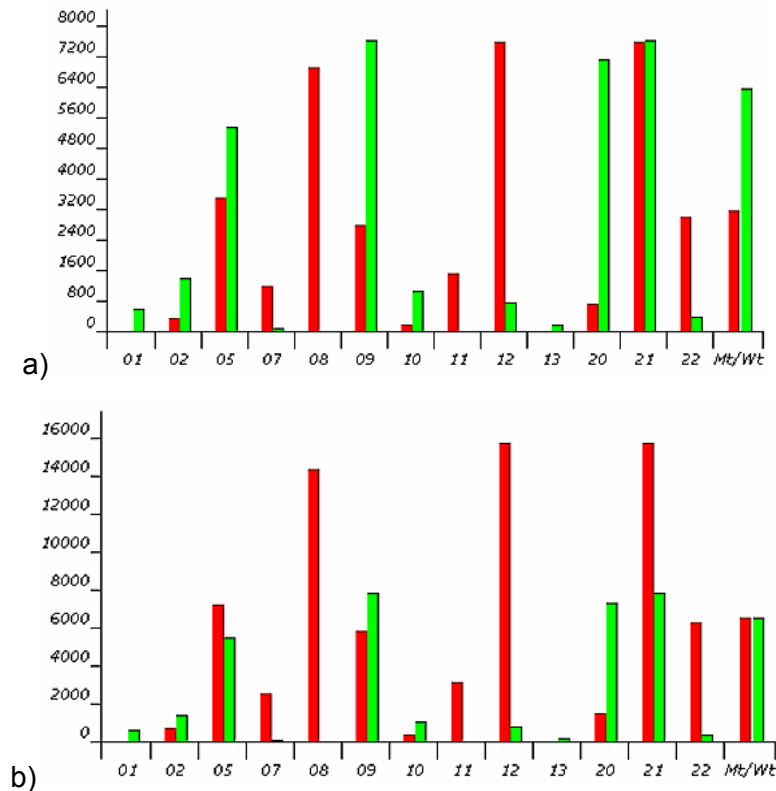


Abbildung 67: Beide Histogramme stellen den gleichen Scan vor, mit dem Unterschied, dass für Histogramm a) keine Heterozygotenkontrolle zur Normalisierung gewählt wurde, während für b) Mt/Wt als Heterozygot definiert wurde. Mt/Wt entspricht einem Pad, wo bei der elektronischen Adressierung zwei Komplementärsequenzen zu den Universal reporter in gleichem Verhältnis geladen wurden.

Als nächstes werden drei Beobachtungen beschrieben, die sich während der Etablierung des Capture down-Assays gezeigt haben und eine Auswirkung auf die weiterführenden Analysen hatten. Obwohl die Vergleichsexperimente der Aufreinigungsmethoden mittels Amplicon down-Assay ergaben, dass die Effizienz der Millipore-Methode der Invitex-Methode gleich zu setzen ist (Abbildung 50 in Kapitel 3.3.1), konnte dies nicht auf das Capture down-Assay übertragen werden. Die Aufreinigung der Singleplex-, Duplex- und Pentaplex-Proben mittels Millipore führte zum totalen Produktverlust, während Invitex problemlos funktionierte. Die Millipore-Aufreinigung der 13plex-Proben wiederum lieferte auswertbare Produkte, wobei im Vergleich zur Invitex-Aufreinigung die SNPs mit geringerer Fragmentlänge an Intensität abnahmen (Abbildung 68). Dies ist darauf zurückzuführen, dass die PCR-Produkte unterhalb 90bp während der Millipore-Aufreinigung verloren gehen, z.B. SNP02 83bp, SNP05 86bp, SNP21 81bp.

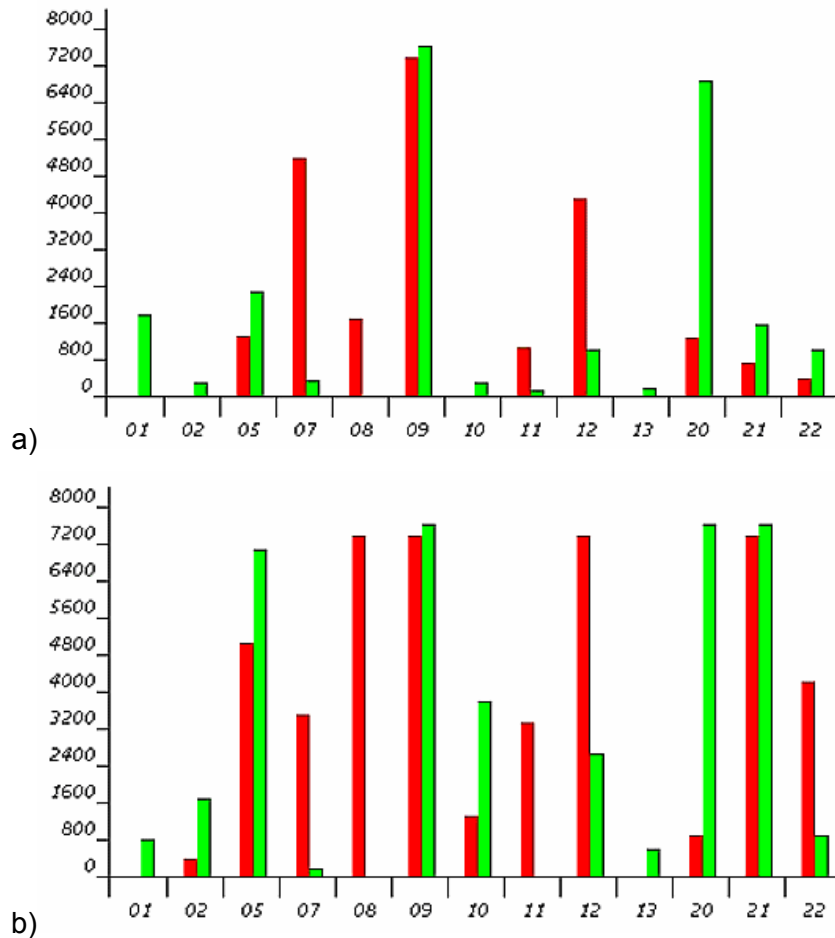


Abbildung 68: Beide Histogramme stellen die Ergebnisse der gleichen 13plex-Probe dar, mit dem Unterschied, dass die Probe in Histogramm a) mittels Millipore-System aufgereinigt wurde während die Probe in Histogramm b) mittels Invitex-Methode aufgereinigt wurde.

Abbildung 69 zeigt ein Beispiel für ein erfolgreiches Stripping. Nachdem die erste Pentaplex-Probe typisiert wurde (Abbildung 69a), folgte das Stripping in drei Schritten. Als Erstes wurde die Temperatur auf 54°C erhöht und das Cartridge mit LSB gewaschen. Als Zweites wurde das Microarray manuell mit NaOH 0,1N gewaschen. Drittens erfolgte ein elektronisches Stripping, indem die Polarität der Pads negativ gestellt wurde. Nach dem Stripping konnten die fünf Capture-Oligos für weitere PCR-Hybridisierungen benutzt werden. Für die zweite PCR-Hybridisierung (Abbildung 69b) wurde eine Probe verwendet, die unterschiedliche Genotypen besitzt im Vergleich zu der ersten PCR-Hybridisierung (Abbildung 69a). Die zweite PCR-Hybridisierung erfolgte parallel auf Pads mit dem gestrippten Capture- (Abbildung 69b) und mit neu adressierten Capture-Oligos (Abbildung 69c), damit konnten die Ergebnisse direkt verglichen werden. Die Ergebnisse der zweiten PCR-Hybridisierung zeigen keine störenden Kontaminationen durch die erste PCR, z.B. ist für SNP01 das erste Produkt heterozygot und die zweite Probe homozygot (Abbildung 69a und b). Ebenso zeigt dieses Experiment keine erheblichen Abweichungen zwischen der PCR-Hybridisierung auf gestripptem Capture- und neu adressiertem Capture-Oligo (Abbildung 69b und c).

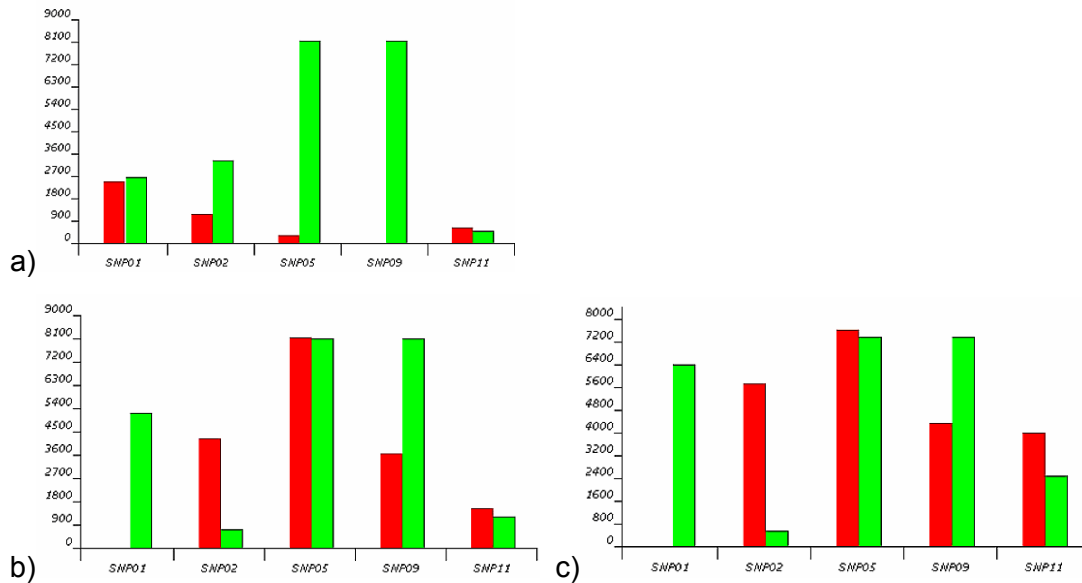


Abbildung 69: Histogramm a) zeigt die Genotypen einer Pentaplex-Probe. Nach dem Scannen wurden die PCR-Produkte gestriipt und die Capture-Oligos für eine zweite PCR-Hybridisierung verwendet (b). Parallel wurden bei der zweiten PCR-Hybridisierung auch Pads beladen, die unbenutzte Capture enthielten (c).

Abbildung 70 weist auf ein Problem hin, das entlang des ganzen Prozesses der Etablierung, Optimierung und Validierung des Capture down-Assays mittels Nanogen-Technologie immer wieder aufgetaucht ist. Es wurden Fehler beim sog. Iscan entdeckt, bei dem sich die Diskriminatoren auf dem Cartridge befinden, aber noch keine thermische Diskriminierung erfolgte. Für Abbildung 70a wurden die Reihen 4 bis 6 komplett gescannt, trotzdem wurden in den Spalten 5 bis 10 nicht vollständige oder keine Signale detektiert. Das gleiche Phänomen zeigte sich, nachdem die Proben gestriipt, neu adressiert und hybridisiert wurden (Abbildung 70b).

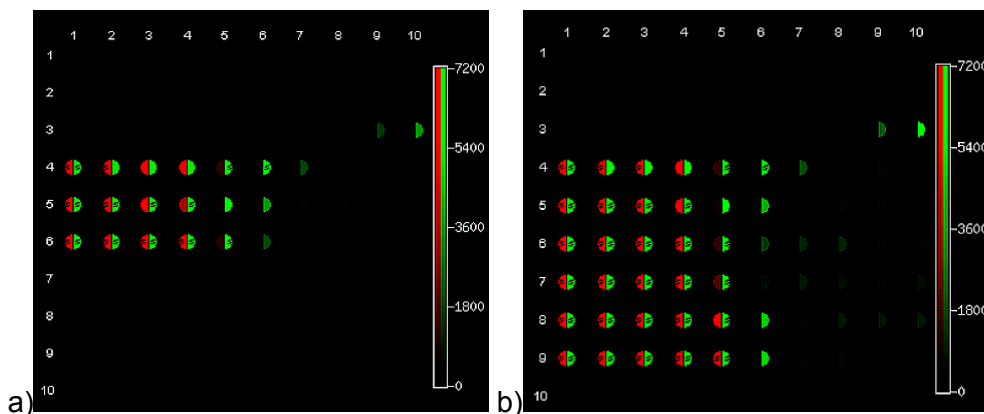


Abbildung 70: Das Bild a) zeigt einen Scan nachdem die Diskriminatoren und Universal reporter Hybridisierung stattfand aber noch kein Readerprotokoll gelaufen ist. Gescannt wurden die Pads 3,9 und 3,10 und die Zeilen 4, 5 und 6. Das Bild b) zeigt den gleichen Mikrochip nachdem gestriipt und gewaschen wurde und eine neue Hybridisierung stattfand. Gescannt wurden die Pads 3,9 und 3,10 und die Reihe 4 bis 9.

Im Laufe des Projektes konnten verschiedene Ursachen ausgemacht werden, die zu solchen Fehlern geführt haben. Eine der häufigsten Ursachen waren Luftblasen, die sich während der automatischen Beladung des Cartridges auf der Mikrochip-Oberfläche befanden und die Hybridisierung verhinderten. Das Luftblasenproblem konnte dadurch beseitigt werden, dass im Waschpuffer kein schäumendes Triton mehr verwendet wurde. Außerdem waren einige Cartridges beschädigt oder nicht voll funktionsfähig geliefert. Darüber hinaus wurde beobachtet, dass elektronische Aktivierungen den Hydrogel Layer des Cartridges beschädigen können, sodass z.B. wiederholte elektronischen Strippings die Beschaffenheit und damit die Eigenschaft des Nanochips beeinträchtigen.

Nach der Etablierung des Capture down-Assays für ein Multiplex aus 13 SNPs hatte sich als weiterführende Aufgabe und Herausforderung herauskristallisiert, die individuelle Diskriminierungstemperatur für jeden SNP zu finden, was von ausschlaggebender Notwendigkeit sein sollte.

### 3.3.3.2 Capture down-Assayoptimierung

Nachdem die Etablierung des Capture down-Assays, bei dem die Abläufe der einzelnen Schritte bestimmt wurden, abgeschlossen war, folgten mehrere Experimente zur Assayoptimierung. Die Ergebnisse sind nicht in chronologischer Abfolge zu verstehen, vielmehr sind es Optimierungen einzelner Schritte, die untereinander in Verbindung stehen. Abbildung 71 zeigt die ersten Ergebnisse des etablierten Assays mit einer parallelen Genotypisierung von 24 SNPs. Mit der Expansion des Assays auf 24 SNPs wird die Erfolgsabhängigkeit vom SNP deutlich, so liefern bestimmte SNPs immer zuverlässige Ergebnisse (z.B. SNPs 01, 07, 08, 32), während andere sehr niedrige Signale ergeben, wie z.B. SNPs 24, 25, 26, 37 (Abbildung 71).

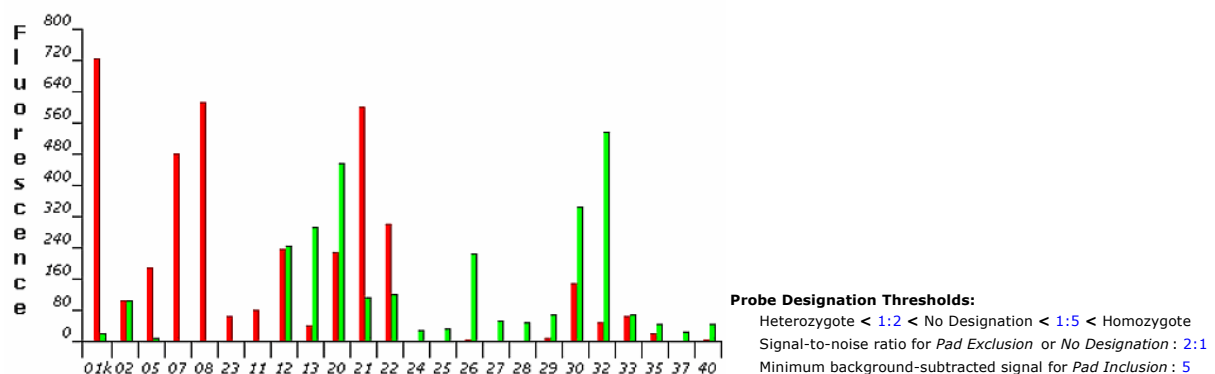


Abbildung 71: Das Histogramm stellt den ersten Versuch der simultanen Genotypisierung von 24 SNPs dar. Auf der X-Achse sind die Namen der SNPs angegeben. Daneben sind die Einstellungen bei der automatischen Auswertung aufgezeigt.

Gleichzeitig wird bei der Betrachtung der Trendgrafiken einzelner SNPs die Abhängigkeit von der Diskriminierungstemperatur und damit auch von dem zu erwartenden Genotyp deutlich. In Abbildung 72 sind die Trends von vier SNPs dargestellt für zwei Personen mit

unterschiedlichen Genotypen und gescannt bei vier verschiedenen Diskriminierungstemperaturen (36°C, 37°C, 38°C und 39°C). Die Diskriminierungstemperaturen, welche nötig sind für eine konkordante Genotypisierung mittels Capture down-Assay, sind grün markiert. Diese vier SNPs repräsentieren die typischen Probleme, die bei dem Diskriminierungsprotokoll auftreten können. Ungleiche Diskriminierungen der zwei Genotypen zeigt SNP20, wo die Diskriminierung der Mutante langsamer erfolgt, sodass ein homozygoter Wildtyp erst bei 39°C detektierbar ist (Probe 1), während der Wildtyp schon bei 36°C diskriminiert (Probe 2, Abbildung 72).

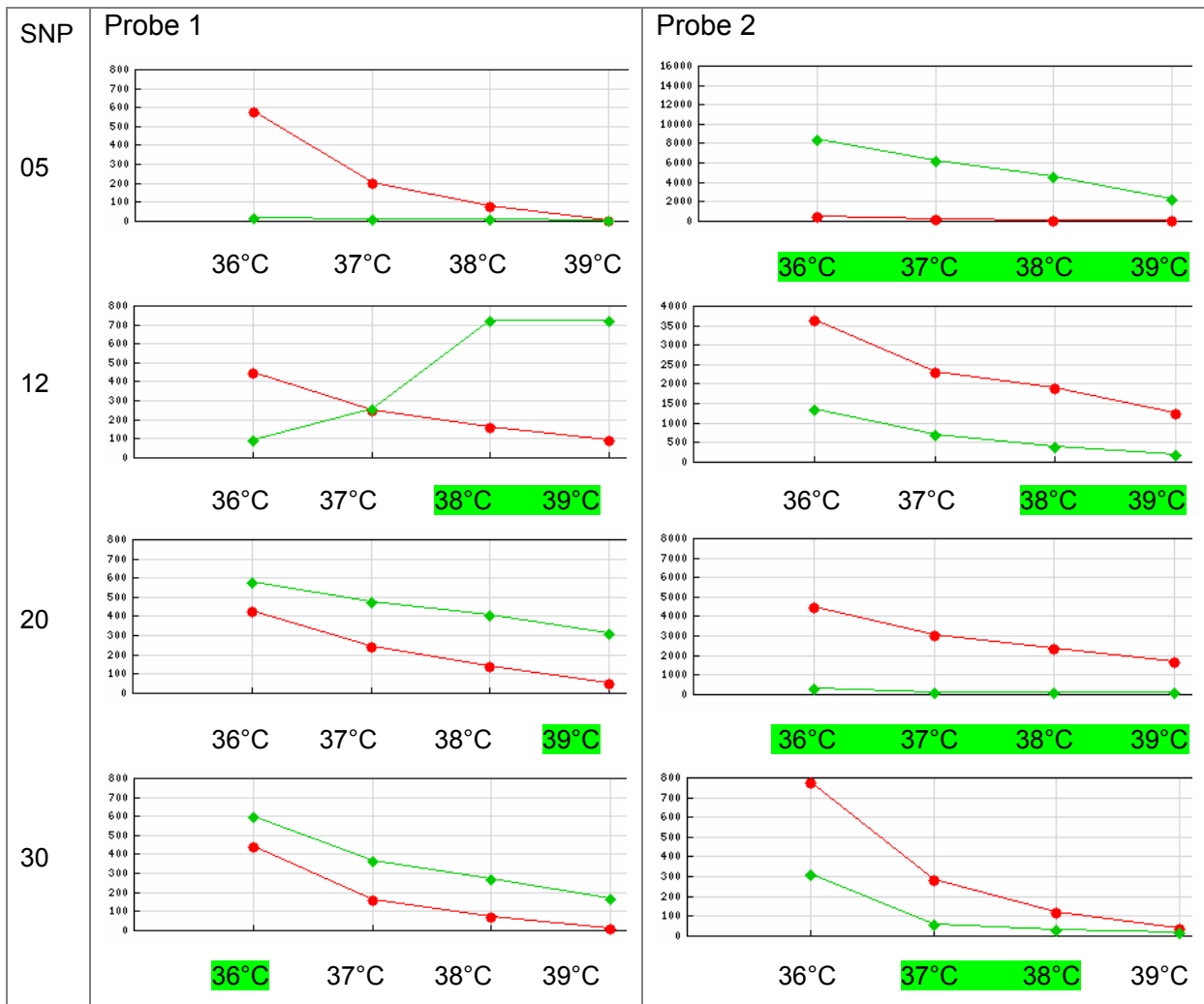


Abbildung 72: Trendgrafiken beispielhaft für vier SNPs (SNP05, SNP12, SNP20 und SNP30) und zwei Proben dargestellt. Die Diskriminierungstemperaturen, die den richtigen Genotyp ergeben, sind jeweils grün markiert.

Die ungleiche Diskriminierung kann schließlich zu falschen Ergebnissen führen, wie die SNPs 05 und 30 zeigen. Für SNP30 kann keine Diskriminierungstemperatur ausgemacht werden, weil diese von dem zu erwartenden Genotyp abhängig ist. So ist eine heterozygote Probe nur bei 36°C zu detektieren (Probe 1), während eine homozygote Probe bei dieser Temperatur keinen zuverlässigen Genotyp ergibt (Probe 2, Abbildung 72). Bei SNP05 diskriminiert der Wildtyp viel schneller als die Mutante, sodass die Detektion einer

heterozygoten Probe nicht möglich ist (Probe 1). Darüber hinaus zeigt das Beispiel von SNP12, wie ausschlaggebend die Ermittlung der richtigen Diskriminierungstemperatur ist. Dieser SNP ergibt für die Probe 1 eine anscheinend korrekte homozygote Mutante bei 36°C und einen heterozygoten Genotyp bei 37°C, dabei erfolgt die richtige Diskriminierung erst bei 39°C (Abbildung 72). Um die Problematik der ungleichen Diskriminierung zu beseitigen, wurden mehrere Experimente zur Assayoptimierung durchgeführt. Als erstes wurden die Konzentrationen der Diskriminatoren entsprechend der ungenauen oder falschen Ergebnisse verändert. Im Falle der vier SNPs aus Abbildung 72 wurden die Konzentrationen der Mutanten-Sonden für die SNPs 05 und 20, der Wildtyp-Sonde für SNP30 und beider Diskriminatoren für SNP12 auf die Hälfte reduziert. Bei SNP05 ist dabei eine eindeutige Verbesserung zu beobachten, indem dadurch die Detektion des heterozygoten Genotyps für die Probe 1 möglich ist (Abbildung 73). Ebenfalls zeigt die balancierte Diskriminierung bei SNP30 eine Verbesserung (Abbildung 73).

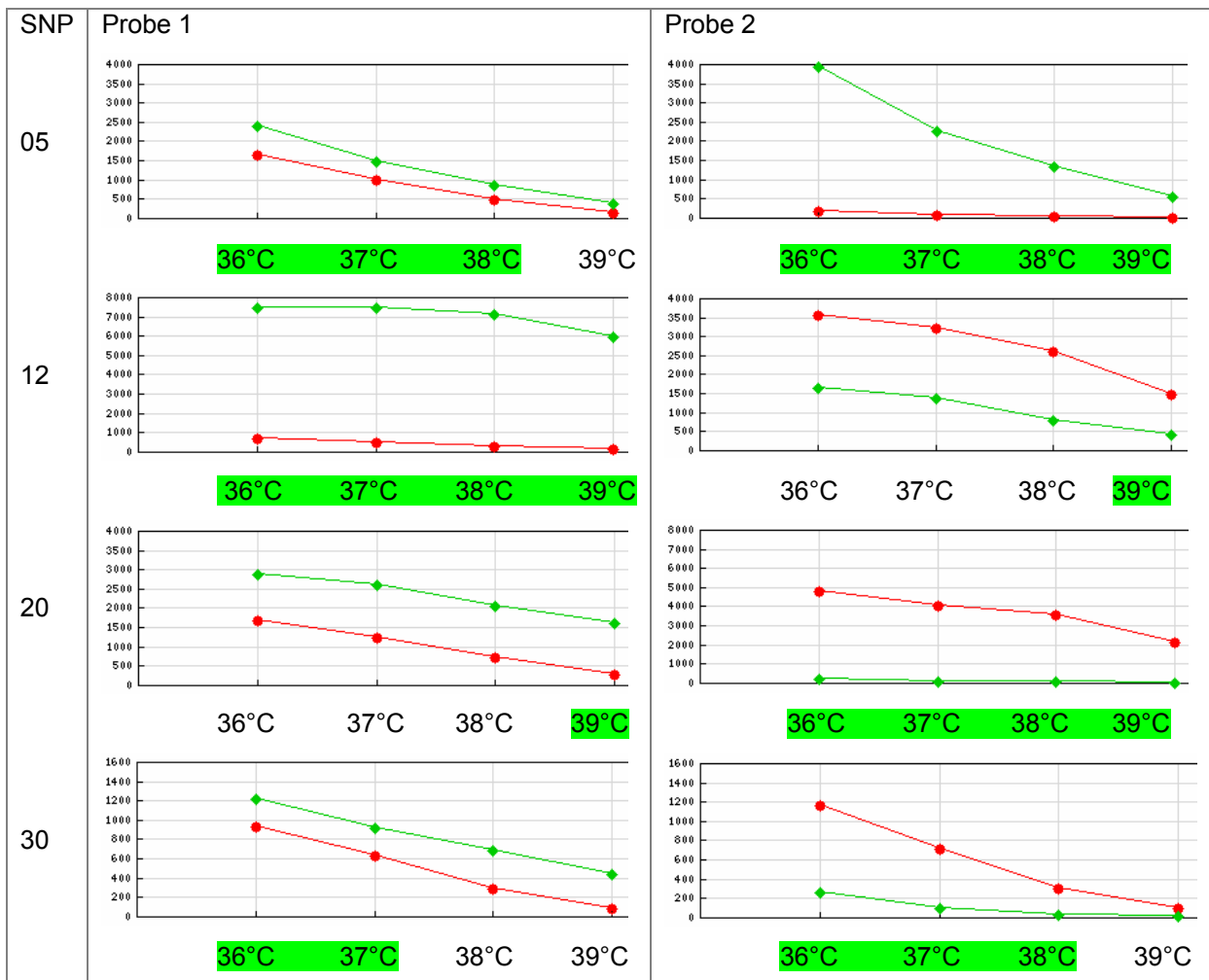


Abbildung 73: Trendgrafiken für vier SNPs und der gleichen zwei Proben wie in Abbildung 72, mit dem Unterschied, dass hier einige Diskriminatoren auf die Hälfte reduziert wurden. Das betrifft folgende Diskriminatoren: Mutante für SNP05 und SNP20, Wildtyp für SNP30 und beide Diskriminatoren für SNP12.

Als zweites Optimierungsexperiment wurden die Konzentrationen der Capture-Oligos für die elektronische Adressierung erhöht. Die zwei Histogramme in Abbildung 74 zeigen die Ergebnisse von unterschiedlich konzentrierten Capture-Oligos. Die Verdoppelung der Capture-Oligos zeigt eine leichte Verstärkung der Signale (z.B. SNPs 33, 35, 37) (Abbildung 74b). Gleichwohl weist aber auch SNP32 eine Verbesserung auf, obwohl die Konzentration von 100nM unverändert blieb.

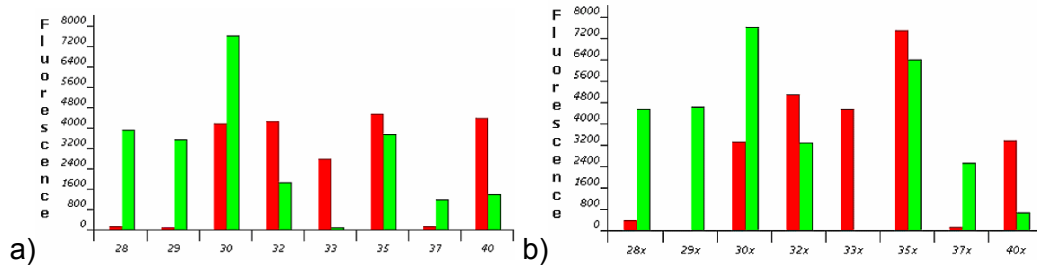


Abbildung 74: Beide Histogramme stellen die Ergebnisse der gleichen Probe dar, nur die Konzentrationen der Capture-Oligos bei der Adressierung sind unterschiedlich. Für Histogramm a) wurden nach Standardprotokoll 100nM Capture adressiert, während die Capture-Konzentrationen für Histogramm b) folgendermaßen erhöht wurden: 150nM für SNPs 28, 30 und 40, 200nM für SNPs 29, 33, 35 und 37, nur für SNP32 blieb die Konzentration bei 100nM.

Zusätzlich wurde ein Touch down-Protokoll getestet, bei dem nur die match Diskriminatoren hybridisieren. Bei dem Touch down-Protokoll ist nur ein Scan vorgesehen, wodurch eine Fehldetektion bei falsch gewählter Temperatur ausgeschlossen wird. Die Ergebnisse der ersten Experimente weisen hohe Backgroundwerte auf, welche zu Ausfällen von vielen SNPs führten (Abbildung 75). Leichte Verbesserung konnte dadurch erreicht werden, dass erst die Diskriminatoren auf das Microarray gegeben wurden, gefolgt von einem Touch down-Protokoll und erst danach wurden die zwei Universal reporter geladen, gefolgt von einem Diskriminierungsprotokoll (Abbildung 75b). Durch weitere Optimierungsschritte, wie stringenteres Waschen, Wahl eines geeigneten Temperaturgradienten und Kombination des Touch down-Protokolls mit dem Diskriminierungsprotokoll, stand schließlich ein optimiertes Readerprotokoll zur Verfügung.

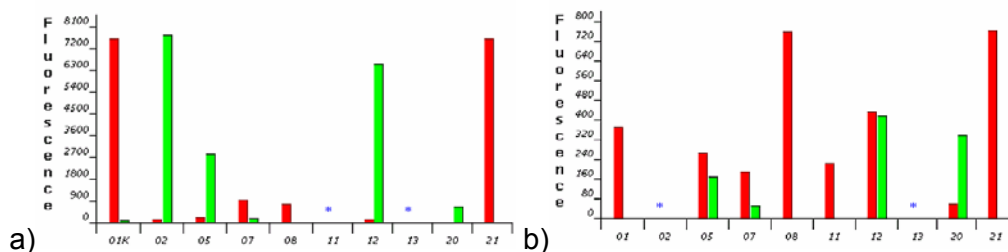


Abbildung 75: Zwei typische Histogramme nach einem sogenannten Touch down-Readerprotokoll. Bei Histogramm a) wurden die Diskriminatoren und Universal reporter gemischt auf dem Microarray gegeben und mittels Touch down-Protokoll diskriminiert. Bei Histogramm b) wurden erst die Diskriminatoren geladen gefolgt von einem Touch down-Protokoll und anschließend die Universal reporter auf dem Microarray gegeben und mittels Diskriminierungsprotokoll behandelt.



Das Readerprotokoll schließlich ist vereinfacht, indem alle Diskriminatoren und die zwei Universal reporter in einem Gemisch auf das Microarray geladen werden und das Touch down-kombiniert mit dem Diskriminierungsprotokoll ohne manuelle Zwischenschritte in einem automatischen Protokoll ablaufen kann (Tabelle 18 in Kapitel 2.2.6.3.2). Durch das optimierte Readerprotokoll konnte die Problematik der ungleichen Diskriminierung und Fehldetektion eliminiert werden, jedoch konnte dadurch nicht die Balance der SNPs innerhalb eines Multiplex hergestellt werden. Aus diesem Grund folgten einige Experimente, die die Aufgabe hatten, die Ausgewogenheit der PCR-Produkte in einem Multiplex zu verbessern. Zunächst wurden die Ergebnisse einer Standard-PCR mit einer asymmetrischen PCR verglichen, in der der Targetstrang im Verhältnis 5:1 und 10:1 amplifiziert oder reamplifiziert wurde (Abbildung 76). Während für SNP02 eine eindeutige Verbesserung mittels asymmetrischer PCR zu verzeichnen war (Abbildung 76b und Abbildung 77), zeigten andere SNPs sogar eine Abnahme in der Signalstärke (Abbildung 76 und Abbildung 77).

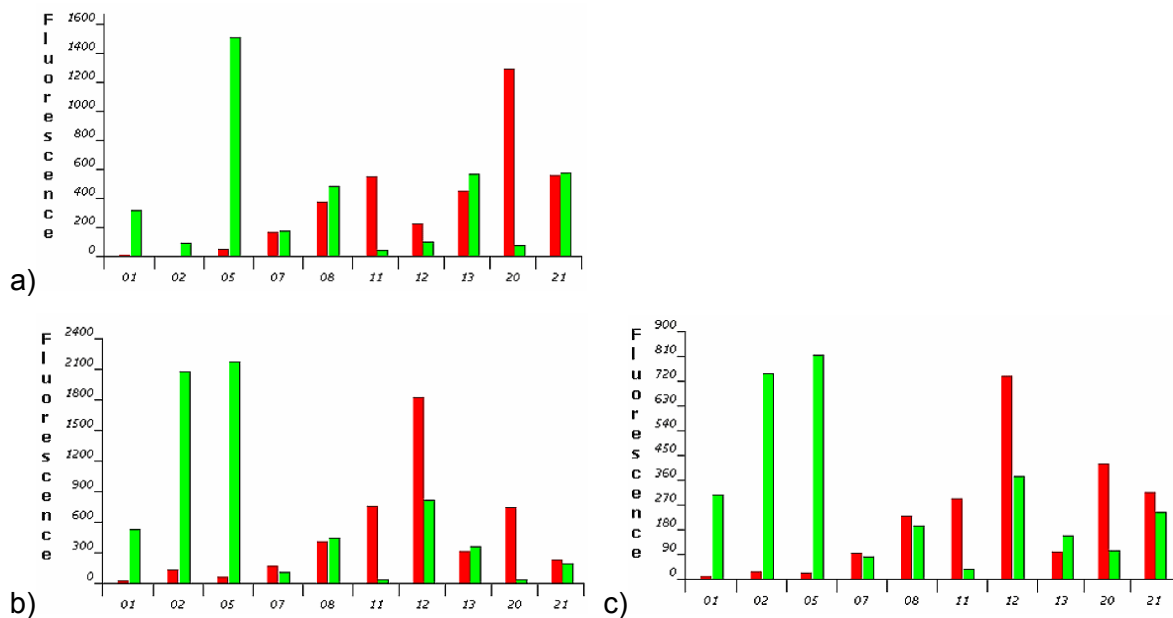


Abbildung 76: Vergleich der verschiedenen PCR-Strategien. Histogramm a) stellt eine Probe dar, die nach Standardprotokoll amplifiziert wurde. Für Histogramm b) wurde eine 5:1 und für Histogramm c) eine 10:1 asymmetrische Amplifikation durchgeführt.

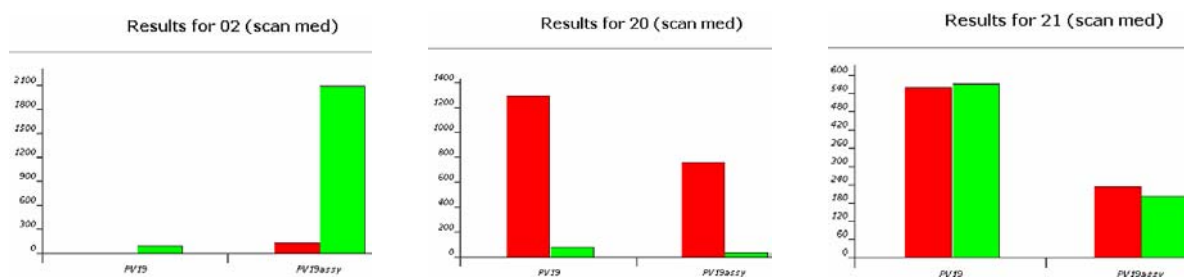


Abbildung 77: Die drei Histogramme verdeutlichen den Unterschied zwischen einer Standard-PCR (erste Signalpaar) und 5:1 asymmetrischer Amplifikation (zweite Signalpaar) für die SNPs 02, 20 und 21.

Weiterhin führten PCR-Zusätze wie Glycerin oder BSA nicht zu einer Ergebnisverbesserung (Abbildung 78). Die Strangtrennung nach der PCR mit Hilfe von Biotin führte zu einem Produktausfall von über 50%, was auf wenig Targetstrang und sehr hohen Background zurückzuführen war (Daten nicht gezeigt).

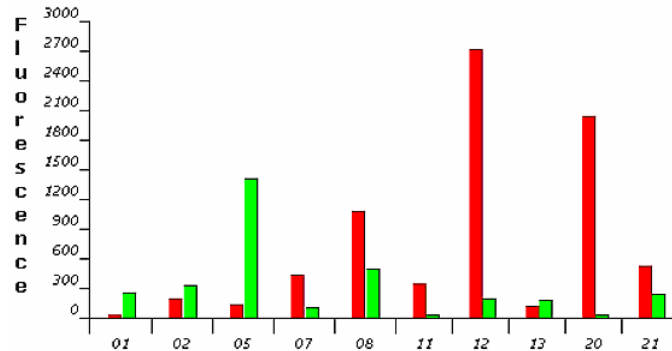


Abbildung 78: Ergebnisse einer PCR-Amplifikation mit Glycerin und BSA Zusätzen.

Eine sichtbare Ergebnisverbesserung und Signalverstärkung bewirkte die Verdoppelung der PCR-Primer (Abbildung 79). Dieses Ergebnis brachte den Beweis, dass die Signalstärken einzelner SNPs, durch eine PCR-Primer Optimierung gezielt verbessert werden können. Mit dieser Methode wurden zu einem späteren Zeitpunkt die PCR-Primerkonzentrationen für jeden validierten SNP festgelegt (Tabelle 9 in Kapitel 2.2.4.4).

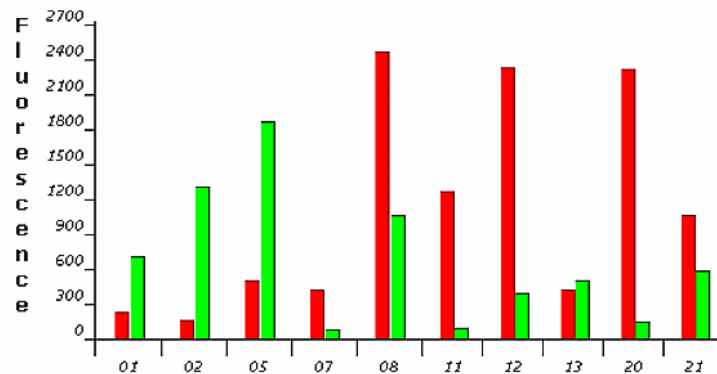


Abbildung 79: Ergebnisse einer PCR-Amplifikation mit verdoppelten PCR-Primern.

Als nächstes folgten vergleichende Experimente zur elektronischen PCR-Produkt-Hybridisierung. Dafür wurde stets Arginin [100mM] als Puffer verwendet mit Einsatz von unterschiedlichen PCR-Produkten. Abbildung 80 stellt Vergleiche zwischen 50µl Invitex gereinigtem (Abbildung 80a), zu 10µl ungereinigtem (Abbildung 80b) und zu 10µl enzymatisch gereinigtem Produkt (Abbildung 80c) dar. Dieses Experiment zeigt, dass sich auch Produkte, die für SNaPshot gereinigt wurden, wie auch die ungereinigten PCR-Produkte, unter Verwendung von Arginin [100mM] für den Capture down-Assay eignen.

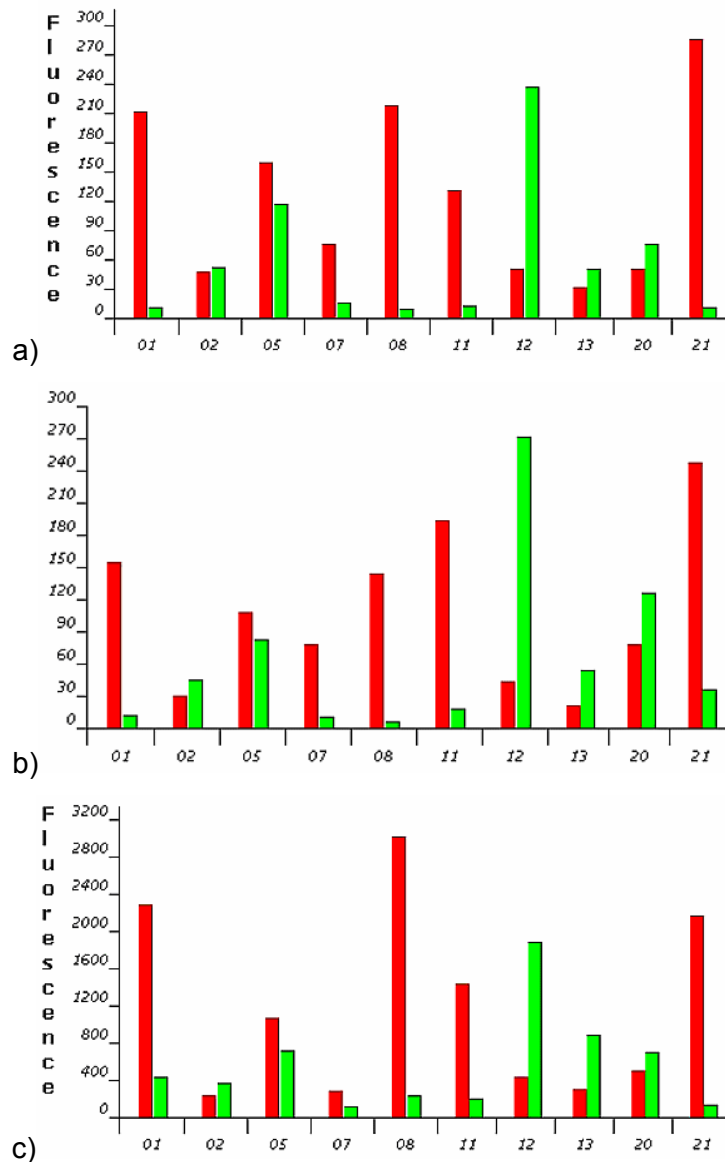


Abbildung 80: Vergleich von verschiedenen PCR-Hybridisierungen. Für Histogramm a) wurden 50µl PCR-Produkte mittels Invitex aufgereinigt und komplett in Arg [100mM] hybridisiert. Für Histogramm b) wurden 10µl ungereinigte PCR-Produkte in Arg [100mM] hybridisiert. Für Histogramm c) wurden 10µl enzymatisch gereinigte Produkte in Arg [100mM] eingesetzt.

Anschließend wurde festgestellt, dass durch Reduzierung des eingesetzten PCR-Volumens die Ergebnisse verbessert werden können (Abbildung 81). Bereits 3µl ungereinigtes PCR-Produkt sind ausreichend für eine zuverlässige Genotypisierung (Abbildung 81b). Das optimale Ergebnis konnte erlangt werden, indem die elektronische Hybridisierung von je 1µl PCR-Produkt dreimal sukzessiv durchgeführt wurde (Abbildung 81c).

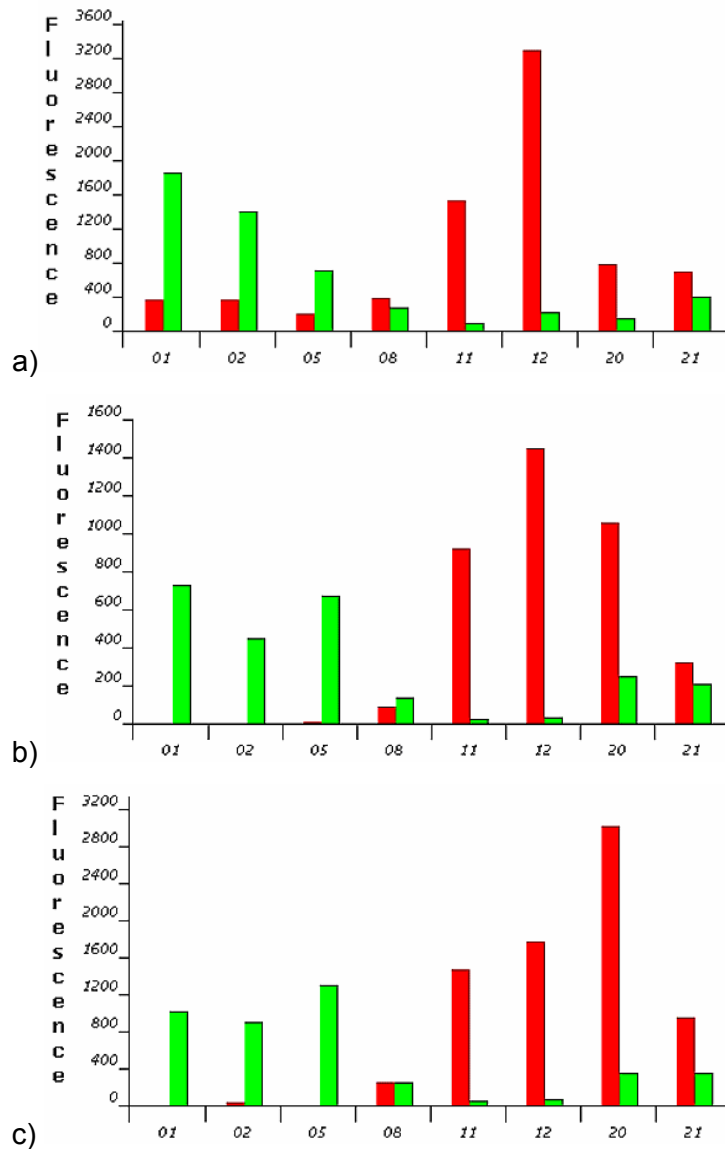


Abbildung 81: Vergleich von verschiedenen Strategien zur PCR-Hybridisierung. Es wurden 10 $\mu$ l (a) oder 3 $\mu$ l ungereinigte PCR-Produkte (b) hybridisiert. Die dreimalige sukzessive Hybridisierung von 1 $\mu$ l ungereinigtem Produkt (c) verbesserte die Ergebnisse.

Während der Optimierungsschritte, insbesondere durch die Einführung von stringenten Waschschritten bei dem Readerprotokoll, wurde beobachtet, dass die nur mit Histidin belegten Background Pads sehr niedrige Signale aufwiesen. Dieses Phänomen bewirkte, dass trotz Subtraktion und Standardeinstellung bei der Auswertung ein falsch positives Signal zurückblieb (Abbildung 82a). Diese Störung kann besonders bei SNPs mit schwachen Signalen zu einem falschen Ergebnis führen (z.B. SNP22 und SNP24 in Abbildung 82a). Ein viel zuverlässigeres und sicheres Ergebnis konnte erlangt werden, indem die Pads als Background definiert wurden, über die bei der Adressierung Histidin und bei der Hybridisierung PCR-Produkte elektronisch geleitet wurden (Abbildung 82b). Abbildung 82 zeigt die Auswertung der gleichen Probe, einmal nur mit Histidin gegenüber Histidin und PCR-Produkt als Background.

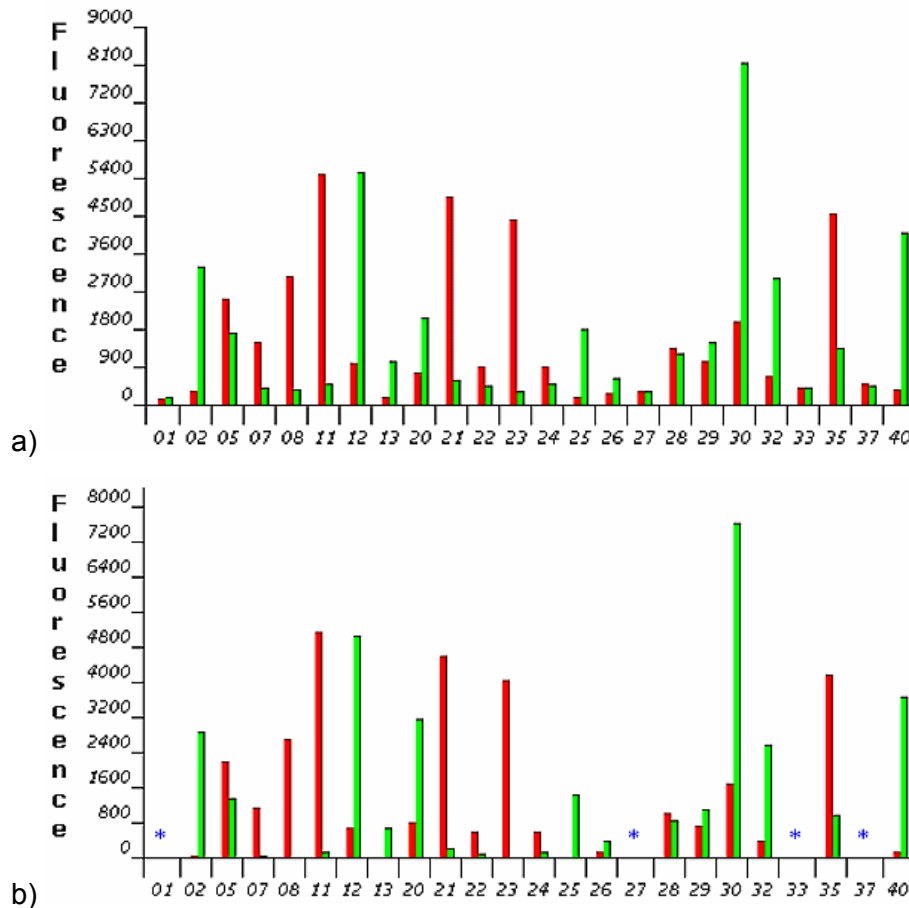


Abbildung 82: Beide Histogramme stellen die Ergebnisse der gleichen Probe dar, mit dem Unterschied, dass für a) ein Pad mit Histidin als Background definiert wurde. Dagegen wurde für Histogramm b) ein Pad mit Histidin und PCR-Produkt als Background gewählt. Die blauen Sternchen deuten auf Ausfälle hin.

Als Nächstes erfolgte die Expansion der Multiplex-Analyse von 24 auf 48 SNPs. Abbildung 83 veranschaulicht die erste parallele Genotypisierung von 48 SNPs. Die Ergebnisse wurden nach ihren zu erwartenden Genotypen sortiert in Abbildung 84. Diese Darstellung gewährleistet eine übersichtliche Validierung einzelner SNPs, indem die homozygoten SNPs, die nicht richtig diskriminieren (z.B. SNPs 19, 21, 36, 41, 46, 15, 25, 26, 32 und 42) und die heterozygoten SNPs, die ungleich diskriminieren (z.B. SNPs 10, 22, 24 und 45), sofort auffallen. Bei den Experimenten mit 48 SNPs wurde die Notwendigkeit eines stringenten Waschschriffs weiterhin deutlich. In Abbildung 85 sind Beispiele für die richtige Bestimmung durch zusätzliche Waschschriffe aufgeführt. Dem korrekten Genotyp entsprechende Ergebnisse lieferte bereits der erste Scan nach 41°C thermischer Diskriminierung und achtmal waschen mit HSW, was durch viermal waschen mit LSW verbessert wurde (Abbildung 85). Damit konnten durch stringenteres Waschen, in diesem Fall von 15 falsch detektierten SNPs, 5 SNPs richtig bestimmt werden.

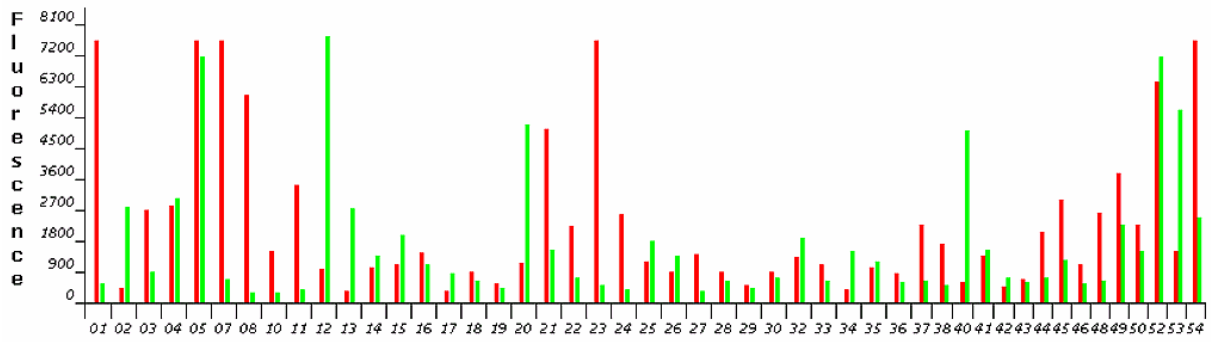


Abbildung 83: Das Histogramm stellt den ersten Versuch der simultanen Genotypisierung von 48 SNPs dar. Auf der X-Achse sind die Namen der SNPs angegeben.

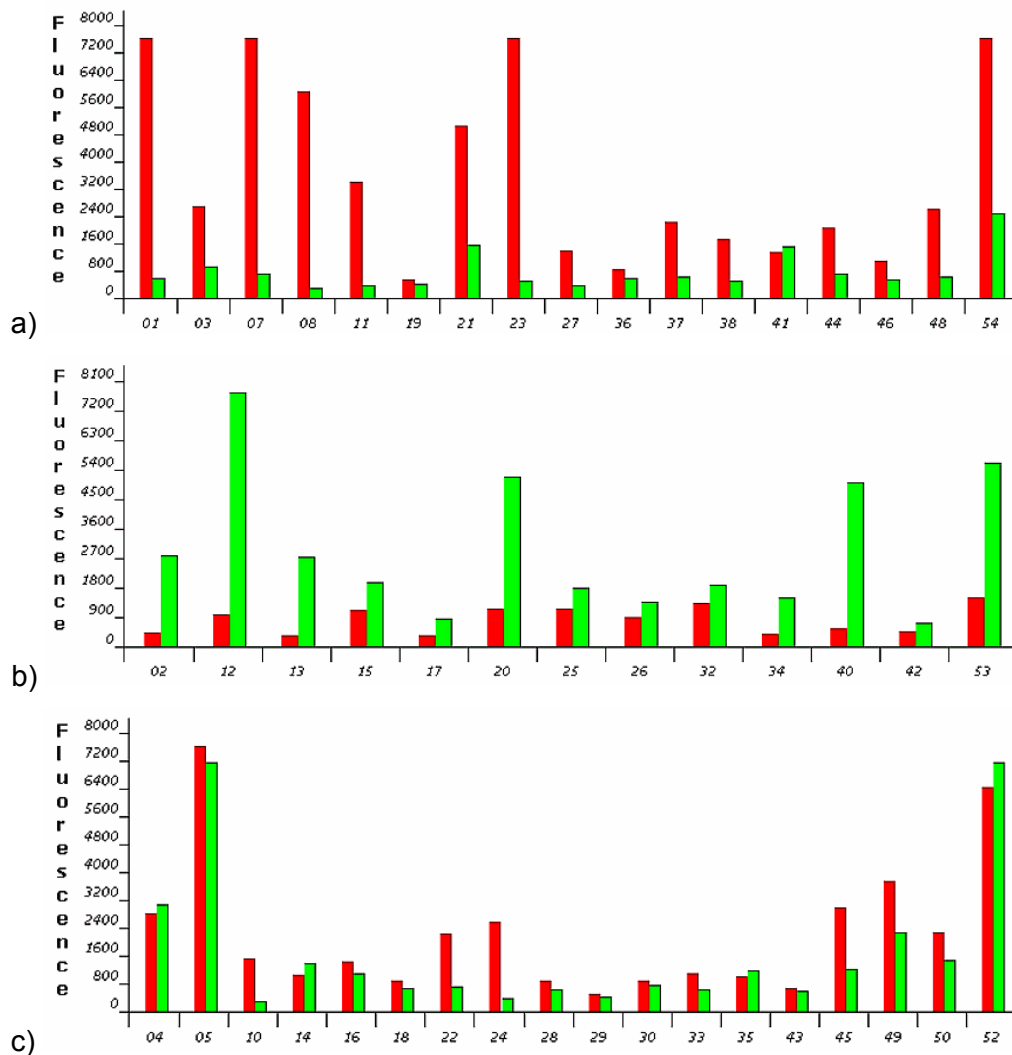


Abbildung 84: Simultane Genotypisierung von 48 SNPs geordnet nach den Genotypen ermittelt mit SNaPshot-Referenzmethode. Die drei Histogramme teilen die SNPs in Mutante (a), Wildtyp (b) und Heterozygote (c) auf.

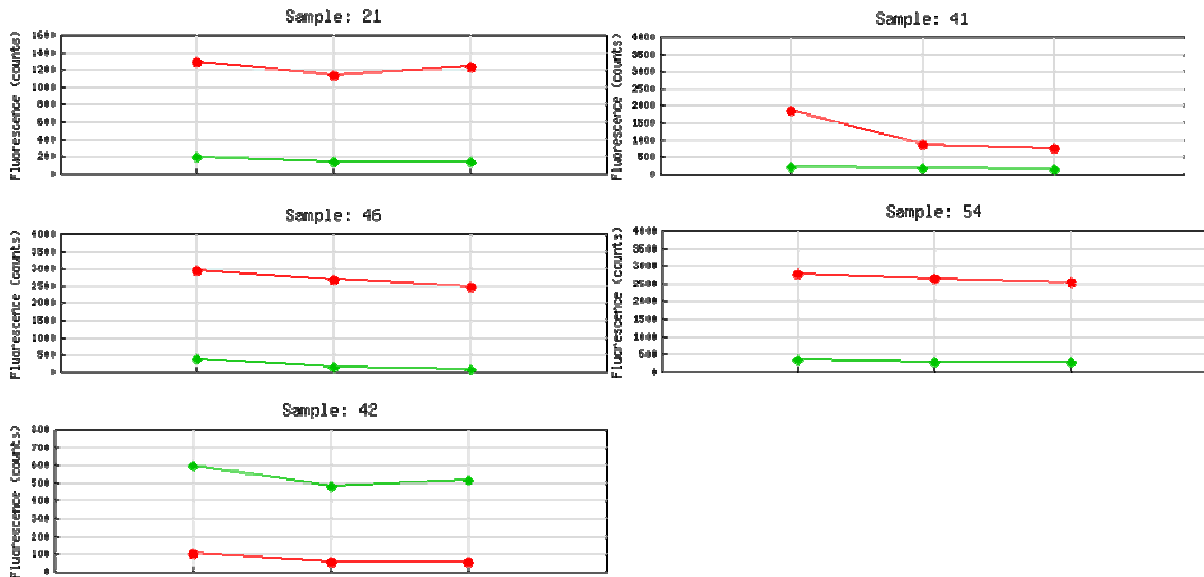


Abbildung 85: Nach stringentem Waschen konnten einige der falsch typisierten Mutanten (SNP21, SNP41, SNP46 und SNP54 in Abbildung 84a) und Wildtypen (SNP42 in Abbildung 84b) zuverlässig genotypisiert werden. Der erste Scan erfolgte bei 41°C, der zweite nach achtmal Waschen mit HSB und der dritte Scan nach anschließenden vier Waschstufen mit LSB.

Fünf weitere Fehldetektionen dieses Experiments konnten korrigiert werden, indem die Analyserichtung geändert wurde (SNP15) oder ein Redesign der Diskriminatoren (für SNPs 22, 24, 32 und 45) erfolgte. Die Ergebnisse in der Abbildung 86 zeigen die SNPs, die durch Oligoredesign korrigiert oder zumindest verbessert wurden.

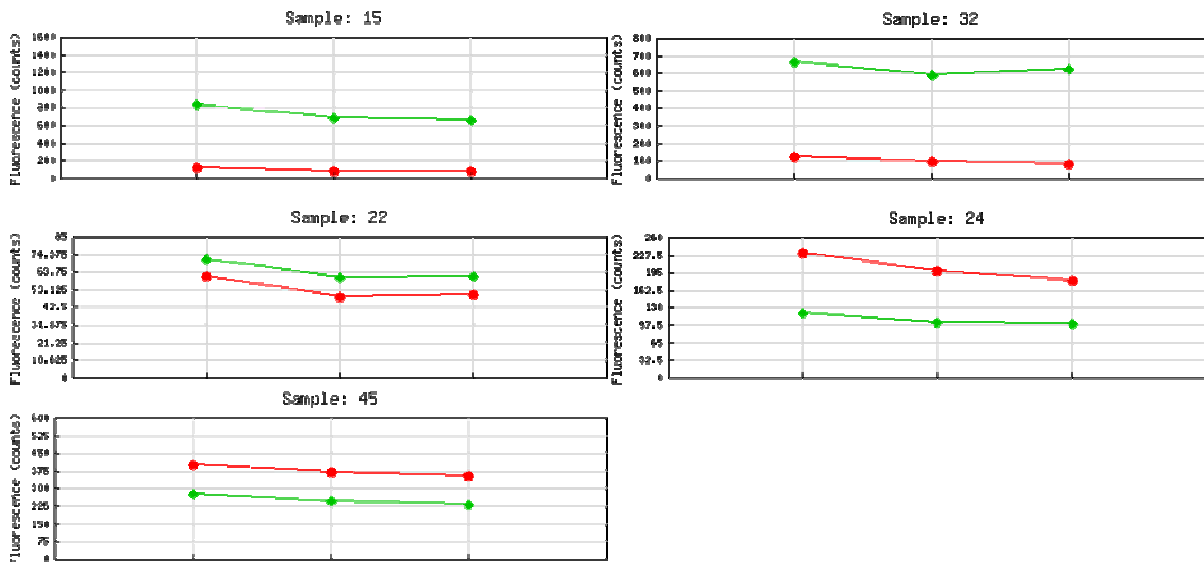


Abbildung 86: Nach Oligoredesign konnten weitere SNPs optimiert werden. Für SNP15 wurde die Leserichtung geändert. Beide Diskriminatoren der SNP22 und SNP24 wurden verlängert und für SNP32 und SNP45 wurde jeweils der Wildtyp verlängert.

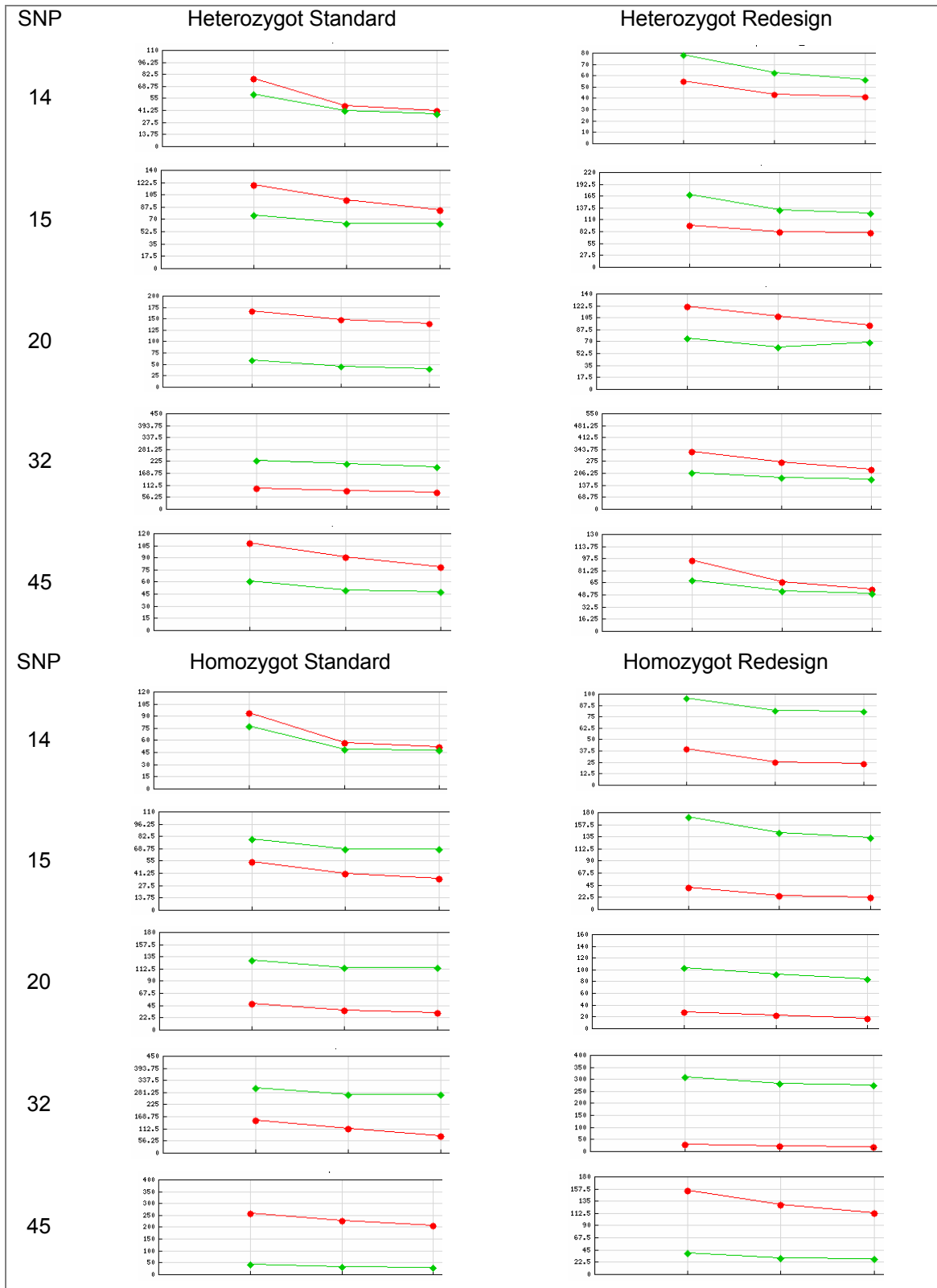


Abbildung 87: Die Trendgrafiken stellen die Ergebnisse für fünf SNPs und verschiedene Proben mit unterschiedlichen Genotypen dar. Gegenüberegestellt sind Typisierungen mit ursprünglich ausgewählten Oligos (Standard) und veränderten Capture-Oligos oder Diskriminatoren (Redesign). Der erste Kasten fasst die Grafiken zusammen für die Proben, die für diese fünf SNPs heterozygot sind. Der zweite Kasten stellt die homozygoten Proben dar.



Die Unterschiede zwischen den Standard Capture-Sonden bzw. Diskriminatoren, die ursprünglich ausgewählt wurden und den beim Redesign veränderten Oligos, wurden anhand mehrerer Experimente mit Personen von unterschiedlichen Genotypen erforscht. Dies war von Bedeutung, um genotypspezifische Fehldetektionen auszuschließen. Abbildung 87 zeigt exemplarisch für fünf SNPs jeweils für eine heterozygote und eine homozygote Probe Trendgrafiken, die die Typisierung mittels Standard Oligos der Typisierung mittels sog. Redesign Oligos gegenüberstellen. Diese Resultate veranschaulichen die Wirkung geringster Veränderungen der Oligos und weisen gleichzeitig auf die Notwendigkeit von präzisiertem Oligodesign hin. So wurden z.B. für SNP20 und SNP32 für beide Genotypen (Heterozygot und Wildtyp) eine Verbesserung durch die veränderten Diskriminatoren erzielt. Für SNP14 und SNP15 konnten die Diskriminierungen erhöht werden, sodass für beide SNPs eine Wildtyp-Typisierung möglich ist und die Heterozygoten sich nicht wesentlich verschlechtert haben. Umgekehrt konnte nach dem Redesign ein Heterozygot für SNP45 ohne Qualitätsverlust bei einem Homozygoten typisiert werden.

Trotz zahlreicher Versuche konnten nicht alle 48 SNPs für das Capture down-Assay optimiert werden. Mit Hilfe der Agilent-Elektrophorese konnte gezeigt werden, dass die Microarray-Genotypisierung maßgeblich von dem Erfolg der PCRs abhängt. Der Analyseerfolg der Triplex-PCRs typisiert mittels Capture down-Assay korreliert mit der Konzentrations- bzw. Molaritätsangabe der einzelnen SNPs, ermittelt mit Hilfe der Agilent-Technologie (Abbildung 88). So sind die Fluoreszenzintensitäten, gemessen während der Nanogen- und Agilent-Analyse, für die meisten SNPs gleich (Abbildung 88 und z.B. die niedrigen SNP13 und SNP26 in Abbildung 89). Gleichzeitig kristallisieren sich die SNPs heraus, die keine PCR-Abhängigkeit zeigen, so zum Beispiel die SNPs 07, 10, 44 und 41 (Abbildung 89).

Die starke PCR-Abhängigkeit demonstriert am besten der Ergebnisvergleich zwischen der Genotypisierung von Triplex- und 48plex-PCR (Abbildung 90a und b). Alle neun Beispiel-SNPs der drei Triplexe liefern einwandfreie und gut diskriminierte Genotypen (Abbildung 90a). Dagegen sind die SNPs einer 48plex-PCR in mehreren Fällen falsch oder nicht bestimmbar (Abbildung 90b). Nachdem die SNPs, die keine PCR-Abhängigkeit gezeigt haben und auch durch Optimierungen nicht verbessert werden konnten, identifiziert waren, wurde die Entscheidung getroffen, diese aus dem Multiplex zu entfernen. Damit stand zuletzt ein 35plex-PCR zur Verfügung. Die Reduktion der Multiplex auf 35 SNPs bewirkte auch eine sichtbare Verbesserung der Microarray-Genotypisierung im Vergleich zu einer 48plex-PCR (Abbildung 90c).

Schließlich wurde die Optimierung mit einem letzten Experiment abgeschlossen. Dabei wurde die elektronische Adressierung des 35plex-PCRs mit unterschiedlichen Volumina ein letztes Mal verglichen. Abbildung 91 bestätigt die frühere Feststellung, dass 10µl einen höheren Background und damit ungenauere Werte liefern als 3µl ungereinigtes PCR-Produkt. Gleichzeitig zeigt die Abbildung 91c, dass die dreifache sukzessive elektronische Hybridisierung von 1µl ungereinigtem PCR-Produkt eine einwandfreie Genotypisierung für alle SNPs erlaubt.

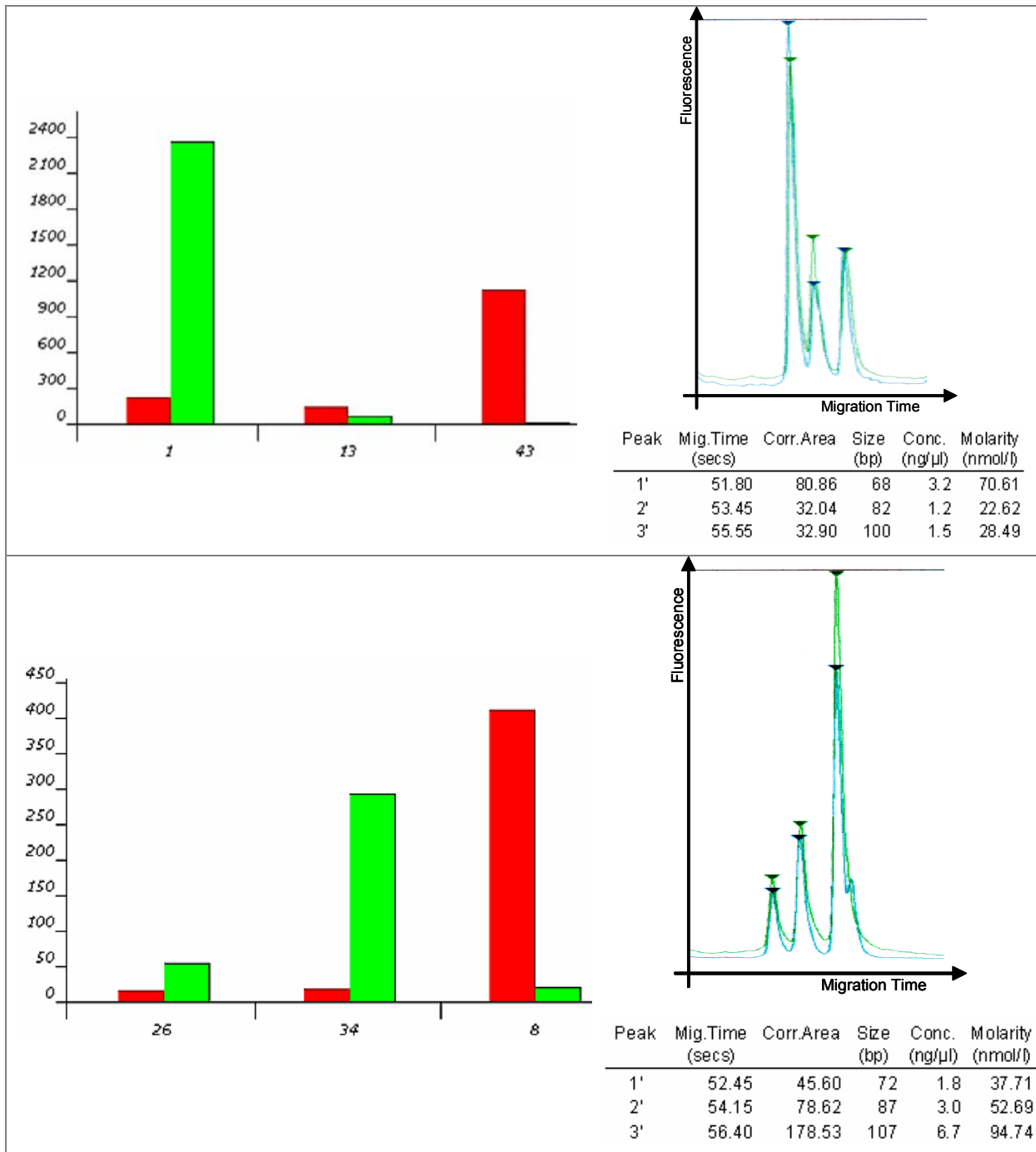


Abbildung 88: Zwei Beispiele für die Korrelation zwischen dem Typisierungserfolg mittels Capture down-Assay und Konzentrations- bzw. Molaritätsangabe der einzelnen SNPs ermittelt mit Agilent-Technologie. Die Histogramme zeigen zwei Triplex-PCR und jeweils rechts davon sind die Ergebnisse der Agilent-Analyse zu denselben Triplex-PCRs in Elektropherogrammen und Rohdatentabellen aufgeführt. Zur Darstellung wurden in einem Agilent-Elektropherogramm zwei Proben übereinander gelegt (blaue und grüne Linie). Die blauen Peaks entsprechen der Probe in der Rohdatentabelle und auch des Nanogen-Histogramms.

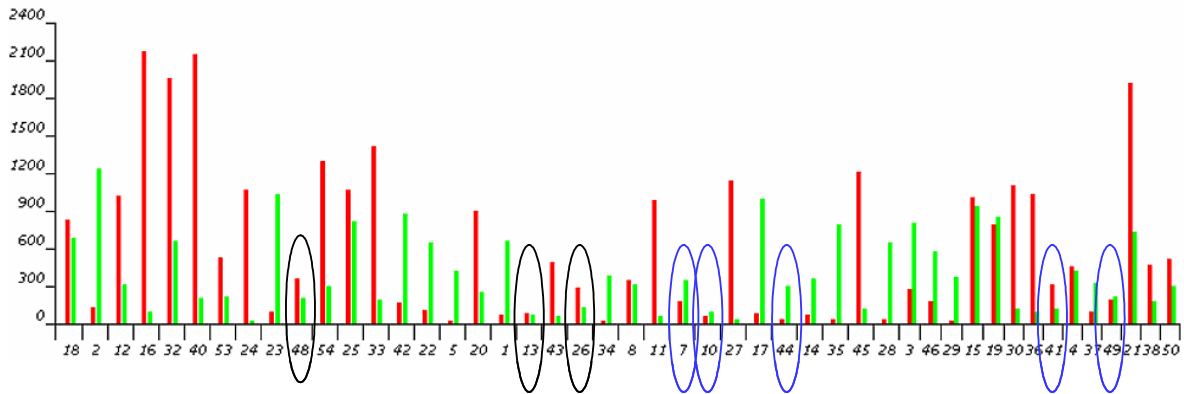


Abbildung 89: Genotypisierung der Triplex-Proben, die in einem Histogramm zusammengeführt wurden. Die schwarzen Kreise geben die SNPs wieder, die PCR abhängige niedrige Signale zeigen. Die blauen Kreise dagegen kennzeichnen die SNPs die keine Korrelation mit den Agilent-Ergebnissen aufweisen und daher niedrige Signale nicht PCR bedingt sein können.

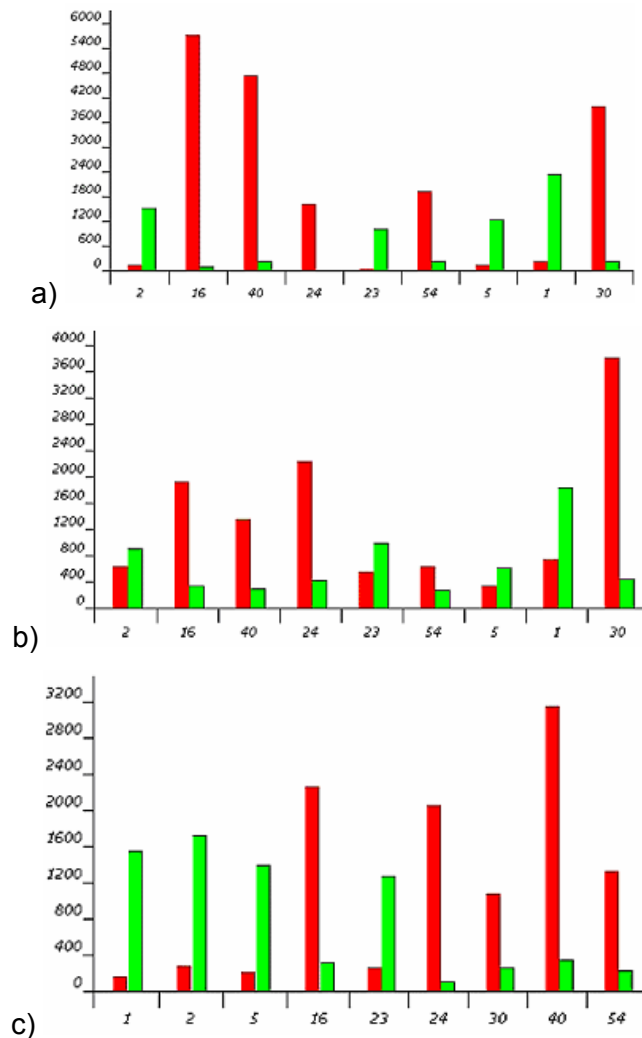


Abbildung 90: Neun Beispiel-SNPs in unterschiedlichen PCR-Ansätzen amplifiziert. Histogramm a) führt die Ergebnisse von Triplexen zusammen, Histogramm b) präsentiert die Ergebnisse einer 48plex-PCR und Histogramm c) zeigt in aufsteigender Reihenfolge die gleichen SNPs einer 35plex-PCR.

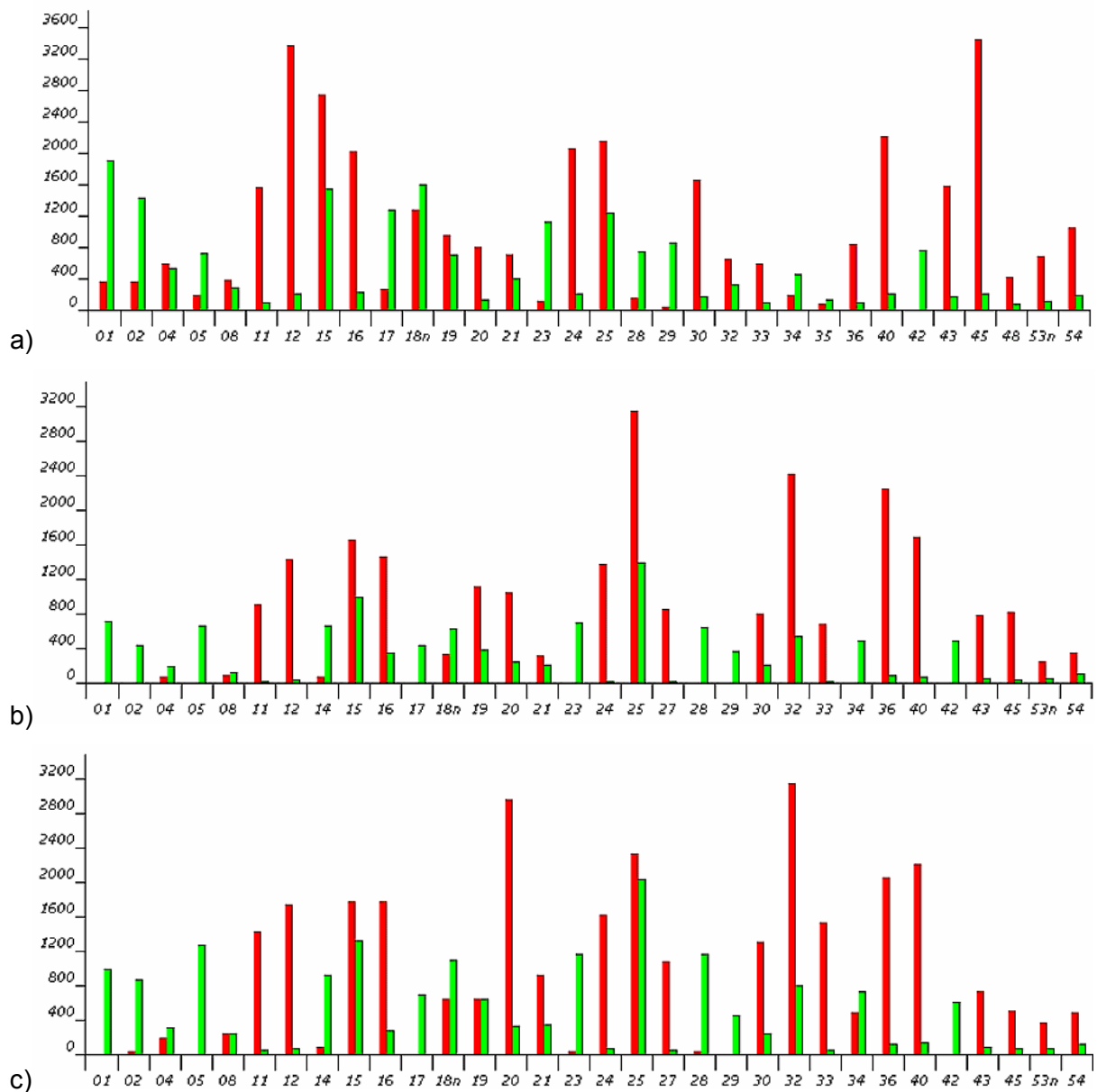


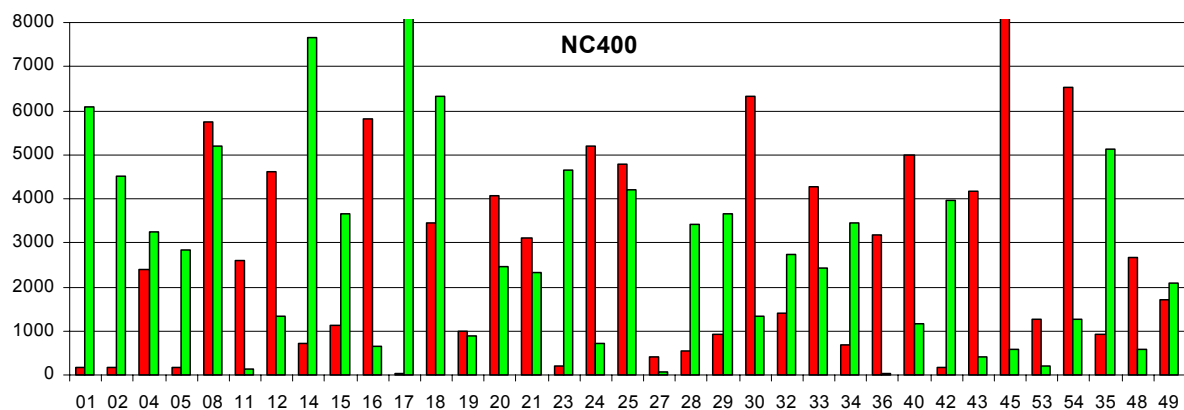
Abbildung 91: Histogramme von simultaner Genotypisierung von 35plex-PCR mittels optimiertem Capture down-Assay. Für alle drei Analysen wurde die gleiche ungereinigte PCR-Probe verwendet, mit etwas unterschiedlicher Hybridisierung. Es wurden 10µl (a) bzw. 3µl (b) PCR-Produkt hybridisiert. Histogramm c) zeigt die Vorteile von dreimal sukzessiver Hybridisierung der gleichen Probe gegenüber den zwei vorhergehenden Histogrammen.

Unter Berücksichtigung der Platzverhältnisse auf dem Mikrochip mit 100 zur Verfügung stehenden Pads fiel die Entscheidung auf 32 SNPs. Nachdem das optimierte Capture down-Assay mit Erfolg wiederholt werden konnte, stand ein Assay zur simultanen Genotypisierung von 32 SNPs für drei Individuen auf einem Microarray bereit für eine systematische Validierung.

Abbildung 92 demonstriert mit zwei Beispielen, dass das für die MBW100 optimierte Capture down-Assay auch auf die fortschrittliche NC400-Plattform übertragbar ist. Abbildung 91c und Abbildung 92a zeigen jeweils die Genotypen der gleichen Person. Dadurch werden auch die

Abweichungen zwischen den zwei Technologien sichtbar. Während die Ergebnisse von 26 SNPs identisch sind, weichen die Genotypen der SNPs 15, 20, 21, 32, 33 und 34 auf der NC400 (Abbildung 92a) von den Ergebnissen der MBW ab (Abbildung 91c). Die meisten der abweichenden SNPs (z.B. SNPs 20, 21, 33 und 34 in Abbildung 92b) zeigen auch Diskriminierungsprobleme bei der Typisierung zweier unterschiedlicher Personen auf einer NC400. Gleichzeitig liefern die restlichen SNPs die Bestätigung, dass die Genotypisierung mittels einer NC400 durchführbar ist, indem eine zuverlässige Diskriminierung stattfindet (z.B. SNPs 01, 05, 08, 12, 14, 16, 23, 28, 29, 30, 40, 45, 53 und 35 Abbildung 92).

a)



b)

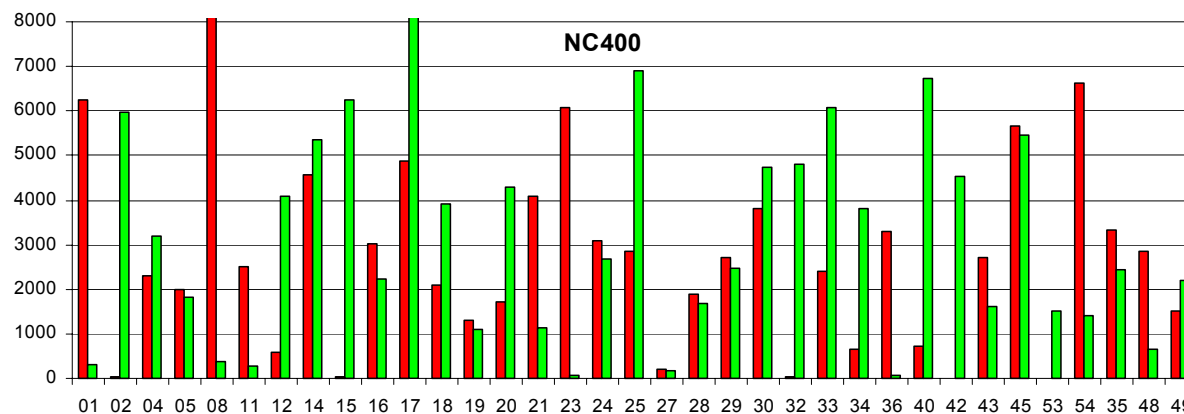


Abbildung 92: Die Ergebnisse von zwei Proben mit unterschiedlichen Genotypen (a und b) typisiert mittels NC400 und das für die MBW-Plattform optimierte Capture down-Assay. Die Probe im Histogramm a) und in der Abbildung 91 sind identisch und die Konkordanzen können direkt bestimmt werden. Bei der NC400-Typisierung konnten alle 35 SNPs typisiert werden, so kommen die SNPs 35, 48 und 49 hinzu, die aus Platzgründen nicht mittels MBW typisiert werden konnten.

### 3.3.3.3 Capture down-Assayvalidierung

Für die Capture down-Assayvalidierung wurden 40 verschiedene Personen in 65 Versuchen mit Hilfe des optimierten Protokolls genotypisiert. Die Ergebnisse wurden anschließend

verwendet, um die optimale statistische Auswertung für jeden SNP zu bestimmen (Kapitel 2.2.6.5). Dabei diente die SNaPshot-Genotypisierung für jede Probe als Referenz. In der linken Hälfte der Tabelle 32 sind die Ergebnisse der statistischen Auswertung mittels selbst programmierter Makros für jeden SNP aufgeführt. Durchschnittlich 88% der statistischen Auswertungen stimmten mit der SNaPshot-Genotypisierung überein. Dabei erreichte SNP35 mit 69% Konkordanz das niedrigste Ergebnis, während alle Proben für SNP15 und SNP40 mit 100%-igem Erfolg typisiert werden konnten.

Tabelle 32: Ergebnisse der Capture down-Assayvalidierung für 32 SNPs. Gegenübergestellt sind die makrobasierte statistische Auswertung und die automatische Auswertung nach Nanogen Onboard Software <sub>N</sub>LAB™ NanoChip® für dieselben Proben. Die Anzahl der Proben variiert zwischen 50 und 64, weil nicht immer alle SNPs analysiert werden konnten aus Gründen wie Platzmangel oder technischem Defekt. Als „Falsch“ wurden die Proben bewertet, die einen vom SNaPshot abweichenden Genotyp ergaben. „NN“ sind die Proben, die nicht eindeutig einem Genotyp zugeordnet werden können. Proben, die nach der Background-Subtraktion keine der beiden Farben >50RFU erreicht haben, gelten als „Ausfall“. Die unterste Grau markierte Zeile nennt Durchschnittswerte für die gesamte 32plex.

SNP	statistische Auswertung						automatische Auswertung					
	Anzahl	Richtig	FALSCH	NN	Ausfall	Konkordanz	Anzahl	Richtig	FALSCH	NN	Ausfall	Konkordanz
SNP01	56	46	3	6	1	82%	56	39	0	16	1	70%
SNP02	58	54	0	3	1	93%	58	53	1	3	1	91%
SNP04	51	40	2	1	8	78%	51	33	1	9	8	65%
SNP05	60	58	0	2	0	97%	60	47	0	5	0	78%
SNP08	63	56	2	3	2	89%	63	55	2	4	2	87%
SNP11	61	54	2	3	2	89%	61	43	5	11	2	70%
SNP12	59	57	1	1	0	97%	59	54	1	4	0	92%
SNP15	59	59	0	0	0	100%	59	57	0	2	0	97%
SNP16	59	53	3	3	0	90%	59	45	4	10	0	76%
SNP17	57	55	0	1	1	96%	57	46	0	10	1	81%
SNP18	56	52	3	1	0	93%	56	36	4	16	0	64%
SNP19	53	51	0	0	2	96%	53	47	1	3	2	89%
SNP20	59	53	0	2	4	90%	59	39	3	13	4	66%
SNP21	62	51	5	3	3	82%	62	38	7	14	3	61%
SNP23	64	56	0	3	5	88%	64	50	0	9	5	78%
SNP24	54	50	1	1	2	93%	54	39	0	13	2	72%
SNP27	53	42	1	2	8	79%	53	40	3	2	8	75%
SNP28	55	51	0	0	4	93%	55	48	0	3	4	87%
SNP29	50	48	0	1	1	96%	50	46	0	3	1	92%
SNP30	59	50	3	2	4	85%	59	47	3	5	4	80%
SNP33	58	47	0	1	10	81%	58	44	0	4	10	76%
SNP34	56	53	0	1	2	95%	56	51	0	3	2	91%
SNP35	55	38	4	4	9	69%	55	32	6	8	9	58%
SNP36	55	45	0	5	5	82%	55	38	0	12	5	69%
SNP40	59	59	0	0	0	100%	59	58	0	1	0	98%
SNP42	54	43	0	7	4	80%	54	32	6	12	4	59%
SNP43	57	44	1	9	3	77%	57	35	6	13	3	61%
SNP45	60	53	1	3	3	88%	60	42	1	14	3	70%
SNP48	55	45	2	1	7	82%	55	36	3	9	7	65%
SNP49	60	53	0	1	6	88%	60	50	0	4	6	83%
SNP53	61	53	0	2	6	87%	61	47	0	8	6	77%
SNP54	61	59	0	0	2	97%	61	58	0	1	2	95%
Anzahl	1839	1628	34	72	105		1839	1425	57	244	105	
Prozent		88%	2%	4%	6%	88%		77%	3%	13%	6%	77%

In der rechten Hälfte der Tabelle 32 ist die automatische Auswertung der statistischen Auswertung (links) gegenübergestellt. Die automatische Auswertung erfolgte nach folgenden Einstellungen:

**Probe Designation Thresholds:**

Heterozygote < 1:2 < No Designation < 1:3,5 < Homozygote

Signal-to-noise ratio for *Pad Exclusion* or *No Designation* : 3:1

Minimum background-subtracted signal for *Pad Inclusion* : 50

Diese Einstellung ergab das beste Ergebnis für die automatische Auswertung, eine richtige Typisierung war dennoch mit einer Trefferquote von 77% viel niedriger als mit der statistischen Methode. Die niedrige Konkordanz ergibt sich vor allem aus dem Zuwachs an nicht typisierbaren Proben (NN=13% in Tabelle 32). Die Unterschiede der zwei Auswertungsmethoden werden in Abbildung 93 deutlich. Diese zeigt die Verteilungen der SNPs nach ihrer Konkordanz. Abbildung 93a zeigt, dass der größte Anteil der SNPs mittels statistischer Auswertung bei >80% und >90%-iger Konkordanz liegt und kein SNP unter 60%, dafür aber 6% mit 100% Konkordanz typisiert werden konnten. Dagegen zeigt Abbildung 93b 6% der SNPs unter 60% und kein SNP bei 100% Konkordanz. Der größte Anteil nach einer automatischen Auswertung liegt hier bei >70% Konkordanz.

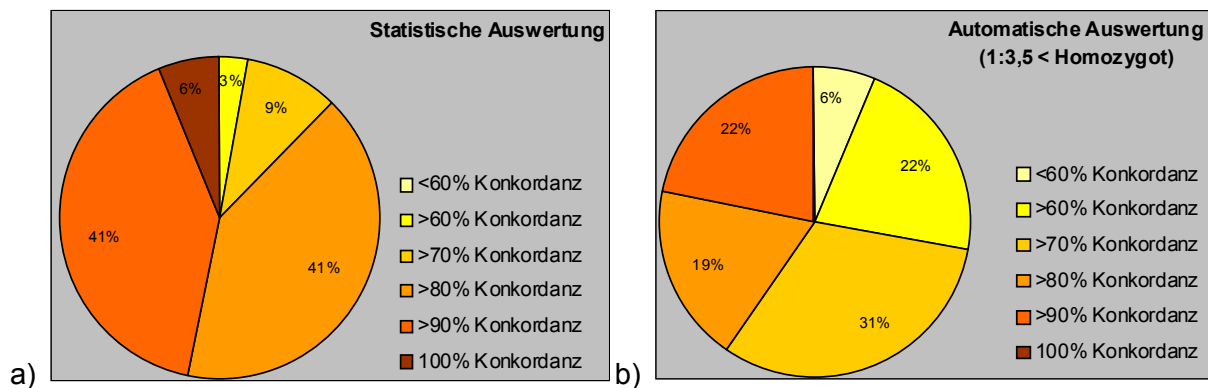


Abbildung 93: Die zwei Kreisdiagramme stellen die Konkordanzen von <60% bis 100%, die durch die statistische (a) bzw. automatische (b) Auswertung erreicht wurden.

Exemplarisch ist anhand des SNP34 der Vorteil und die Funktionalität der statistischen Auswertung dargestellt (Abbildung 94). Abbildung 94 zeigt die Einstellungen der Konfidenzbänder zwischen den einzelnen Genotypen, die eine optimale Auswertung für den SNP34 erlauben. Die Winkel der einzelnen Konfidenzbänder können verschoben werden, indem die Irrtumswahrscheinlichkeit verändert wird. Das Diagramm b) und die dazu gehörende Tabelle veranschaulichen die Auswirkung, wenn die Irrtumswahrscheinlichkeit für Genotyp M von 0,001% auf 1% erhöht wird. Das Konfidenzband für Genotyp M entfernt sich dementsprechend von dem Gaußschnitt zwischen M und MW (der Winkel des Konfidenzbandes für M verringert sich von <math><13^\circ</math> auf <math><9^\circ</math> Abbildung 94b). Dabei verkleinert sich der genotypbestimmende Bereich für Homozygot M und drei als Referenz typisierte Mutanten gelangen in die Grauzone und können nicht als Mutante richtig erkannt werden, d.h. die Irrtumswahrscheinlichkeit für diese Mutanten eine Fehlentscheidung zu treffen, ist bei dem in Abbildung 94b gewählten Winkel erheblich größer.

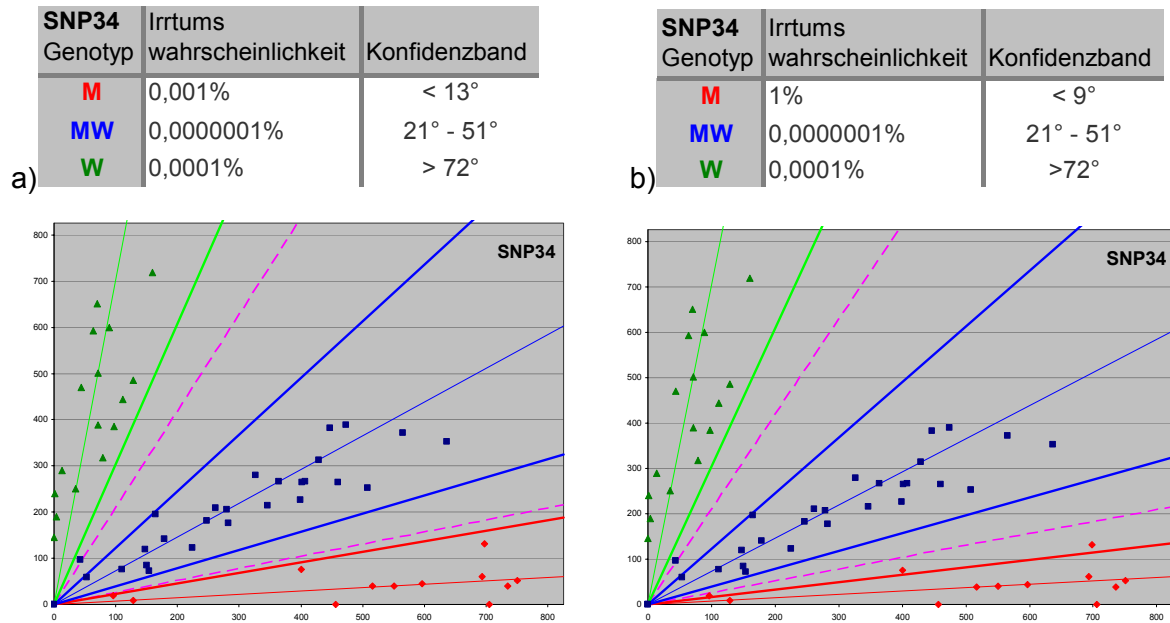


Abbildung 94: Die zwei Diagramme stellen die drei Genotypkorridore (fette Linien, M = rote Linie, MW = blaue Linien und W = grüne Linie) für den SNP34 dar. Die darüber liegenden Tabellen geben die vorgewählten Wahrscheinlichkeiten der zugelassenen Fehldeutungen und die dadurch bestimmten Konfidenzbänder für die drei Genotypen an. Diagramm b) zeigt die Auswirkung von einer Erhöhung des zugelassenen Irrtums auf 1% anhand des mutanten Genotyps. Das hat zur Folge, dass der mutante Genotypkorridor kleiner geworden ist.

In Tabelle 33a können die Ergebnisse der verschiedenen Auswertungsmethoden für SNP34 abgelesen werden. Die Genotypen, die mittels SNaPshot-Methode typisiert wurden, gelten als Referenz, die nächsten drei Spalten beinhalten die Ergebnisse, die mittels Nanogen-Technologie automatisch ermittelt oder nach Makro statistisch ausgewertet wurden. Die automatische Auswertung liefert, abgesehen von zwei Ausfällen, drei Abweichungen zur SNaPshot-Referenz (Proben 22, 32 und 38), während die statistische Auswertung nur eine Abweichung (Probe 32) ergibt. Wenn aber der genotypbestimmende Bereich für die homozygote Mutante verkleinert wird, erhöht sich die Irrtumswahrscheinlichkeit und es kommen drei Abweichungen hinzu (Proben 48, 49 und 50). Die zweite Hälfte der Tabelle 33b zeigt die Ergebnisse und die Vorteile der statistischen Auswertung gegenüber der automatischen Auswertung anhand des SNP35, der mit Abstand die niedrigste Konkordanz bei der Capture down-Genotypisierung erlangte (Tabelle 32). Sieben Proben, die mittels automatischer Auswertung keinem Genotyp zugeordnet wurden, konnten mittels statistischer Auswertung richtig typisiert werden. Dennoch bilden die genotypbestimmenden Bereiche in der Abbildung 95 keinen zuverlässigen Korridor für die drei Genotypen. Weil die Standardabweichungen der Heterozygoten und der homozygoten Wildtyp-SNP zum Median sehr groß sind, ergeben sich für diese zwei Genotypen sehr flache Gaußkurven (Abbildung 95).



Tabelle 33: Die Ergebnisse der verschiedenen Auswertungen sind für SNP34 (a) und SNP35 (b) als Beispiel aufgeführt. Mit Hilfe der SNaPshot-Referenz werden die Konkordanzen bestimmt. Die Auswertungen erfolgten automatisch und statistisch für beide SNPs. Für SNP34 wurden in einer zweiten statistischen Auswertung die Ergebnisse aufgeführt, nachdem der Genotypkorridor Mutante von 13° auf 9° eingengt wurde.

a) SNP34 Probe	SNaPshot Referenz	Auswertung			b) SNP35 Probe	SNaPshot Referenz	Auswertung	
		Automatisch	Statistisch M<13°	Statistisch M<9°			Automatisch	Statistisch
1	w	w	w	w	1	w	NN	w
2	w	w	w	w	2	w	mw	mw
3	w	w	w	w	3	w	w	w
4	w	w	w	w	4	w	w	w
5	w	w	w	w	5	w	w	w
6	w	w	w	w	6	w	w	w
7	w	w	w	w	7	w	w	w
8	w	w	w	w	8	w	w	w
9	w	w	w	w	9	w	mw	mw
10	w	w	w	w	10	w	mw	mw
11	w	w	w	w	11	w	NN	w
12	w	w	w	w	12	mw	NN	NN
13	w	w	w	w	13	mw	mw	mw
14	w	w	w	w	14	mw	mw	mw
15	w	w	w	w	15	mw	mw	mw
16	w	w	w	w	16	mw	mw	mw
17	mw	mw	mw	mw	17	mw	NN	mw
18	mw	mw	mw	mw	18	mw	mw	mw
19	mw	mw	mw	mw	19	mw	NN	mw
20	mw	mw	mw	mw	20	mw	mw	mw
21	mw	mw	mw	mw	21	mw	mw	mw
22	mw	NN	mw	mw	22	mw	mw	mw
23	mw	mw	mw	mw	23	mw	mw	mw
24	mw	mw	mw	mw	24	mw	m	NN
25	mw	mw	mw	mw	25	mw	NN	mw
26	mw	mw	mw	mw	26	mw	mw	mw
27	mw	mw	mw	mw	27	mw	mw	mw
28	mw	mw	mw	mw	28	mw	mw	mw
29	mw	mw	mw	mw	29	mw	mw	mw
30	mw	mw	mw	mw	30	mw	mw	mw
31	mw	mw	mw	mw	31	mw	mw	mw
32	mw	NN	NN	NN	32	mw	mw	mw
33	mw	mw	mw	mw	33	mw	NN	mw
34	mw	mw	mw	mw	34	mw	NN	mw
35	mw	mw	mw	mw	35	mw	mw	mw
36	mw	mw	mw	mw	36	mw	m	NN
37	mw	mw	mw	mw	37	mw	m	m
38	mw	NN	mw	mw	38	mw	mw	mw
39	mw	mw	mw	mw	39	m	m	m
40	mw	mw	mw	mw	40	m	m	m
41	mw	mw	mw	mw	41	m	m	m
42	mw	mw	mw	mw	42	m	m	m
43	m	m	m	m	43	m	m	m
44	m	m	m	m	44	m	m	m
45	m	m	m	m	45	m	m	m
46	m	m	m	m	46	m	m	NN
47	m	m	m	m	47	mw	Ausfall	Ausfall
48	m	m	m	NN	48	mw	Ausfall	Ausfall
49	m	m	m	NN	49	mw	Ausfall	Ausfall
50	m	m	m	NN	50	mw	Ausfall	Ausfall
51	m	m	m	m	51	mw	Ausfall	Ausfall
52	m	m	m	m	52	mw	Ausfall	Ausfall
53	m	m	m	m	53	mw	Ausfall	Ausfall
54	m	m	m	m	54	mw	Ausfall	Ausfall
55	mw	Ausfall	Ausfall	Ausfall	55	mw	Ausfall	Ausfall
56	mw	Ausfall	Ausfall	Ausfall				

NN Signal unbestimmt.

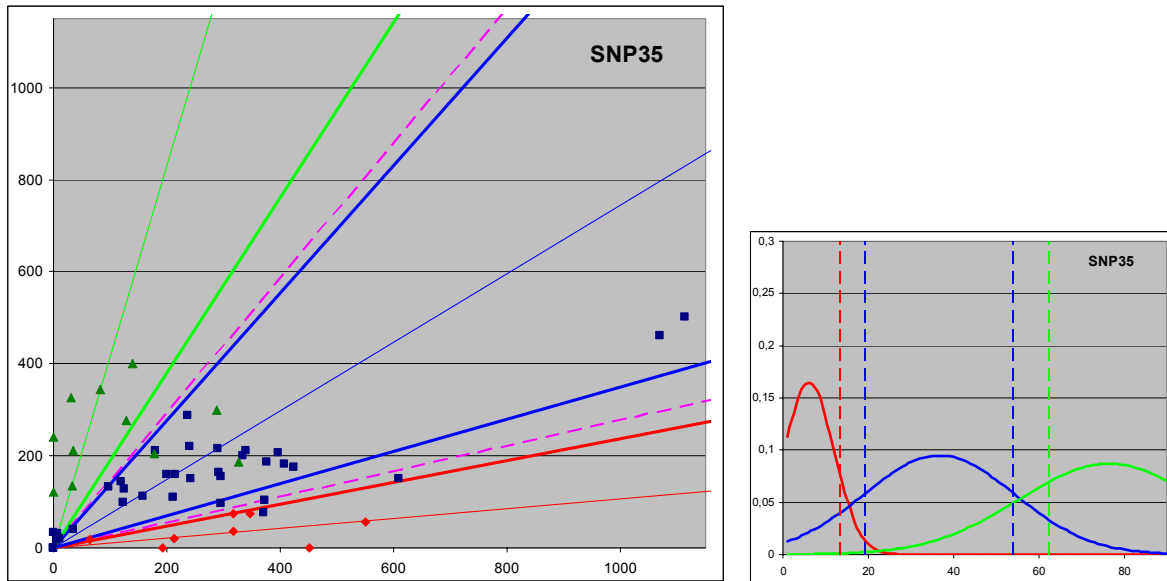


Abbildung 95: SNP35 stellt mit Abstand die Auswertung mit der kleinsten Konkordanz dar (Tabelle 32). Das wird auch deutlich bei der Betrachtung der Gaußkurven. Die grüne und blaue Gaußkurve, die die Abweichungen der Wildtyp Proben und Heterozygoten zum Median darstellen, erstrecken sich sehr flach und ineinander überfließend, deshalb ist eine genaue Korridorbestimmung für die Genotypen nicht möglich.

Die Ergebnisse der SNPs 54, 18, 36 und 17 sind als Beispiel aufgeführt, um die Vorteile der statistischen Auswertung zu betonen (Abbildung 96). Während die Winkel der Konfidenzbänder für den Genotyp Heterozygot bei den SNPs 54 und 18 groß sind, liegen sie für die SNPs 36 und 17 mit  $16^\circ < MW < 40^\circ$  bzw.  $< 44^\circ$  sehr niedrig (Abbildung 96). Das ist auf die ungleiche Signalstärke der zwei Fluoreszenzfarben zurückzuführen, das heißt bei den SNPs 54 und 18 sind die RFUs für den Wildtyp verhältnismäßig höher als die Signale für die Mutante. Umgekehrt gilt für die SNPs 36 und 17, dass die RFUs für die Mutante verhältnismäßig höher sind als die Signale für den Wildtyp. Die statistische Auswertung gleicht die Unterschiede der zwei Fluoreszenzfarben aus, indem sie die genotypbestimmenden Bereiche dementsprechend definiert (Tabelle 34).

Tabelle 34: Genotypbestimmende Winkel für die in Abbildung 96 dargestellten vier SNPs und Genotyp Mutante, Heterozygot und Wildtyp.

Genotyp	Genotypbestimmende Winkel			
	SNP54	SNP18	SNP36	SNP17
<b>M</b>	< 17°	< 27°	< 11°	< 9°
<b>MW</b>	37° - 56°	43° - 81°	16° - 40°	16° - 44°
<b>W</b>	> 69°	> 82°	> 62°	> 65°

In Tabelle 35 sind die genotypbestimmenden Winkel für alle 32 SNPs aufgeführt, die anhand der statistischen Makros ermittelt wurden. Schließlich wurden mittels dieser Winkel die Genotypen einer Probe für 32 SNPs ausgerechnet und bestimmt.

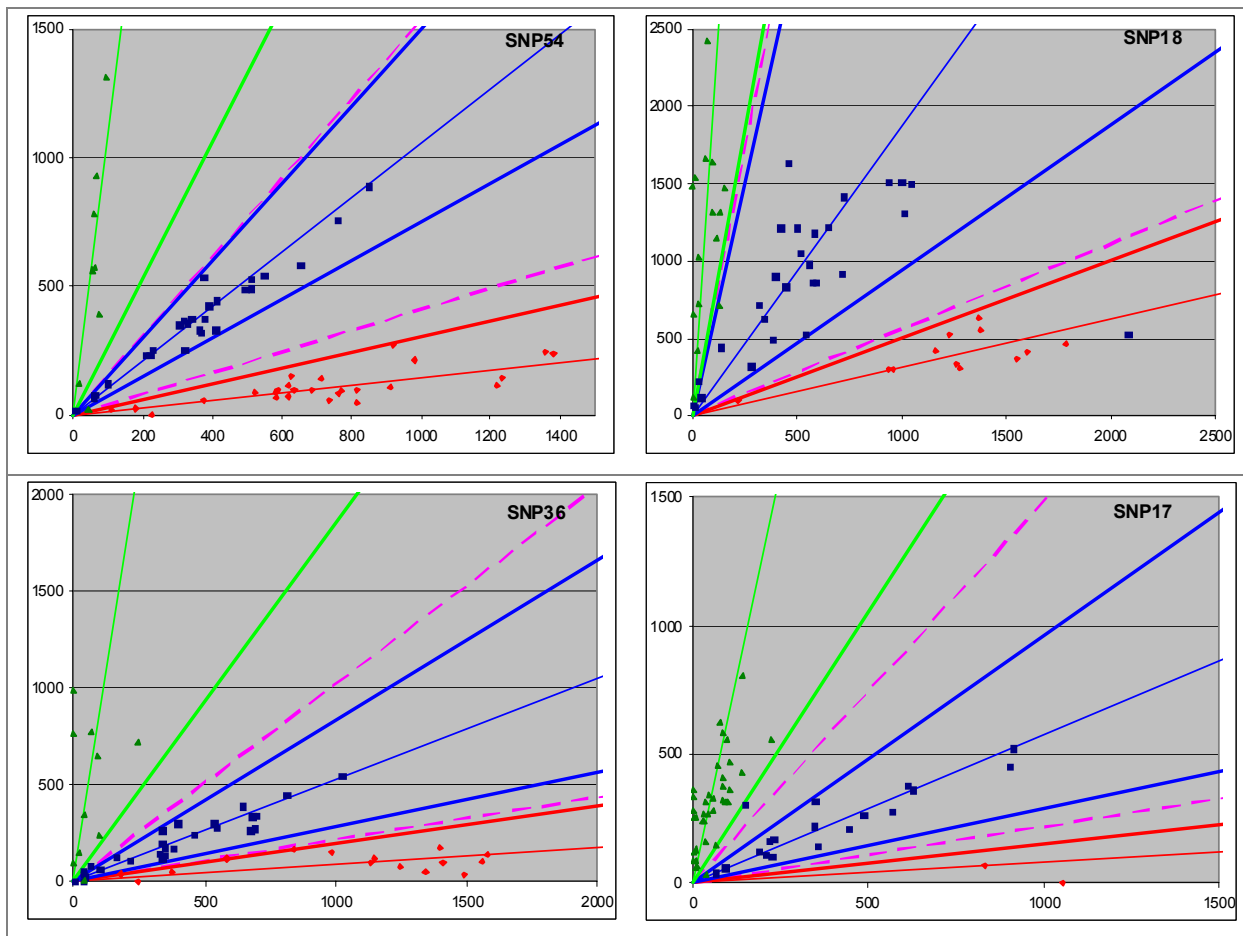


Abbildung 96: Die vier Diagramme stellen dar, warum die Auswertung nach festgelegten einfachen Verhältnissen (wie bei der automatischen Auswertung) nicht erfolgreich sein kann. Die oberen Diagramme zeigen zwei SNPs, die allgemein hohe Signale für den Wildtyp liefern im Vergleich zur Mutante. Aus dem Grund fallen die genotypbestimmenden Winkel größer aus als der Durchschnitt (Tabelle 35). Die unteren Diagramme zeigen dagegen zwei Beispiele für SNPs mit hohen Mutanten und niedrigen Wildtyp Signalen und damit auch kleinen genotypbestimmenden Winkeln.

In der Tabelle 36 sind die Konkordanzen für 40 verschiedene Personen aufgeführt, die für die Validierung des optimierten 32plex Capture down-Assays herangezogen wurden. Für die Beurteilung wurde, jeweils die beste Analyse einer DNA ausgesucht, falls sie mehrmals typisiert wurde. Mittels statistischer Auswertung konnte für 40 Proben durchschnittlich eine Konkordanz von 94% erzielt werden (Tabelle 36). Insgesamt 3% der SNPs konnten nicht einem Genotyp zugeordnet werden und jeweils 1% der SNPs ist ausgefallen oder wurde falsch interpretiert. Im Vergleich dazu erreichten 40 Proben nach einer automatischen Auswertung im Durchschnitt nur 82% Konkordanz, denn 13% der SNPs konnten keinem Genotyp zugeordnet werden und 3% wurden falsch gewertet (Tabelle 36).

Tabelle 35: Genotypbestimmende Winkel für 32 SNPs und Genotypen Mutante, Heterozygot und Wildtyp nach den Namen der SNPs sortiert.

SNP	Genotypbestimmende Winkel		
	M <sup>1)</sup>	MW	W <sup>2)</sup>
01	< 20°	30° - 77°	> 79°
02	< 14°	20° - 68°	> 73°
04	< 19°	20° - 61°	> 63°
05	< 19°	24° - 60°	> 66°
08	< 13°	20° - 52°	> 61°
11	< 16°	18° - 45°	> 61°
12	< 15°	20° - 56°	> 68°
15	< 13°	20° - 67°	> 74°
16	< 12°	14° - 52°	90°
17	< 9°	16° - 44°	> 65°
18	< 27°	43° - 81°	> 82°
19	< 15°	19° - 65°	> 65°
20	< 15°	20° - 48°	> 57°
21	< 14°	18° - 46°	> 54°
23	< 18°	21° - 44°	> 67°
24	< 10°	12° - 43°	> 66°
27	< 22°	28° - 53°	> 74°
28	< 9°	22° - 49°	> 67°
29	< 19°	29° - 66°	> 76°
30	< 14°	23° - 47°	> 57°
33	< 24°	35° - 63°	> 72°
34	< 13°	21° - 51°	> 72°
35	< 13°	19° - 54°	> 62°
36	< 11°	16° - 40°	> 62°
40	< 8°	16° - 58°	> 71°
42	0°	19° - 53°	> 59°
43	< 9°	12° - 60°	90°
45	< 13°	15° - 32°	> 61°
48	< 21°	28° - 56°	> 65°
49	< 15°	19° - 53°	> 70°
53	< 23°	29° - 62°	> 68°
54	< 17°	37° - 56°	> 69°

1) 0° bedeutet, dass kein Genotyp M zu Verfügung stand um den Winkel M zu bestimmen.

2) 90° bedeutet, dass kein Genotyp W zu Verfügung stand um den Winkel W zu bestimmen.

Zur Veranschaulichung wurden in Abbildung 97 die Ergebnisse der DNA33 dargestellt. Die statistische Auswertung einerseits und die Automatische andererseits weichen bei dieser Probe am stärksten voneinander ab. Während die statistische Auswertung einen 100%-igen Genotyp dieser Probe lieferte, konnte mittels automatischer Auswertung nur 75% Konkordanz erzielt werden (Tabelle 36 und Tabelle 37). So zum Beispiel konnten die SNPs 16, 17, 24, 35, 42, 48 und 53 nach ihren Verhältnissen zwischen den zwei Fluoreszenzfarben nicht richtig genotypisiert werden (Abbildung 97 und Tabelle 37). Nach dem Verhältnis der Farbintensitäten zu urteilen ist SNP20 ein Heterozygot, jedoch die Referenzmethode SNaPshot ergab einen Homozygoten Wildtyp für die DNA33 (Abbildung 97). Dieser Genotyp konnte anhand der statistischen Auswertung richtig ermittelt werden (Tabelle 37). Das bedeutet, dass die automatisierte Betrachtung der Verhältnisse zwischen den zwei Fluoreszenzfarben in einigen Fällen zu falschen Ergebnissen führen kann.

Tabelle 36: Ergebnisse der Capture down Assayvalidierung für 32 SNPs und 40 Proben. Gegenübergestellt sind die makrobasierete statistische Auswertung und die automatische Auswertung für dieselben Proben. Die Anzahl der typisierten SNPs variiert zwischen 27 und 32, weil nicht immer alle SNPs analysiert werden konnten aus Gründen wie Platzmangel oder technischem Defekt. Als „Falsch“ wurden die SNPs bewertet, die einen vom SNaPshot abweichenden Genotyp ergaben. „NN“ sind die SNPs, die nicht eindeutig einem Genotyp zugeordnet werden können. SNPs, die nach der Background-Subtraktion keine der beiden Farben >50RFU erreicht haben gelten als „Ausfall“. Die unterste, grau markierte Zeile gibt die Durchschnittswerte für alle 40 Proben an.

DNA	statistische Auswertung						automatische Auswertung					
	Anzahl	Richtig	FALSCH	NN	Ausfall	Konkordanz	Anzahl	Richtig	FALSCH	NN	Ausfall	Konkordanz
DNA1	27	26	0	1	0	96%	27	22	1	4	0	81%
DNA2	32	31	0	1	0	97%	32	30	0	2	0	94%
DNA3	32	32	0	0	0	100%	32	31	0	1	0	97%
DNA4	30	29	1	0	0	97%	30	25	2	3	0	83%
DNA5	32	31	0	1	0	97%	32	26	1	5	0	81%
DNA6	32	32	0	0	0	100%	32	26	0	6	0	81%
DNA7	32	31	0	0	1	97%	32	23	1	7	1	72%
DNA8	31	27	0	1	3	87%	31	23	0	5	3	74%
DNA9	32	29	0	1	2	91%	32	25	1	4	2	78%
DNA10	29	25	1	1	2	86%	29	18	2	7	2	62%
DNA11	31	30	0	1	0	97%	31	27	0	4	0	87%
DNA12	32	31	0	1	0	97%	32	27	1	4	0	84%
DNA13	31	28	0	3	0	90%	31	26	1	4	0	84%
DNA14	31	22	1	4	4	71%	31	22	2	3	4	71%
DNA15	32	26	2	0	4	81%	32	25	1	2	4	78%
DNA16	32	31	0	1	0	97%	32	24	1	7	0	75%
DNA17	30	26	0	2	2	87%	30	23	2	3	2	77%
DNA18	32	32	0	0	0	100%	32	32	0	0	0	100%
DNA19	32	31	0	1	0	97%	32	29	1	2	0	91%
DNA20	32	32	0	0	0	100%	32	28	0	4	0	88%
DNA21	32	32	0	0	0	100%	32	29	1	2	0	91%
DNA22	32	30	0	1	1	94%	32	27	1	3	1	84%
DNA23	32	31	0	1	0	97%	32	23	1	8	0	72%
DNA24	32	31	0	1	0	97%	32	26	0	6	0	81%
DNA25	32	32	0	0	0	100%	32	28	0	4	0	88%
DNA26	29	28	0	1	0	97%	29	25	0	4	0	86%
DNA27	32	28	1	3	0	88%	32	27	2	3	0	84%
DNA28	32	26	3	3	0	81%	32	27	3	2	0	84%
DNA29	32	32	0	0	0	100%	32	25	0	7	0	78%
DNA30	32	28	3	1	0	88%	32	24	1	7	0	75%
DNA31	32	31	1	0	0	97%	32	26	0	6	0	81%
DNA32	32	32	0	0	0	100%	32	26	1	5	0	81%
DNA33	32	32	0	0	0	100%	32	24	1	7	0	75%
DNA34	32	30	1	1	0	94%	32	28	0	4	0	88%
DNA35	32	29	2	1	0	91%	32	26	4	2	0	81%
DNA36	32	28	1	3	0	88%	32	27	3	2	0	84%
DNA37	32	30	1	1	0	94%	32	27	2	3	0	84%
DNA38	32	29	1	2	0	91%	32	26	2	4	0	81%
DNA39	32	30	0	2	0	94%	32	28	2	2	0	88%
DNA40	32	32	0	0	0	100%	32	28	1	3	0	88%
Anzahl	1261	1183	19	40	19		1261	1039	42	161	19	
Prozent		94%	1%	3%	1%	94%		82%	3%	13%	1%	82%

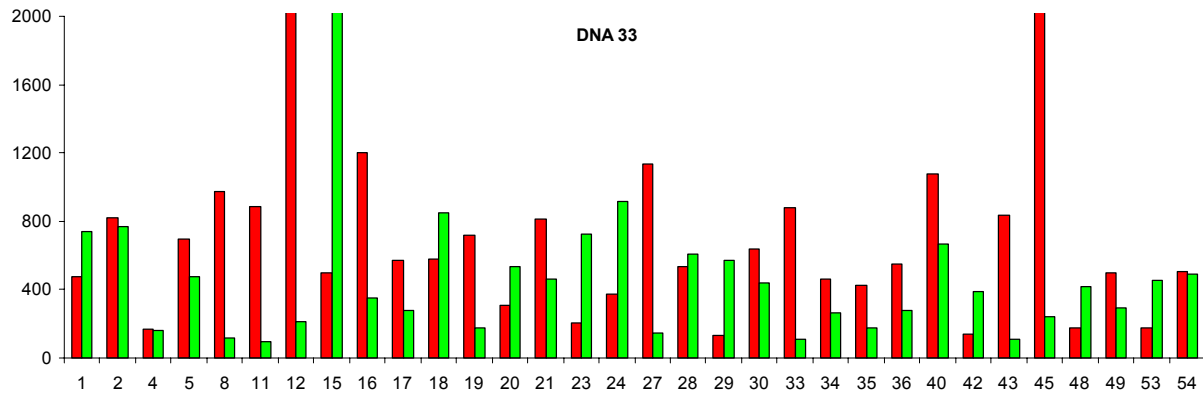


Abbildung 97: Die Fluoreszenzsignale der 32 SNPs für die Probe 33 aus der Tabelle 36.

Tabelle 37: Ergebnisse der Probe 33 sortiert nach den Namen der 32 SNPs. Die ersten zwei Spalten geben die Rohdaten für die roten und grünen Fluoreszenzsignale wieder. Die nächsten drei Spalten führen die Ergebnisse der SNaPshot-Typisierung und des Capture down-Assays nach statistischer Auswertung bzw. nach automatischer Auswertung auf. Grau markierte NN konnten keinem Genotyp zugeordnet werden, rote Markierung deutet eine Abweichung zur SNaPshot Referenz an.

SNP	Fluoreszenz RFU		SNaPshot Referenz	Statistische Auswertung	Automatische Auswertung
	M	W			
1	478,36	739,38	CT	CT	CT
2	821,92	767,50	GA	AG	AG
4	165,46	158,99	AT	TA	TA
5	695,47	474,13	AG	AG	AG
8	977,92	116,36	A	A	A
11	888,95	94,33	A	A	A
12	2253,01	209,54	A	A	A
15	498,3	3228,42	A	A	A
16	1199,09	349,67	GT	GT	NN
17	568,8	278,60	GA	GA	NN
18	576,7	849,80	CA	CA	CA
19	716,86	178,28	A	A	A
20	304,68	531,98	C	C	AC
21	812,1	459,55	AG	AG	AG
23	206,39	726,88	G	G	G
24	373,69	917,49	G	G	NN
27	1136,02	144,33	C	C	C
28	535,85	606,37	AG	AG	AG
29	131,05	568,77	A	A	A
30	638,95	440,41	AC	CA	CA
33	881,79	108,35	A	A	A
34	460,2	264,41	CG	CG	CG
35	423,46	175,65	AG	GA	NN
36	547,67	278,34	TA	AT	AT
40	1074,84	667,73	AG	GA	GA
42	136,71	391,19	T	T	NN
43	835,63	108,18	T	T	T
45	2457,22	245,28	C	C	C
48	175,08	420,9	C	C	NN
49	496,37	293,16	AG	GA	GA
53	176,67	452,81	A	A	NN
54	508,87	491,57	AG	AG	AG
<b>Konkordanz DNA 33</b>				<b>100%</b>	<b>75%</b>

## 4 Diskussion

### 4.1 Diskussion zur Y-chromosomalen Validierungsstudie

Mittels Y-chromosomaler Marker können räumliche und zeitliche Verläufe einer demographischen Expansion von Populationen dargestellt werden, wobei die genetische Drift mit zufälligen Veränderungen der Gen- oder Allelhäufigkeit den wichtigsten Evolutionsfaktor darstellt. Flaschenhalseffekte, Gründereffekte, Migration und Isolation beeinflussen die patrilineare Vererbung des Y-chromosomalen Markers stärker als die autosomalen Merkmale und lösen damit auch stärkere Differenzierungen der Y-chromosomalen Allel- bzw. Haplogruppenfrequenzen aus.

Wegen des Ausbleibens von genetischer Rekombination im Genom können die Y-chromosomalen Polymorphismen in Haplogruppen zusammengefasst werden, die einen phylogenetischen Baum bilden (Jobling und Tyler-Smith 2003). Im Jahr 2002 veröffentlichte das Y-Chromosome Consortium (YCC) 245 binäre Polymorphismen, die 153 Haplogruppen, davon 18 Hauptgruppen (Haplogruppen A-R) definieren (Y-Consortium 2002). Darauf folgende Publikationen erhöhten die Anzahl der SNPs, die aber vollständig den vom YCC vorgeschlagenen Haplogruppen zugeordnet werden konnten (Jobling und Tyler-Smith 2003, Cinnioglu et al. 2004, Rootsi et al. 2004). Die Eigenschaften der binären Y-chromosomalen Marker bringen in verschiedenen Bereichen hohen Nutzen. Ihre genetische Diversität liefert nicht nur Informationen über die männliche Migration in der Vergangenheit (Rosser et al. 2000, Arredi et al. 2004, Brion et al. 2004, Semino et al. 2004, Bouakaze et al. 2007), sie leisten auch wertvolle Hilfe in der forensischen Molekulargenetik bei speziellen Fragestellungen, in denen die gängigen forensischen Marker der Routine (z.B. STRs) keine ausreichenden Antworten liefern, so zum Beispiel bei Sexualstraftaten oder bei schwierigen Vaterschaftsbestimmungen (Jobling et al. 1997, Gusmao et al. 1999, Jobling 2001, Schneider 2008).

Das Ziel des ersten Arbeitsschrittes war die Selektion Y-chromosomaler SNPs aus der großen Sammlung der bereits veröffentlichten Loci. Unter Berücksichtigung der publizierten Eigenschaften der SNPs wurden 29 Y-chromosomale Loci ausgewählt, die 31 Haplogruppen, darunter 12 Haupthaplogruppen definieren. Demnach wurde die Gruppe aus 29 SNPs Major Y-chromosome haplogroup typing kit genannt (Brion et al. 2005). Um in den unterschiedlichen Laboreinrichtungen zuverlässige Ergebnisse generieren zu können, war anschließend die Entwicklung und Etablierung einer stabilen Multiplexstrategie nötig. In erster Linie wurde ein Standardprotokoll für die PCR-Amplifikation von 29 SNPs in einer einzigen Reaktion eingeführt. Zweitens wurde eine Minisequenzierung mittels SNaPshot™-Kit zur Detektion der 29 SNPs etabliert. Von den vielen existierenden Genotypisierungsmethoden zeichnete sich der SNaPshot als das geeignete Verfahren für eine Validierung aus, denn für diese Methode wird nur eine Kapillarelektrophorese-Plattform benötigt, welche in den meisten Routinelaboratorien vorhanden ist und somit eine Neuanschaffung erspart. Weiterhin war von Vorteil, dass die beteiligten Arbeitsgruppen bereits Erfahrungen mit der

Minisequenzierungsreaktion gesammelt hatten, was die Standardisierung der Analyse erleichterte.

Nachdem die Etablierung der Minisequenzierung abgeschlossen war, folgte die Validierung des Major Y-chromosome haplogroup typing kits und des methodischen Verfahrens in den drei Laboratorien mittels einer Sammlung von 1126 DNA-Proben, die von nicht verwandten, männlichen Probanden aus zwölf unterschiedlichen Populationen stammten. Dabei wurde die Genauigkeit und Reproduzierbarkeit der Typisierungsmethode als zuverlässig bewertet.

## 4.1.1 Methodische Anmerkungen

### 4.1.1.1 Segmentduplikationen

Bekanntlich bestehen 30-40% des Y-Chromosoms aus inter- oder intrachromosomalen Verdoppelungen mit 90 bis 98% Übereinstimmung (Bailey et al. 2002, Skaletsky et al. 2003). Die Anzahl der Duplikationen mit mehr als 98%-iger Übereinstimmung ist schwierig zu bestimmen, da 80% dieser Gruppe in den Datenbanken als falsch-positiv identifiziert wurden (Bailey et al. 2001, Lander et al. 2001) und gleichzeitig wahre Segmentverdoppelungen fehlen können, weil sie nicht als solche erkannt wurden (Bailey et al. 2002). Schätzungsweise 10% der SNPs aus den Datenbanken sind paralogous sequence variants (PSVs), abgeleitet von Segmentverdoppelungen (Estivill et al. 2002).

Bei der Genotypisierung des Kollektivs aus Deutschland mittels des Major Y-chromosome haplogroup typing kits wurden bei 43% der Proben für den SNP P25 Doppelpeaks detektiert. SNP P25 wurde als PSV erkannt, abgeleitet von einer polymorphen Segmentverdoppelung. Daraus ergibt sich, dass bei der Detektion des mutanten A-Allels zusätzlich das Stammallel C erscheint. Diese Beobachtung stimmt mit den Ergebnissen einer systematischen Studie überein, in der 600 männliche Individuen mit unterschiedlicher Abstammung für SNP P25 mittels SNaPshot-Minisequenzierung analysiert wurden und dabei kein Allel A ohne Begleitung von Allel C detektiert wurde (Sanchez et al. 2004). Dieses Phänomen wurde unter anderem auch mit Hilfe der Nanogen-Technologie untermauert (Borsting et al. 2004) und auch die Sequenzierungen dieser Proben ergaben immer einen Doppelpeak, bestehend gleichzeitig aus den A- und C-Basen (Abbildung 98).

Weiterführende Veröffentlichungen (Kuroda-Kawaguchi et al. 2001, King et al. 2005, Adams et al. 2006) zeigten, dass P25 in der AZFc Region des Y-Chromosoms (q-Arm) in Form einer dreifachen paralogen g-Typ Wiederholung vorliegt (Abbildung 99). Die in Europa sehr häufig vorkommende Haplogruppe R1b\* ist definiert durch die Detektion der Allele C/C/A. Während g1 und g2 ein Stammallel C aufweisen, besitzt die distale g3-Wiederholung ein Allel A.

Ein Teil des PSVs liegt in einer Palidromsequenz, wo bekannterweise häufig eine Genkonversion in Form von Rückmutationen stattfindet (Rozen et al. 2003). Die Analyse eines großen Kollektivs zeigte Konversionen (Adams et al. 2006), d.h. bei der Rückmutation wurde nur das Allel C (C/C/C) detektiert, obwohl die Probe, durch eine Mutation bei SNP M269, eindeutig der R1b Haplogruppe zugeordnet werden konnte. Weitere Genkonversionen wurden durch ein verstärktes Allel A beobachtet, immer in Begleitung von Allel C (C/A/A), wobei nie eine reine Allel A (A/A/A) Konversion auftauchte. Die möglichen Konfigurationen



der AZFc Region des Y-Chromosoms, die zu einer Genkonversion des P25 PSVs führen können, sind in Abbildung 100 aufgeführt.

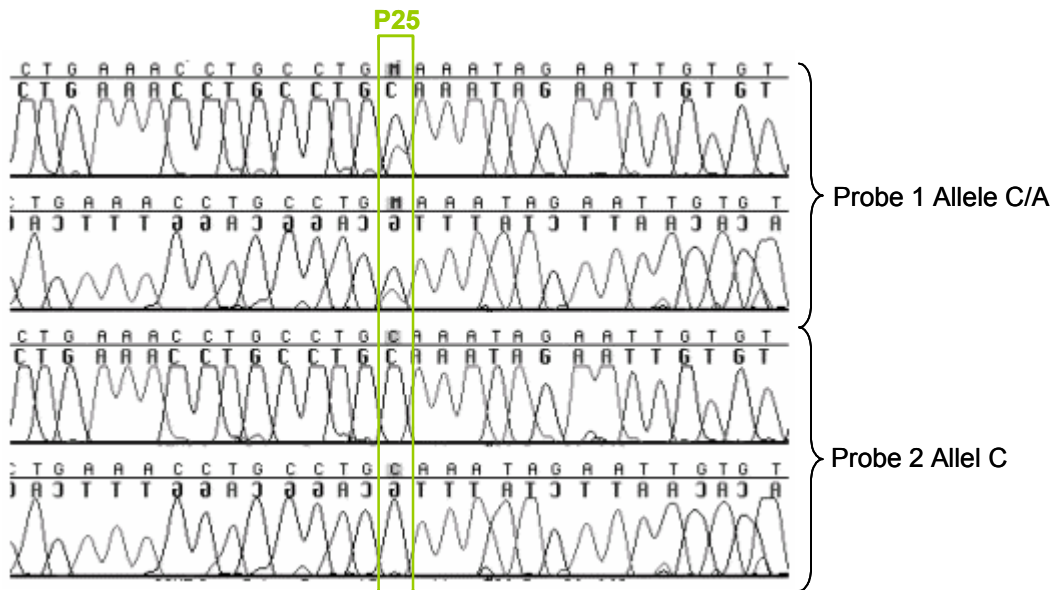


Abbildung 98: Forward und reverse Sequenzierungen von zwei PCR-Fragmenten, die SNP P25 enthalten. Probe 1 ergibt Signale für das A- und C-Allel, Probe 2 besitzt nur Allel C (Sanchez et al. 2004, modifiziert).

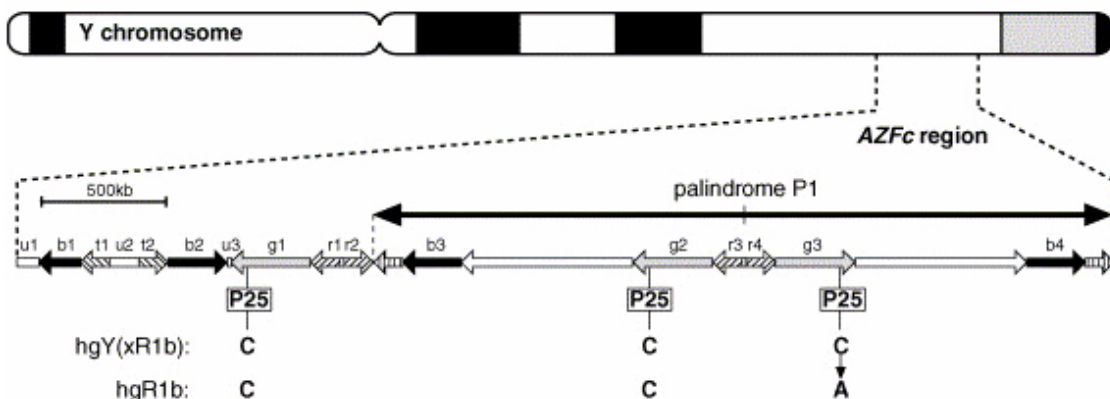


Abbildung 99: Position der AZFc Region auf dem Y-Chromosom, wo sich der PSV des SNPs P25 in drei Wiederholungseinheiten (g1, g2 und g3) befindet (Kuroda-Kawaguchi et al. 2001, King et al. 2005, Adams et al. 2006).

Die Häufigkeit einer Genkonversion pro Duplikation und Generation liegt schätzungsweise bei  $2,2 \times 10^{-4}$  (Rozen et al. 2003). Dieser Umstand bestätigt die unsichere bzw. unzuverlässige Genotypisierung des P25 PSVs, die bereits mehrfach beschrieben und diskutiert wurde (Rozen et al. 2003, Sanchez et al. 2004, Adams et al. 2006). Aus diesen Gründen erscheint der Vorschlag gerechtfertigt, die Haplogruppe R1b durch am phylogenetischen Baum untergeordnete SNPs zu bestätigen, oder gleich durch SNP M269 zu ersetzen (Adams et al. 2006).

Bei der Genotypisierung der Proben aus Asien wurden in 13 Fällen für den Locus M174 ebenfalls Doppelpeaks detektiert, wobei das mutante Allel C nur in Kombination mit Stammallel T auftauchte.

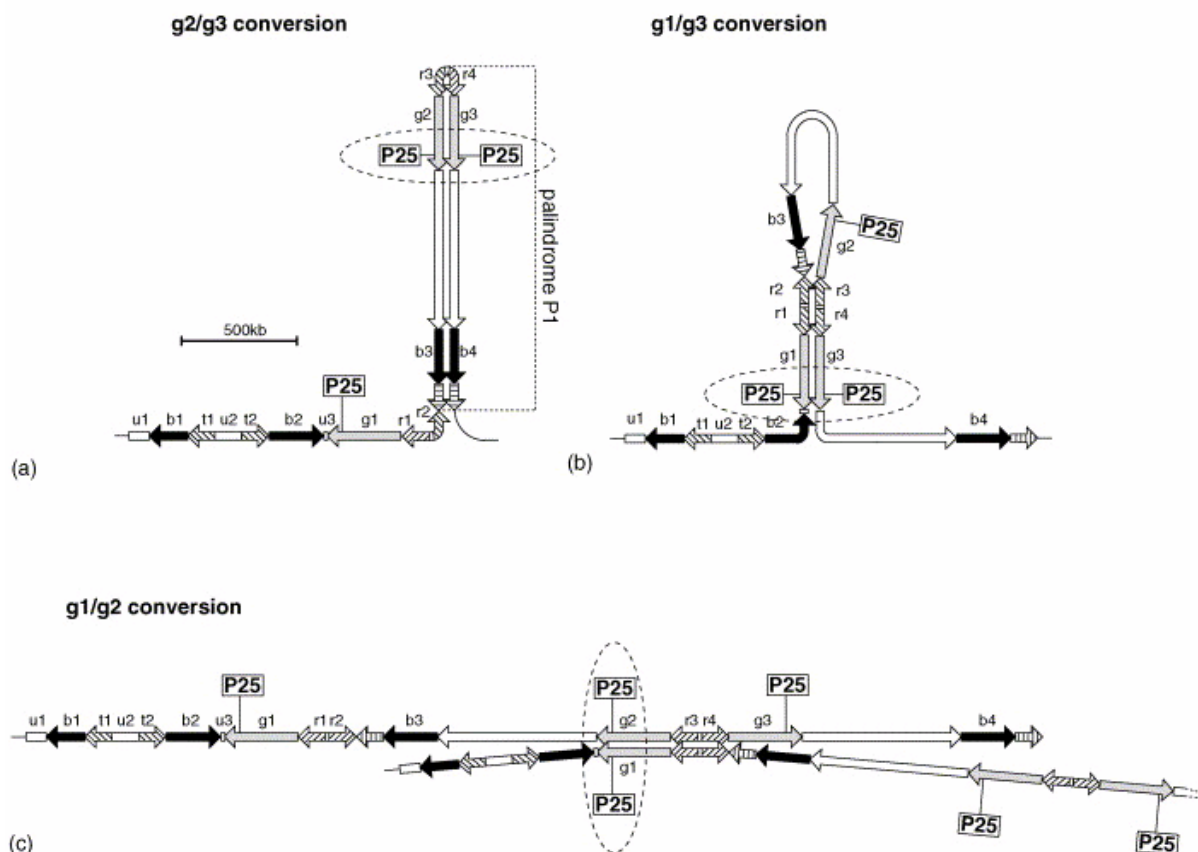


Abbildung 100: Mögliche Konfiguration der AZFc Region durch Genkonversion des PSVs (Adams et al. 2006). Durch eine hairpin-Bildung in der Palindromsequenz kann es zur Konversion zwischen g2 und g3 kommen (a), oder durch eine Schleife entsteht eine Konversion zwischen g1 und g3 (b). Weitere Möglichkeit ist die Konversion zwischen g1 und g2 indem eine ungleiche Paarung der Schwesterchromatiden stattfindet (c).

Beim Alignment der Sequenz des SNPs M174 mit Hilfe von BLAST wurde festgestellt, dass diese sich auf dem USP9Y-Gen (ubiquitin-specific protease 9) befindet, das eine homologe Sequenz auf dem X-Chromosom, dem USP9X-Gen, besitzt. Die Sequenz auf dem X-Chromosom stimmt mit der auf dem Y-Chromosom überein und die Minisequenzierungsreaktion zeigt ein T-Allel an. Aus diesem Grund wurde immer ein Allel T detektiert, auch wenn die Person ein Allel C für den Y-chromosomalen SNP M174 besitzt. Mittels Analyse einer Probe weiblicher Herkunft konnte gezeigt werden, dass die X-chromosomale Sequenz in der PCR amplifiziert wurde, indem sich ein T-Signal für Locus M174 ergab (Brion et al. 2005). Außerdem korrelierte die Signalstärke des T-Allels bei einem Test mit männlichen und weiblichen Mischproben mit dem Mischungsverhältnis der Probe. Die zwei Peaks für SNP M174 in den 13 Proben der Y-chromosomalen Validierungsstudie resultierten nicht aus einer Vermischung mit weiblicher DNA; Kontaminationen dieser Proben konnten mittels STR-Analysen ausgeschlossen werden.

Unter dem forensischen Aspekt, DNA-Fragmente kurz zu halten, ist das Redesign der PCR-Primer nicht möglich, wenn die Homologie die Länge des erwünschten PCR-Produktes überschreitet. Um eine Unsicherheit bei der Haplogruppenzuordnung zu vermeiden, sollte die Haplogruppe D durch weitere SNPs bestätigt werden.

#### 4.1.1.2 Whole Genome Amplification

Für die forensische Molekulargenetik ist die Qualität und Quantität der zu untersuchenden DNA fundamental. Geringe Probenmengen in degradiertem Zustand und die Vielzahl der durchzuführenden Analysen stellen meist die Gründe für die Limitierung dar. Die Whole Genome Amplification versprach eine sichere Methode zur allgemeinen Vervielfältigung genomischer DNA noch vor der locusspezifischen PCR (Lovmar und Syvanen 2006).

Diese Methode konnte erfolgreich eingesetzt werden, um die Quantität der über mehrere Jahre gelagerten und eingetrockneten Proben aus Japan für die Y-chromosomale Validierungsstudie zu erhöhen. Nach der WGA-Reaktion lag die DNA hochmolekular (1-12kb), in über 15-facher Konzentration vor. Dennoch war die SNaPshot™-Genotypisierung einer WGA-behandelten Probe gegenüber der Originalprobe mit deutlich mehr Problemen verbunden. Die unbalancierten, lückenhaften bzw. fehlerhaften Ergebnisse waren so unzuverlässig, dass 58% der Proben unter Einsatz der originalen, nicht WGA-behandelten DNA reanalysiert werden mussten. Diese Resultate sind Konsequenzen inhomogener Amplifikationen, die durch eine differenzielle Affinität der Phi29 Polymerase bei der WGA-Reaktion, insbesondere bezüglich geringer Qualität oder Quantität des Ausgangsmaterials entstehen und deutliche Unter- oder Überrepräsentation einzelner Marker bewirken. Diese Beobachtungen decken sich sowohl mit den Erkenntnissen, die bei zwei aufeinander aufgebauten Studien über die Verwendbarkeit von WGA-behandeltem Material für die Standard STR-Typisierung gewonnen wurden (Schneider et al. 2004, Balogh et al. 2006), als auch mit den Schlussfolgerungen anderer Arbeitsgruppen (Bergen et al. 2005, Sun et al. 2005, Barber und Foran 2006). Mittels der microarray-based comparative genomic hybridisation (aCGH) Technologie, welche chromosomale copy number variations (CNVs) über das ganze Genom verteilt analysiert, konnten die Abweichungen zwischen WGA und original DNA aufgezeigt werden (Arriola et al. 2007). Dafür wurden hochqualitative DNA-Verdünnungen (50ng/µl und 0,5ng/µl) amplifiziert und mit einer nicht amplifizierten Probe verglichen (Abbildung 101). Die Plots demonstrieren deutlich die Abweichungen zwischen WGA-behandelter und nicht WGA-behandelter Probe im gesamten Genom, die mit der Reduzierung der DNA-Konzentration erheblich zunehmen.

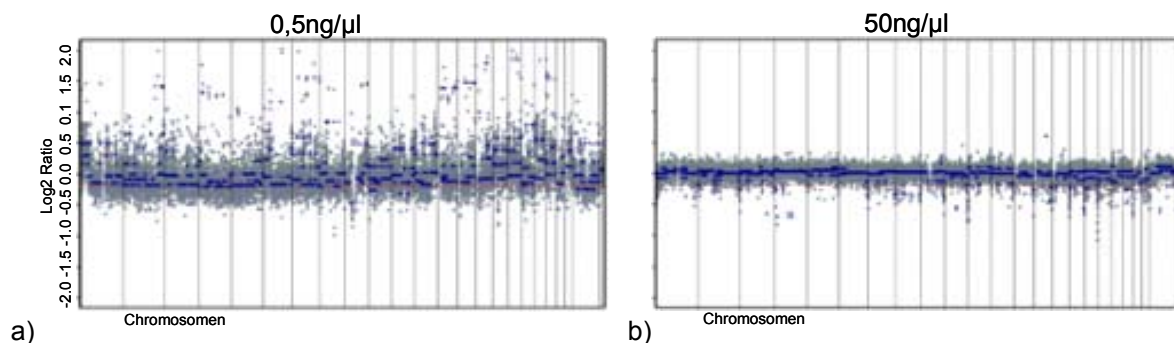


Abbildung 101: Zwei WGA-Reaktionen mit unterschiedlichen Konzentrationen des gleichen Ausgangsmaterials a) 0,5ng/µl und b) 50ng/µl, in Kombination mit nicht WGA-behandelter Referenzprobe. Die Y-Achse gibt die log<sub>2</sub> ratio von -2 bis 2 an und die X-Achse weist auf die Chromosomenlokalisation im Genom hin (Arriola et al. 2007).

Abbildung 101 zeigt die durch einen WGA-Prozess induzierten CNVs auf der genomischen Karte. Auch wenn diese Experimente die differentielle WGA bestätigen und bildlich darstellen, liefert die Arbeit außer der bereits beschriebenen Korrelation mit dem GC-Gehalt (Bredel et al. 2005) keine weiteren Erklärungen für die inhomogene Amplifikation (Arriola et al. 2007).

Die schwerwiegenden Auswirkungen der inhomogenen Reaktion der Phi29 Polymerase bei der WGA-Reaktion wurden bei der Genotypisierung der amplifizierten Proben aus Japan deutlich und zeigten über die viel berichteten Defizite der WGA-Methode im Bereich forensischer Proben hinaus auch die eingeschränkte Verwendbarkeit selbst für asserviertes Material in Standardqualität. Nur 42% der WGA-behandelten Proben eigneten sich für eine SNaPshot-Genotypisierung, obwohl die Ausgangs-DNA einwandfreie und vollständige Genotypen lieferte. Dagegen ergaben die amplifizierten Proben unbalancierte und unvollständige Genotypen mit unter- oder überrepräsentierten Markern. Die inhomogene Amplifikation wird bei der Betrachtung des SNPs M175 noch deutlicher. SNP M175 ergab für 17 WGA-Proben einen Doppelpeak, was jedoch nie bei der Analyse von originaler DNA beobachtet werden konnte. Das Allel A, das eine fünf Basen Deletion für SNP M175 als Ursache besitzt, wurde in jeder WGA-Probe detektiert, auch wenn das Allel T vorhanden war. Dabei war das Signal für das falsch-positive Allel A immer höher als das für das Stammallel T, welches in einigen Fällen nur einen Peak unter 30RFU ergab. Das stellt einen Grund dar, warum es teilweise bei der automatischen Auswertung zu einer Fehldeutung gekommen ist.

Eine mögliche Erklärung für die Doppelpeakdetektion der WGA-Proben liefert das Sequenzalignment mit Hilfe von BLAST. Dabei wurde eine Homologie für die Sequenz des SNPs M175 (Tetratricopeptide repeat protein) auf dem X-Chromosom gefunden (Information nach Celera), welche bei der Minisequenzierungsreaktion das mutante Allel A ergibt.

Die Tatsache, dass eine Doppelpeak- oder Fehldetektion nur im Zusammenhang mit einer WGA entstand, kann auf stochastische Effekte zurückgeführt werden, die aus einer differentiellen Reaktion der Phi29 Polymerase resultieren.

Die Aussicht auf Etablierung einer einfachen und zuverlässigen Methode zur Vervielfältigung des Ausgangsmaterials veranlasste zahlreiche Arbeitsgruppen zur Optimierung des WGA-Protokolls hinsichtlich seiner Eigenschaft zur differentiellen Amplifikation. Mehrere Veröffentlichungen demonstrierten, dass das Mischen mehrerer WGA-Produkte, insbesondere von geringer Ausgangskonzentration (z.B. 100pg/µl), die Homogenität der Probe verbessert, mit dem Ergebnis von weniger Ausfällen und balancierten Inter- bzw. Intra-Loci (Tzvetkov et al. 2005, Nagy et al. 2006).

Mit Hilfe einer blunt-end ligation-mediated-WGA (BL-WGA) unter Einsatz einer T4 DNA-Polymerase und einer T4 Ligase konnten kurze DNA-Fragmente erfolgreich analysiert werden und Artefakte oder Ausfälle reduziert werden (Li et al. 2006). Diese von Li et al. beschriebene DNA-Behandlung vor der WGA-Reaktion könnte die Methode hinsichtlich ihrer Anwendbarkeit in der forensischen Routine revolutionieren. Dabei wäre ihre Eignung zur Eliminierung der stochastischen Effekte und Verbesserung der Homogenität, die für LCN-DNA in der Forensik typisch ist, noch zu überprüfen (Panelli et al. 2006).

## 4.1.2 Populationsspezifische Anmerkungen

Die AMOVA-Analyse der Stichproben der gesamten Validierungsstudie ergab, dass mittels des Major Y-chromosome haplogroup typing kits hohe Varianz (33,7%,  $p < 0,001$ ) zwischen den Kontinenten und nur geringe Varianz (6,8%,  $p < 0,001$ ) zwischen den Populationen eines Kontinents zu bestimmen sind. Die höchste Varianz mit 59,5% ( $p < 0,001$ ) wurde jedoch innerhalb einer Population errechnet (Brion et al. 2005).

Insgesamt 26 der 31 Haplogruppen, die durch die 29 Y-chromosomalen SNPs definiert werden, wurden in den zwölf analysierten Populationen detektiert. Mit Ausnahme der Populationen aus Südamerika und Grönland, die keine  $F_{ST}$ -Distanzen zu den analysierten Populationen aus Europa zeigten, konnten sechs Haplogruppen nur in Proben aus Asien, drei Haplogruppen nur in Proben aus Afrika und sieben Haplogruppen nur in Proben aus Europa nachgewiesen werden.

Unter der Annahme, dass die 179 Stichproben aus China, Thailand und Japan Südostasien repräsentieren, können 93% der Proben mit südostasiatischer patrilinearer Vererbung mittels des 29 Y-chromosomalen Kits einer der sechs Haplogruppen zugeteilt werden. Nach der Haplogruppenverteilung in Europa können mit der gleichen Methode durchschnittlich 73% der Individuen mit westeuropäischer patrilinearer Vererbung (repräsentiert durch 430 Proben aus Deutschland, Dänemark und Spanien) aber nur 37% der Individuen aus der Türkei (51 Proben) den betreffenden sieben Haplogruppen zugewiesen werden. Die Haplogruppenbestimmung der Stichproben aus Afrika (repräsentiert durch 235 Proben aus Mosambique und Somalia) lässt folgern, dass das Major Y-chromosome haplogroup typing kit 97% der Individuen mit patrilinearer Vererbung bezogen auf Südafrika aber nicht auf ganz Afrika den drei Haplogruppen zuteilt.

Unter Berücksichtigung dieser Erkenntnisse ist die Haplogruppenzuordnung mittels des Major Y-chromosome haplogroup typing kits nicht mit der geographischen Abstammung gleichzusetzen. Detaillierte Informationen dazu sind in der Originalveröffentlichung (Brion et al. 2005) im Anhang 7.3 nachzulesen.

## 4.1.3 Schlussfolgerungen der Y-chromosomalen Validierungsstudie

Die Vorstellung, die geographische Herkunft eines unbekanntes Spurenlegers zu bestimmen, eröffnet neue Perspektiven in der forensischen Genetik. Eine Populationsstudie in diese Richtung ist jedoch schon im Vorfeld kritisch zu betrachten und setzt die Berücksichtigung bestimmter Grundprinzipien voraus. Ethische Aspekte sind ebenso wie das Missbrauchspotential und eine prospektive Zielbeschreibung mit definierter Sammelstrategie der Proben zu hinterfragen. Übertragen auf das hier beschriebene Projekt ist festzuhalten, dass der Schwerpunkt für die Motivation einer Y-chromosomalen Studie allein in der multizentrischen Validierung ausgewählter Y-chromosomaler SNPs liegt. Die Aussage der Y-chromosomalen Validierungsstudie ist durch die geringe Vielfalt und Anzahl der Proben eingeschränkt. Schließlich handelte es sich um Stichproben, die aus den Hauskollektiven der beteiligten Arbeitsgruppen stammten und nur nach Selbsteinschätzung einer bestimmten Population zugeordnet waren. Die oben geäußerte Vorstellung in Richtung einer reinen

Populationsstudie würde ein größeres Probenkollektiv in Verbindung mit einer qualitätsgesicherten Sammlungsstrategie erfordern. Darüber hinaus wären komplexe Problemlösungen dringend erforderlich, die nicht nur aus der Analyse des patrilinearen Y-Chromosoms bestehen, sondern weitere Informationen der matrilinearen mtDNA (Quintans et al. 2004) und populationsspezifische Erkenntnisse über autosomale Chromosomen (Phillips 2007) integrieren. Diese sogenannten Ancestry-Informative Markers (AIMs), die unter anderem durch populationspezifische SNPs (Shriver et al. 2004, Phillips et al. 2004a) oder tri-allelische SNPs (Phillips et al. 2004b) definiert sind, können bei entsprechender Anzahl Populationsdifferenzen ausmachen (Lao et al. 2008).

Als Fazit der Y-chromosomalen Validierungsstudie kann festgehalten werden, dass eine Einschätzung der geographischen Abstammung einer unbekannt Person erst möglich ist, wenn die genetischen Informationen dieser drei unabhängigen Systeme als Grundlage für die Berechnung eines kombinierten Likelihood-Quotienten verwendet werden können.

## 4.2 Diskussion zur forensischen Validierungsstudie

In den Jahren 2004 und 2005 wurden erste Arbeiten zum Einsatz von autosomalen SNP-Markern bei der Personenidentifikation veröffentlicht (Inagaki et al. 2004, Dixon et al. 2006) und lösten großes Interesse an der Etablierung und Validierung diverser SNP-Genotypisierungsmethoden aus (Huang et al. 2004, Quintans et al. 2004, Sobrino et al. 2005). Die Ansammlung von Erkenntnissen über SNPs und von Erfahrungen mit Typisierungsmethoden vor allem auch aus der Genomforschung, veranlasste die Erforschung der Verwendbarkeit dieser Marker zur Analyse von forensisch relevanten Materialien.

Nach der erfolgreichen Etablierung der SNaPshot™-Methode für die Genotypisierung der 52 autosomalen SNPs (Sanchez et al. 2006a, Sanchez et al. 2006b) erfolgte die Validierung dieser Methode für forensische Zwecke im Rahmen des SNPforID-Projektes. In einer systematischen Studie wurde die Anwendbarkeit der Methode zur Genotypisierung von forensisch relevanten Materialien wie degradierter DNA, Proben von geringer Quantität und schlechter Qualität, oder auch Mischproben, validiert. Zur Auswertung wurden die Ergebnisse mit STR-Multiplex-Systemen verglichen. Die direkte Gegenüberstellung sollte die Vor- bzw. Nachteile beider Methoden offenbaren und die theoretische Erwartung belegen, dass die SNP-Analyse besonders für die forensische Molekulargenetik strukturbedingte Vorteile mit sich bringt, wenn das Untersuchungsmaterial entsprechende Herausforderungen stellt.

### 4.2.1 Methodische Anmerkungen

Bei der Typisierung der Verdünnungsreihen und der Mischproben aus qualitativ guter Standard-DNA wurden keinerlei methodische Besonderheiten beobachtet. Die Genotypisierungserfolge dieser Proben nahmen erwartungsgemäß ab, indem die Allelausfälle sich mit der Reduzierung der DNA-Ausgangskonzentration häuften (Kapitel 3.2.1.2). Dennoch konnten diese Proben unabhängig von der DNA-Konzentration keine zufriedenstellende Genotypisierung liefern. Selbst die 1ng/µl konzentrierten Verdünnungsproben konnten nicht zu 100% genotypisiert werden, obwohl die frühere Methodvalidierung zeigte, dass selbst 500pg/µl DNA ausreicht für eine vollständige SNaPshot-Genotypisierung der 52 autosomalen SNPs (Sanchez et al. 2006b). Für die unvollständige Genotypisierung waren einzelne SNPs verantwortlich, die in der Mehrheit der Analysen kontinuierlich ausgefallen sind. Die eindeutige Beobachtung systematischer Ausfälle einzelner SNPs bei den Mischproben und selbst bei DNA aus frischen Speichelproben stärkte den Verdacht, dass diese durch Probleme mit einzelnen Reagenzien, wie PCR-Primer oder Analyseprimer, verursacht wurden. Die Bestätigung brachte schließlich die Prüfung der PCR-Primer mittels elektronischer Microarray-Technologie, indem Amplifikationen mit alten und mit neuen Primern gegenübergestellt wurden.

Über die allgemein schwache Detektion hinaus bereiteten die künstlich degradierten Proben, die gleichzeitig mangelnde Qualität und Quantität aufwiesen, erhebliche methodische Schwierigkeiten bei der SNaPshot-Genotypisierung. Oftmals mussten die elektro-phoretischen Rohdaten gesichtet und manuell editiert werden. Insbesondere machten zwei Phänomene eine automatische Auswertung unmöglich: Zum einen wurde eine Art Agglomeration an starken Signalen am Anfang eines Laufes, die aber rapide abnahmen bzw. ausfielen, beobachtet (Abbildung 36 in Kapitel 3.2.1.1). Zum anderen ist es häufiger zu einer Laufverschiebung der Signale gekommen (Abbildung 35 in Kapitel 3.2.1.1). Das Verschiebungsphänomen konnte trotz verschiedener Optimierungsmaßnahmen nicht eliminiert werden. Theoretisch konnten die PCR, die Aufreinigung und die SNaPshot-Reaktionen als Verursacher ausgeschlossen werden, denn deren Wiederholung brachte keine Verbesserung. Nicht zuletzt wurden die Reaktionen auch immer für mehrere Proben aus demselben Reagenzien-Mix angesetzt, und dennoch konnte keine Kontinuität innerhalb eines solchen Ansatzes festgestellt werden. Gleichzeitig konnte ebenfalls kein technischer Fehler im Bereich der Software oder der Einstellung entdeckt werden. Instrumentenspezifische Ursachen für die Verschiebungen konnten ausgeschlossen werden, indem die kapillarelektrophoretische Auftrennung derselben SNaPshot-Proben im rechtsmedizinischen Institut Köln wiederholt wurde, mit dem gleichen Ergebnis wie in Mainz.

Das Phänomen der Agglomeration kann evtl. auch auf die Gegebenheiten der Proben zurückgeführt werden. Die geringe Quantität im Zusammenhang mit dem qualitativ schlechten Zustand der natürlich degradierten Proben konnten als Ursache ausgemacht und beseitigt werden. Durch die Erhöhung der DNA-Ausgangskonzentration mittels Vakuum und damit der eingesetzten DNA-Menge (auf 1,2ng) für die PCR-Amplifikation verschwand dieses Phänomen, was eine automatische Auswertung ermöglichte (Abbildung 36 in Kapitel 3.2.1.1).

Daraus folgt für die natürlich degradierten Proben, dass die methodischen Schwierigkeiten der SNaPshot-Analyse zunehmen, wenn geringe DNA-Quantität gepaart mit schlechter Qualität bei einer Probe auftreten.

Weitere Optimierungen vorzunehmen war an dieser Stelle nicht möglich, denn ausgehend von dem ursprünglichen Ziel, eine bereits etablierte Methode für forensische Einsätze zu validieren, wurden die Proben nur in limitiertem Volumen bereitgestellt und sollten wie Spurenmaterial in der Routine bearbeitet werden.

## **4.2.2 Gesamtdiskussion der forensischen Validierungsstudie**

Zur Sensitivitätsbestimmung der 52plex-SNaPshot-Methode mittels der Verdünnungsreihen lieferten beide beteiligten Laboratorien wesentlich niedrigere Genotypisierungsraten als die STR-Analysen. Diese Ergebnisse sind widersprüchlich zu den bisher gesammelten Erkenntnissen über die Genotypisierung der 52 autosomalen SNPs aller beteiligten Arbeitsgruppen (Sanchez et al. 2006b). Darüber hinaus meldeten beide Laboratorien durchgehend reagenzienbezogene Probleme (z.B. Kapitel 3.2.2). Unter Berücksichtigung dieser zwei Aspekte sollten die Ergebnisse zur Sensitivität nur relativiert und einzelfallbezogen betrachtet werden.



Im Gegensatz dazu konnten alle qualitativ schlechten Proben, die sich zusammensetzten aus künstlich degradierter DNA, zersetztem Muskelgewebe, Haarschaft-Extrakten oder auch aus natürlich degradierten Proben, die zusätzlich noch eine geringe Quantität aufwiesen, im Vergleich zur STR-Typisierung höhere Genotypisierungserfolge erreichen. Diese Ergebnisse weisen eindeutige Vorteile der SNP-Genotypisierungsmethode gegenüber der in der forensischen Routine verwendeten STR-Methode nach und versprechen gerade bei Analysen von qualitativ schlechtem Material eine mögliche Alternative.

Verteilt über die gesamte Validierungsstudie zeigten sich auffällig große Unterschiede zwischen den fünf beteiligten Laboratorien. Das heißt, unabhängig vom Ausgangsmaterial, lieferten einige Laboratorien konsequent hohe Genotypisierungsraten bei niedriger Artefaktzahl, während andere im Vergleich regelmäßig niedrige Genotypisierungsquoten meldeten.

Diese Erkenntnisse weisen auf die hohe Anfälligkeit der 52plex-SNaPshot-Genotypisierung hin. Nur bei optimalen Gegebenheiten ist auch eine optimale Nutzung der Methode garantiert. Faktoren wie technische Laborausrüstung, einwandfreie Reagenzien und nicht zuletzt Erfahrung bei der Bearbeitung von forensisch relevanten Proben spielen dabei eine wesentliche Rolle.

### **4.2.3 Schlussfolgerungen der forensischen Validierungsstudie**

Die Relevanz der strukturbedingten Vorteile der SNPs gegenüber den STR-Markern für die forensische Spurenkunde führte bereits zu zahlreichen theoretischen Diskussionen (Gill 2001b, Gill et al. 2004, Amorim und Pereira 2005, Dixon et al. 2006). Die vom SNPforID-Konsortium durchgeführte systematische Studie zur Validierung der 52 autosomalen SNPs und der SNaPshot-Genotypisierungsmethode für forensisch relevante Proben lieferte dazu eine Reihe von praktischen Erkenntnissen (Musgrave-Brown et al. 2007).

Die Ergebnisse der forensischen Validierungsstudie erbrachten den Beleg, dass die SNP-Analyse zur Bearbeitung von Proben in qualitativ schlechtem Zustand einen größeren Genotypisierungserfolg verspricht. Damit kann die Ergänzung bzw. der Ersatz der STR-Analyse durch die SNP-Typisierung in solchen Fällen gewinnbringende Bedeutung haben. Neueste Veröffentlichungen unterstreichen diese Erkenntnisse, indem sie den SNPs deutliche Vorteile gegenüber den STR-Markern bei der Untersuchung von schwierigem Ausgangsmaterial zuweisen (Dixon et al. 2006, Roper et al. 2007). Weitere Arbeiten zeigten sogar, dass die SNP-Analyse in besonderen Fällen unbedingt erforderlich sein kann sowohl zur Lösung von Identifikationsfragen oder ambivalenten Verwandtschaftsanalysen (Fondevila et al. 2008, Lareu et al. 2008) als auch bei anthropologischen Fragen (Bouakaze et al. 2007). Gleichzeitig lieferte die Studie den Beweis dafür, dass die SNP-Genotypisierung nicht in allen Bereichen der Spurenkunde eine Alternative darstellt. So konnte zum Beispiel bei der Untersuchung von Mischproben keine Korrelation zwischen Signalstärke und Probe festgestellt und damit auch keinerlei Rückschluss auf die Mischungsverhältnisse gezogen werden. Darüber hinaus verlangt die Multiplex-SNP-Typisierung viel Erfahrung, in erster Linie bezüglich eines präzisen Oligodesigns. Dabei stellt die große Anzahl der Primer, die für eine adäquate Multiplex-Analyse nötig sind, eine Herausforderung dar. Diese müssen nicht

nur alle unter einheitlichen Konditionen funktionieren, sondern sie dürfen auch keine Interaktionen wie z.B. Heterodimere bilden, die gerade bei der Analyse von niedrig konzentrierten Proben stören würden. Zusätzlich erfordert die SNP-Analyse viel Praxis und technische Erfahrung. Außerdem sind die optimalen Gegebenheiten des Labors gerade zur Vermeidung von Artefaktbildung von großer Bedeutung.

Abschließend kann zusammengefasst werden, dass die forensische Validierungsstudie anwendungsbezogene Vorteile der SNP-Analyse in der forensischen Spurenkunde gezeigt hat. Gleichzeitig wurden die Sensibilität und auch die Anfälligkeit der Methode sichtbar, welche die Bedeutung optimaler Analysebedingungen hervorhebt.

## 4.3 Diskussion zur elektronischen Microarray-Technologie

Die Nanogen-Plattform stellt nicht nur die einzige elektronische Microarray-Genotypisierungsmethode dar, die im Rahmen des SNPforID-Projektes erforscht wurde, sondern sie ist die erste elektronische Microarray-Technologie überhaupt, die für forensische Analysen etabliert wurde. Obwohl für die Nanogen-Plattform bereits kommerziell erhältliche ASR (Analyte Specific Reagents) Kits<sup>11</sup> zur Detektion von SNPs, Insertionen, Deletionen existieren, wurde die Methode bisher nicht auf ihre Einsetzbarkeit für forensische Analysen überprüft. Dieses Arbeitsfeld verlangt von einer Analyseplattform spezielle Eigenschaften wie Multiplexfähigkeit, Effizienz, Schnelligkeit, Sensitivität, Reproduzierbarkeit und Automation im Sinne einer Hochdurchsatzkapazität. Vor diesem Hintergrund wurde in der vorliegenden Arbeit die Etablierung dreier unterschiedlich aufgebauter Assays (Amplicon down, Capture down und APEX) behandelt. Anschließend wurde das für forensische Zwecke besonders geeignete Capture down-Assay optimiert und validiert.

### 4.3.1 Amplicon down-Assay

Der Aufbau des Amplicon down-Assays basiert auf der Adressierung biotinylierter PCR-Produkte, gefolgt von einer Hybridisierung der Stabilizer und Reporter-Oligos (Kapitel 2.2.6.3.1). Mit der Etablierung des Assays konnte erst begonnen werden, nachdem von dem SNPforID-Konsortium 23 autosomale SNPs ausgewählt und in eine Multiplex-PCR zusammengefasst wurden. Die ersten fünf SNPs aus dem Multiplex, die die theoretischen Designkriterien erfüllten, wurden für die Etablierung des Amplicon down-Assays eingesetzt. Schon die ersten Designversuche zeigten, dass die Kriterien nur Richtlinien darstellen, die nicht kompromisslos befolgt werden können. Für die fünf SNPs konnten störende Schleifen in der Sekundärstruktur, insbesondere im Bereich der Reporter verhindert werden, aber die BSE (Base Stacking Energy) zwischen den Reportern und dem Stabilizer konnten nicht immer klein und ausgewogen gehalten werden. So ist die BSE des Wildtyps für SNP01 mit -5,37 unerwünscht hoch und der Unterschied zwischen Wildtyp mit -14,59 BSE und Mutante mit -8,26 BSE des SNP09 sehr groß (Tabelle 22 in Kapitel 2.2.6.4). Auch beim Reporterdesign liegt  $\Delta T_m$  von 1,7°C für SNP11 etwas hoch. Die Konsequenz einer hohen Temperaturdifferenz ist die ungleiche Diskriminierung der Wildtyp- und Mutante-Reporter wie in Abbildung 49 (Kapitel 3.3.1) dargestellt. Während der Etablierung des Amplicon down-Assays, bei dem die Analysen der SNPs einer Multiplex sukzessiv und nicht gleichzeitig erfolgen, konnte gezeigt werden, dass eine Heterozygotenkontrolle solche kleinen Abweichungen beim Design korrigieren kann.

Das Multiplex PCR-Standardprotokoll sieht einen ungewöhnlich hohen  $MgCl_2$  Gehalt (Endkonzentration 8mM) vor. Um zu verhindern, dass die hohen Salzkonzentrationen später bei der elektronischen Adressierung stören, wurden die PCR-Produkte aufgereinigt. Die passende Aufreinigungsmethode zu finden wurde durch die Besonderheit erschwert, dass

---

<sup>11</sup> [www.nanogen.com](http://www.nanogen.com)

die Produkte bis zu 68bp kurz sind. Erst nach dem Vergleich der Methoden, die auf unterschiedlichen Prinzipien basieren, konnten reproduzierbare Differenzen aufgezeigt werden. Die Methoden von Qiagen und ChargeSwitch, die durch zusätzliche Waschschriffe die vollständige Entsalzung der Probe garantieren, führen zu erheblichen Produktverlusten. Ein gutes Ergebnis hingegen zeigte das Millipore Vacuum-System, das die günstigste Aufreinigungsmethode darstellt. Nach der einmaligen Anschaffung eines Vacuum Separations Systems bedarf diese Methode nur noch der 96-er Mikrotiterplatten mit Filtermembranboden. Dennoch konnte sich diese Methode wegen ihrer Abhängigkeit vom Zustand der Filtermembrane in der Mikrotiterplatte nicht durchsetzen. Um reproduzierbar gute Ergebnisse zu gewährleisten muss bei jedem Filtern eine neue Mikrotiterplatte angebrochen werden, unabhängig von der Anzahl der zu reinigenden PCR-Produkte. Das liegt daran, dass sich die Beschaffenheit der Filtermembran nach Erstnutzung dahingehend verändert, dass es zu erhöhten Produktverlusten oder eingeschränkter Entsalzung kommt. Zur Entscheidung, die Millipore-Aufreinigung nicht weiterhin als validierte Methode einzusetzen, führten auch die Erkenntnisse aus der Etablierung des Capture down-Assays (Kapitel 3.3.3). Während die biotinylierten Pentaplex-Proben für das Amplicon down-Assay mittels Millipore aufgereinigt werden konnten, sind die nicht-biotinylierten Pentaplex-Produkte für das Capture down-Assay vollständig verloren gegangen. Eine Aufreinigung von nicht-biotinylierten Produkten konnte erst bei einem Multiplex von 13 SNPs erfolgen. Die Abhängigkeit von der Multiplexgröße zeigt, dass diese Methode unzuverlässig ist, sie verlangt bei jeder PCR-Veränderung eine neue Überprüfung und Bewertung. Diese Unschärfe war bei der Aufreinigung mittels Invitex-Methode nicht zu beobachten. Damit stand abschließend fest, dass die Säulenaufreinigung nach Invitex den Anforderungen am besten entspricht. Die Aufreinigung mittels dieser Methode befreit die PCR von störenden Salzen, führt gleichzeitig zu den geringsten Fragmentverlusten und die Ergebnisse können reproduziert werden.

Für die Adressierung der PCR-Produkte, die elektronisch in 90 Sekunden erfolgt, wurde die ganze 50µl Probe verwendet. Die darauf folgende Hybridisierung und Diskriminierung benötigt pro SNP maximal 15 Minuten. Demnach ist die Genotypisierung einer Person für fünf SNPs mittels Amplicon down-Assay innerhalb von 1,5 Stunden erbracht und das Ergebnis kann zeitnah abgelesen werden. Die kurze Typisierungsdauer von fünf SNPs ergibt sich nicht zuletzt auch daraus, dass beim hier etablierten Amplicon down-Assay eine elektronische Hybridisierung des Stabilizers jeweils vor jedem Reportingschritt entfällt. Auf diesen von der Firma Nanogen empfohlenen und durch Veröffentlichungen (Feng 1999, Gilles et al. 1999, Sosnowski et al. 2002) bestätigten Schritt, der eine Multiplexanalyse erheblich erschwert und verlängert, kann verzichtet werden, indem der Stabilizer mit den zwei Reportern in einem Schritt passiv hybridisiert wird.

Die Kapazität des Mikrochips eignet sich anhand des Amplicon down-Protokolls für eine simultane Genotypisierung von 98 verschiedenen Personen mit fünf SNPs. Dabei erhöht sich nur die Zeit der PCR-Adressierung auf 98x90 Sekunden. Das bedeutet, dass in vier Stunden 98 Individuen für fünf SNPs genotypisiert werden können und jeder weitere SNP den Prozess um 15 Minuten verlängert. Damit zeigt sich, dass die Vorteile dieses Assays darin liegen, viele Personen für wenige SNPs in kürzester Zeit mit geringem Aufwand zu typisieren.

Ebenfalls kommt die hohe Sensitivität der Methode zu Gute. Während der Amplicon down-Assay etablierung wurde zielorientiert das ganze zur Verfügung stehende PCR-Produkt eingesetzt. Die Sensitivität der Microarray-Technologie konnte aber in einer parallelen Studie in Kopenhagen mittels Y-chromosomalen SNPs gezeigt werden (Borsting et al. 2004). Die fehlende Heterozygotie der Y-chromosomalen SNPs determinierte die maximale Sensitivität des Systems. Es wurden gemäß des Amplicon down-Assays fünf SNPs einer Multiplex auf einem Pad nacheinander detektiert. Bei dieser Studie konnten je 100pg DNA von 400 männlichen Individuen mit 100%-iger Genauigkeit typisiert werden. Zusätzlich konnte ermittelt werden, dass die Sensitivität mit der Länge der zu untersuchenden Amplikons im Zusammenhang steht. Deutliche Unterschiede wurden festgestellt bei der Typisierung derselben SNPs in zwei unterschiedlich langen PCR-Produkten. Für die Detektion eines 50bp langen Fragments wurden nur 10-30fmol PCR-Produkt und für ein 200bp langes Fragment mehr als 200fmol benötigt (Borsting et al. 2004). Die PCR-Produkte, die im Rahmen dieser Arbeit für das Amplicon down-Assay etabliert wurden, liegen zwischen 68bp und 95bp und benötigen damit erwartungsgemäß wenig PCR-Volumen. Die Bestätigung dafür liefert die Optimierung des Capture down-Assays (Kapitel 3.3.3.2), welches bereits 1µl PCR-Produkt typisieren kann.

Nach der Etablierung des Readerprotokolls erfolgt die thermische Diskriminierung der Reporter sehr selektiv, ohne Anzeichen auf Kreuzkontaminationen. Die fünf SNPs können mit großer Zuverlässigkeit bei drei Diskriminierungstemperaturen nacheinander gescannt werden. Die dazwischen liegenden thermischen und passiven Strippingschritte garantieren, dass die PCR-Produkte für die nächste passive Hybridisierung reporterfrei bereitliegen. Bei der Etablierung des Amplicon down-Assays für fünf autosomale SNPs konnte keine Korrelation zwischen der Signalstärke und der Anzahl der aufeinander folgenden Hybridisierungen festgestellt werden. Vielmehr waren die Fluoreszenzintensitäten SNP-abhängig, so liegen die Signale bei der zweiten Hybridisierung von SNP02 meist unter 1500RFU (Abbildung 51b, Kapitel 3.3.1), während die dritte Hybridisierung von SNP05 meist über 2400RFU hinausgeht (Abbildung 51c, Kapitel 3.3.1). Von der Arbeitsgruppe in Kopenhagen hingegen wurde im Laufe der Optimierung festgestellt, dass die Fluoreszenzintensität mit jeder neuen Hybridisierung exponentiell abnimmt, sodass schätzungsweise nicht mehr als 10-15 SNPs auf einem Pad unter Verwendung des Amplicon down-Assays analysierbar sind (Borsting et al. 2004).

Als Schlussfolgerung bleibt festzustellen, dass das Amplicon down-Assay nicht einsetzbar ist für die simultane Genotypisierung von vielen SNPs, gleichzeitig eignet es sich aber für die Typisierung einer hohen Individuenzahl in einer Analyse. Dabei ist die hier etablierte Methode zuverlässig und in der Handhabung schnell und einfach. Aufgrund dieser Eigenschaften, verbunden mit der leichten automatischen Auswertung, könnte diese Methode effizient eingesetzt werden, zum Beispiel für einen schnellen Vorausschluss bei einer Massentypisierung, um den Verdächtigenkreis für spezifischere und daher aufwendigere Identifizierungsanalysen einzuengen.

### 4.3.2 APEX-Assay

Der Aufbau des APEX-Assays basiert auf der Adressierung der Capture-Sonden mit anschließender Hybridisierung des PCR-Produktes und bildet die Voraussetzung für die Einzelbasen-Extension. Dafür wurden die Capture-Sonden verwendet, die nach den Kriterien für den Capture down-Assay entwickelt wurden. Um nicht ein Stripping des PCR-Produktes zu bewirken, wurde die Extension bei einer Temperatur von 48°C durchgeführt und nicht wie empfohlen bei 52°C (Tonisson et al. 2000, Tonisson et al. 2002). Obwohl bei den Experimenten nicht die optimalen Bedingungen für die DNA-Polymerase geschaffen werden konnten, lieferte das APEX-Assay einige positive Ergebnisse. Zur Optimierung müssten jedoch für den APEX-Assay die Capture-Oligos nach speziellen Kriterien entworfen werden. Auf diesen Aufwand wurde jedoch verzichtet, da kein Interesse daran bestehen kann, eine SBE-Standardmethode (Jaakson et al. 2003) auf einem elektronischen Microarray durchzuführen und dabei die Eigenschaften der Plattform nicht zu nutzen.

Die zwei APEX-Experimente ohne große Investition und Zeitaufwand, reichten aus, um die Flexibilität der elektronischen Microarray-Technologie zu zeigen. Die Nanogen-Technologie liegt vorteilhaft als offenes System vor und gewährt dem Nutzer völlige Freiheit bei der Gestaltung des Assays je nach Genotypisierungsschwerpunkt (Sosnowski et al. 2002, Ferrari et al. 2003).

Der kurze Exkurs ergänzt und untermauert das Potential der Nanogen-Technologie für verschiedene Anwendungen, indem unterschiedlich aufgebaute Assays etabliert werden (Feng 1999, Ferrari et al. 2003). Dazu gehören unter anderem die Identifikation von STR-Allelen (Feng 1999, Radtkey et al. 2000), Protein-Assays z.B. zur direkten Detektion von Staphylokokkus Enterotoxin B (SEB) und Cholera Toxin B (CTB) (Ewalt et al. 2001) und auch die Strand Displacement Amplifikation (SDA) direkt auf dem Mikrochip (Westin et al. 2000, Huang et al. 2004).

### 4.3.3 Capture down-Assay

#### 4.3.3.1 Methodische Anmerkungen

Der Aufbau des Capture down-Assays basiert auf einer elektronischen Adressierung der Capture-Sonden, gefolgt von der elektronischen Hybridisierung des PCR-Produktes und der passiven Hybridisierung der Reporter. Nach zahlreichen Optimierungen steht ein etabliertes Capture down-Assay für 32 autosomale SNPs zu Verfügung, die im Durchschnitt mit 88%-iger Sicherheit analysiert werden können (Tabelle 32 in Kapitel 3.3.3.3). Die Genotypisierung von 40 Personen mittels des validierten 32plex Capture down-Assays ergibt eine 94%-ige Übereinstimmung mit der SNaPshot-Referenzmethode (Tabelle 36 in Kapitel 3.3.3.3).

Trotz zahlreicher Reporter- und Capture-Redesigns erfüllten von 52 autosomalen SNPs nur 35 SNPs sowohl die theoretischen Designkriterien als auch die praktischen Erwartungen. Daraus resultiert, dass die Adaptation an das elektronische Microarray nur für 67% der ausgewählten SNPs erfolgreich war. Die Konsequenz ist, dass ca. 33% mehr SNPs als die

Zielzahl zur Auswahl stehen sollten um unter Berücksichtigung der Designkriterien eine bestimmte Multiplexgröße zu erreichen.

Darüber hinaus weisen die Ergebnisse darauf hin, dass der Genotypisierungserfolg unmittelbar von der Qualität der PCR-Produkte abhängt. Alle Produkte einer Multiplex-PCR müssen ausgewogen sein und in ausreichender Konzentration vorliegen, um zuverlässige Ergebnisse zu erzielen. Aus diesem Grund ist die Analyse von geringer oder degradiert DNA nur dann erfolgreich, wenn alle PCR-Produkte einer Multiplex-PCR in entsprechender Menge und balanciert amplifiziert sind. PCR-Optimierungen anhand von Zusätzen oder asymmetrischer Amplifikation brachten nur unwesentliche Verbesserungen. Entsprechend der PCR-Konzentrationsbestimmung mittels der Agilent-Technologie konnte das validierte PCR-Standardprotokoll optimiert werden, indem die Primerkonzentrationen für einzelne SNPs erhöht oder reduziert wurden (Tabelle 9 in Kapitel 2.2.4.4). Beim ersten Schritt des validierten Capture down-Assays werden Capture-Sonden für 32 SNPs jeweils auf drei Pads eines Microarrays adressiert. Durch die Verringerung des Multiplexes von 35 auf 32 SNPs können auf einem Microarray drei PCR-Produkte simultan typisiert werden und es bleiben noch vier Pads für die notwendige Backgroundbestimmung auf dem Array übrig.

Der statistische Informationsgehalt einer 32plex-Analyse im Vergleich zu einer 52plex ist geringer. Die Frequenzdaten der 52 SNPs für die vier Populationsgruppen (Europa, Afrika, Amerika und Südasien) wurden einer im Internet frei verfügbaren Datenbank<sup>12</sup> entnommen, die auf den Daten des SNPforID-Projektes basiert (Amigo et al. 2008). Während die Wahrscheinlichkeit, dass eine zufällig gewählte Person einen vordefinierten Genotyp für die 52 SNPs aufweist, zwischen  $9,8 \times 10^{-19}$  (europäische Bevölkerung) und  $9,4 \times 10^{-24}$  (afrikanische Bevölkerung) liegt, erreichen die 32 SNPs eine Wahrscheinlichkeit zwischen  $4,6 \times 10^{-12}$  und  $8,4 \times 10^{-16}$  (Tabelle 38). Dennoch ist die Wahrscheinlichkeit, dass jemand, der zufällig aus einer Population herausgegriffen würde, den vorgegebenen Genotyp nicht aufweist, auch für die 32 SNPs und alle vier Bevölkerungsgruppen über 99,99999%.

Tabelle 38: Informationsgehalt des 32plexes im Vergleich zu der 52plex-Analyse.

Populationen <sup>1)</sup>	PM <sup>2)</sup>		PD <sup>3)</sup>	
	52plex	32plex	52plex	32plex
<b>Europa</b> n=750	9,77486E-19	4,62073E-12	>99,99999%	>99,99999%
<b>Afrika</b> n=407	9,4289E-24	8,40306E-16	>99,99999%	>99,99999%
<b>Amerika</b> n=328	3,25563E-21	1,32626E-13	>99,99999%	>99,99999%
<b>Südasien</b> n=146	2,07456E-23	2,05034E-14	>99,99999%	>99,99999%

1) Frequenzdaten zu den Populationen aus SNPforID browser entnommen.

2) Probability of a match (PM) gesamt für 52plex bzw. 32plex.

3) Power of discrimination (PD) gesamt für 52plex bzw. 32plex.

Trotz der Reduktion des Multiplexes auf nur 32 SNPs erreichen diese eine statistische Identifizierungskraft von einer Person aus 216 Milliarden mit europäischer Abstammung und sogar einer aus 1,2 Billionen für Proben aus Afrika.

Die 32 Capture-Adressierungen erfolgen automatisch innerhalb einer Stunde und können auf Vorrat erfolgen. Die Dauer der Adressierung verlängert sich nur unwesentlich, wenn parallel

<sup>12</sup> <http://spsmart.cesga.es/>

zwei, drei oder vier Microarrays beladen werden. Zur Adressierung werden maximal 100nM Capture-Oligo pro Microarray benötigt. Die Erhöhung der Capture-Oligo-Konzentration hat keinen Einfluss auf den Genotypisierungserfolg, hingegen könnte sie die Kontaminationsgefahr steigern. Die elektronische Adressierung von nur 100nM Capture mit Unterstützung des His-Puffers kann so selektiv durchgeführt werden, dass auf benachbarten Pads keine Kreuzkontaminationen zu beobachten sind.

Im zweiten Schritt des validierten Capture down-Assays werden jeweils drei verschiedene PCR-Produkte auf einem Microarray elektronisch hybridisiert. Die elektronische Hybridisierung war nicht nur die entscheidende Voraussetzung für eine parallele Analyse von drei PCR-Produkten, sondern zeigte auch gegenüber der passiven Hybridisierung Signale von höherer Fluoreszenzintensität. Andere Arbeitsgruppen beobachteten für die elektronische Hybridisierung sogar eine 30-40fach höhere Rate im Vergleich zur passiven Hybridisierung (Edman et al. 1997). Dabei konnten die besten Ergebnisse erzielt werden, indem 1µl ungereinigtes PCR-Produkt in Arg-Puffer sukzessiv dreimal hybridisiert. Durch den vollautomatischen Ablauf lassen sich Multiplexe von drei Personen auf einem Microarray innerhalb von 20 Minuten hybridisieren. Die Tatsache, dass das Capture down-Assay nach der erfolgreichen Validierung keine PCR-Aufreinigung mehr benötigt und bereits 3µl an Produkt für eine zuverlässige Genotypisierung ausreichen, erfüllt einige Erwartungen an eine Hochdurchsatzmethode. So wird die Schnelligkeit durch die Reduzierung des Assayablaufs um einen kompletten Arbeitsschritt erhöht. Ebenfalls entfällt dadurch die Gefahr des Produktverlustes und durch Reduzierung manueller Handhabungen verringern sich auch die Fehler- oder Kontaminationsquellen. Darüber hinaus bleibt nach einer Genotypisierung mittels Capture down-Assay so viel Restprodukt übrig, dass bei Unstimmigkeiten die Analyse wiederholt werden kann und eine Option auf begleitende oder ergänzende Untersuchungen erhalten bleibt. Da die Assayvalidierung unter Einsatz von  $\geq 10$ ng DNA in 25µl Volumen erfolgte, kann keine Aussage getroffen werden über die Typisierung von Proben mit geringer DNA-Konzentration oder im degradierten Zustand. Aber die erfolgreiche Reduzierung des PCR-Produkteinsatzes auf 3µl kann als Zeichen gedeutet werden, dass die Nanogen-Plattform ein sensibles System darstellt, das nur gering konzentrierte PCR-Produkte zur vollständigen Genotypisierung benötigt (Borsting et al. 2004).

Beim dritten Arbeitsschritt, der passiven Hybridisierung der Diskriminatoren, nimmt die Verwendung von zwei fluoreszenzmarkierten Sonden mit Random-Sequenzen (Goering 2003) als Universal Reporter keinen Einfluss auf die Dauer der Hybridisierung, da alle nötigen Diskriminatoren und die zwei Universal Reporter in HSB vor der Hybridisierung gemischt werden. Ebenfalls zeigen sie keine negativen Auswirkungen auf den Genotypisierungserfolg. Im Gegensatz dazu senken sie deutlich die anfallenden Kosten. Das bedeutet, dass mit dem Einsatz von Universal Reportern nur zwei markierte Oligos notwendig sind, wodurch die kostspielige Fluoreszenzmarkierung aller 64 Oligos, die für eine 32plex-Genotypisierung benötigt werden, erspart bleibt. Damit verringert diese Methode die Kosten des Capture down-Assays erheblich, ohne die Zuverlässigkeit des Assay zu beeinträchtigen.

Die größte Herausforderung des Capture down-Assays ist das Design der Diskriminatoren. Dabei sind die genaue Erfüllung theoretischer Designkriterien und die praktische Erprobung von gleicher Bedeutung. Während der Assayoptimierung konnten anhand des Diskriminatoren-Redesigns mehrere SNPs im Hinblick auf die Balance zwischen intra- und



inter-SNPs optimiert werden. Das heißt, damit wurde nicht nur erreicht, dass innerhalb eines SNPs die zwei Diskriminatoren bei gleicher Temperatur diskriminieren, sondern auch, dass sich alle SNPs einer 32plex unter bestimmten Konditionen annähernd gleich verhalten. Durch eine theoretische und praktische Validierung des Designs wurde die Effizienz der Plattform dahingehend erhöht, dass 32 SNPs parallel nach einer einzigen kurzen Hybridisierung mit großer Genauigkeit analysiert werden können. Der abschließende Arbeitsschritt, die Diskriminierung mit Detektion erfolgt automatisch nach vordefiniertem Readerprotokoll. Nach zahlreichen Experimenten und Optimierungen gelang es ein Protokoll zu validieren, das nur einen Scan nach mehreren Temperatur- und Waschschritten vorsieht. Dadurch konnte die Zuverlässigkeit des Capture down-Assays maßgeblich erhöht werden. Denn mehrere Scans bei unterschiedlichen Temperaturen bedeuten nicht nur, dass die SNPs einer Multiplex nicht alle bei einer Temperatur auswertbar sind, sondern es besteht die Gefahr einer ungenauen Auswertung. Dabei kann zur Auswertung ein Scan zu früh gewählt werden, wenn die Diskriminierung noch nicht abgeschlossen ist und fälschlicherweise z.B. ein homozygoter Genotyp für heterozygot gehalten wird, oder verspätet, wenn es bereits zu Oligoverlust gekommen ist und ein heterozygoter Genotyp als homozygot gedeutet wird. Obwohl das validierte Readerprotokoll sich aus zwei Protokollen, Touch down- und Diskriminierungsprotokoll, zusammensetzt und aus vielen Temperatur- und Waschschritten besteht, liefert es in nur 15 Minuten die Ergebnisse in einer Form, die eine automatische Auswertung in Sekunden ermöglicht. Die Rohdaten können anhand des Konvertierungsmakros und des Genotypisierungsmakros ebenso schnell nach statistisch ermittelter Methode ausgewertet werden.

Die Genauigkeit und Zuverlässigkeit der Auswertung wurde durch die Definition einer optimalen Background-Korrektur erheblich gesteigert. Das validierte Capture down-Assay sieht vor, ein Pad als Background zu wählen, welches nicht bei der Capture-Adressierung, dafür aber bei der PCR-Produkt Hybridisierung aktiviert wird. Da sich die PCR-Produkte in einem Multiplex befinden, kann die elektronische Hybridisierung nicht selektiv erfolgen, sondern es werden alle Pads, die 32 verschiedene Capture-Oligos enthalten, gleichzeitig aktiviert. Die Produkte, die ohne Komplementärsequenz möglicherweise an dem Pad haften bleiben und das Ergebnis verfälschen könnten, werden durch den Background-Pad aufgezeigt und können von der Probe abgezogen werden. Durch diese Maßnahme ist das Signal des Backgrounds zwar höher als das Signal eines Pads, das keine Capture-Adressierung und keine PCR-Hybridisierung beinhaltet, dafür sind aber die Ergebnisse genauer und zuverlässiger auch dann, wenn die Signal-To-Noise-Ratio meist nur 3:1 erreicht.

Während der Capture down-Assayvalidierung konnte gezeigt werden, dass eine weitere Komponente die Zuverlässigkeit der Auswertung beeinflusst. Trotz genauen Designs bleibt in manchen Fällen ein geringes Ungleichgewicht zwischen den zwei Diskriminatoren eines SNPs, die zusätzlich durch die unterschiedlichen Eigenschaften der zwei Fluoreszenzfarben verstärkt wird. Um die Unterschiede auszugleichen, müssten für jeden SNP einzelne Heterozygotenkontrollen definiert werden, was die Reservierung eines Drittels des Microarrays als ineffiziente Maßnahme erforderlich machen würde. Die intra-SNP Diskrepanz zwischen den zwei Diskriminatoren während des Capture down-Assays konnte mit Hilfe der für diese Plattform neu entwickelten statistischen Auswertung eliminiert werden. Dafür wurden anhand von 40 Referenzproben mit bekannten Genotypen und einer

statistischen Berechnung der Konfidenzintervalle die Genotypenbereiche für jeden SNP einzeln bestimmt. Die statistisch gewonnenen Ergebnisse dienen der Definition eines Genotypisierungsmakros, welches jeden SNP komfortabel nach individuellen Vorgaben auswertet. Damit kann diese Methode als Normalisierung angesehen werden und verleiht dem Capture down-Assay eine hohe Zuverlässigkeit. Während die System eigene automatische Auswertung, die alle SNPs pauschal nach dem gleichen Verhältnis errechnet, 16% der Referenzproben nicht bestimmt oder sogar falsch bewertet, sind es nur 4% Fehlerquote mittels selbst programmierter statistischer Auswertung (Tabelle 36 in Kapitel 3.3.3.3). Hier wird die Zuverlässigkeit des Capture down-Assays mit Hilfe der statistischen Auswertung um 12% gesteigert und erreicht eine hohe Genauigkeit von 94%.

Damit stellt das validierte Capture down-Assay eine sichere Genotypisierungsmethode dar, die durch die Minimierung der Capture-Oligos, der PCR-Produkte und der Diskriminatoren die Kontaminationsgefahr während der einzelnen Arbeitsschritte verhindert. Vor diesem Hintergrund erfüllt die Methode die Voraussetzung für eine Parallelbeladung von vier Cartridges. Das Loaderprotokoll erlaubt die Capture-Adressierung der 32 SNPs simultan auf vier Microarrays in nur einer Stunde und die Hybridisierung von zwölf verschiedenen PCR-Multiplexen sukzessiv in 1,5 Stunden. Da die gleiche Diskriminatoren- und Universal Reporter-Mischung benötigt wird, kann die passive Hybridisierung auch parallel in 10 Minuten erfolgen. Allein das Readerprotokoll kann nicht simultan ablaufen und aus diesem Grund werden für jedes Microarray 15 Minuten benötigt. Dennoch ist die Analyse von vier Cartridges schnell, in nur vier Stunden, abgeschlossen und die Genotypen für zwölf Personen und 32 SNPs können zeitnah abgelesen werden. Durch die Parallelbeladung von vier Cartridges wird die Schnelligkeit und Effizienz der elektronischen Microarray-Technologie erhöht, indem die Analysedauer pro Mikrochip von einer Stunde und 45 Minuten auf eine Stunde reduziert wird.

Die Stripping Eigenschaft der Methode bringt weitere Vorteile im Hinblick auf die Schnelligkeit und Effizienz, wobei die Diskriminatoren und die PCR-Produkte vollständig entfernt werden, während die Capture-Sonden mittels starker Biotin-Streptavidin-Bindung stabil bleiben. Ein benutzter Microarray kann mindestens einmal gestrippt und ohne Kontaminationsgefahr neu beladen werden. Mit jedem weiteren Stripping steigt die Gefahr der Kontamination oder der Beschädigung des Layers durch wiederholte elektronische Aktivierungen. Die Stripping-Methode ermöglicht, dass ein Cartridge zweimal benutzt werden kann und ist damit vor allem eine kostensparende Maßnahme, denn ein Microarray Cartridge von 150€ stellt in diesem Arbeitsprozess das teuerste Verbrauchsmaterial dar.

#### **4.3.3.2 Bewertung des Capture down-Assays**

Als Fazit kann zusammengefasst werden, dass die elektronische Nanogen Microarray-Plattform und das dafür validierte Capture down-Assay geeignet sind eine SNP-Genotypisierung in speziellen Fragen der Forensik durchzuführen:

Die erfolgreiche Validierung des 32plex Protokolls unterstreicht die Multiplexfähigkeit der Methode. Dabei sind 32 SNPs zur simultanen Typisierung nicht als oberstes Limit anzusehen, denn bei einer entsprechend größeren Auswahl an SNPs ( $\approx 130$ ) kann das

Assay theoretisch bis zu 98 SNPs verarbeiten. Dafür spricht die Bewährung der präzise festgelegten Oligodesignkriterien, die ganz besonders bei der eins zu eins Adaptation des validierten Assays an der fortschrittlichen, vollautomatischen NC400-Plattform zum Ausdruck kommt. Diese praktischen Erfahrungen gepaart mit der Tatsache, dass bei einem Capture down-Assay kein sukzessives Reporting stattfindet, welches beim Amplicon down-Assay ein Limit darstellte, weisen auf unproblematische Erweiterungsoptionen hin. Der einzige Schwachpunkt, der eine Limitation in der Multiplexgröße darstellt, betrifft die PCR-Amplifikation, denn für eine vollständige Genotypisierung müssen alle Produkte einer Multiplex stabil und in vergleichbaren Mengen amplifiziert sein.

Die Schnelligkeit der Methode konnte im Laufe der Optimierung mit großem Erfolg gesteigert werden. Dazu tragen Faktoren bei, wie Protokollvalidierung, die eine schnelle, vollautomatische Beladung, Detektion und Auswertung garantiert, oder die Parallelanalyse von vier Microarrays, aber auch die Reduzierung des Assayablaufs um den Aufreinigungsschritt. Bei all diesen Komponenten ist die manuelle Handhabung auf ein Minimum reduziert und damit die potenziellen Fehlerquellen eingeschränkt und die Zuverlässigkeit erhöht.

Abgesehen von der Abschaffung eines manuellen Schrittes machen weitere Erkenntnisse und Optimierungen die Methode zuverlässig. So wird die Kreuzkontaminationsgefahr durch Minimierung der Capture-Konzentration und des Einsatzvolumens des PCR-Produktes praktisch eliminiert. Weiterhin gewährleistet das validierte Readerprotokoll mit nur einem Scan eine sichere Detektion.

Die Genauigkeit der Methode ist gegeben durch die optimale Definition des Backgrounds und vor allem durch die neu entwickelte statistische Auswertung. Diese ersetzt die Heterozygotenkontrollen und dient für jeden SNP individuell als Normalisierung. Die Auswertung erfolgt mittels eines Makros, das keine manuelle oder subjektive Editierung der Ergebnisse erfordert und damit eine genaue Genotypisierung liefert.

Eine elektronische Microarray-Methode ist allein wegen der Anschaffung der Cartridges immer kostenaufwendig. Hier konnten aber die weiteren anfallenden Kosten durch mehrere Maßnahmen stark reduziert werden. Die Minimierung der benötigten Reagenzien und der Wegfall einer Aufreinigungsmethode tragen unmittelbar dazu bei. Die Einführung der Universal Reporter, wodurch die teure Fluoreszenzmarkierung der Diskriminatoren erspart geblieben ist, reduziert die Kosten maßgeblich. Darüber hinaus ist die Möglichkeit des Strippings und der Wiederverwendung eines benutzen Cartridges zu nennen, wodurch sich der Cartridge-Verbrauch auf die Hälfte reduzieren lässt.

Als Schlussfolgerung bleibt festzuhalten, dass die elektronische Microarray-Technologie eine multiplexfähige, flexible, effiziente und schnelle Methode darstellt, welche sich für die Hochdurchsatztypisierung von SNPs eignet. Jedoch konnte das dafür validierte Capture down-Assay mit einer Genauigkeit von bis zu 94% nicht die Zuverlässigkeit erreichen, die für den Einsatz in der forensischen Routine erforderlich ist.

#### 4.3.3.3 Vergleichende Beurteilung mit anderen Genotypisierungsmethoden

Seit einigen Jahren werden kontinuierlich neue Plattformen und Genotypisierungsmethoden für SNPs entwickelt (Syvanen 2001, Sobrino et al. 2005). Alle diese Methoden weisen, mit unterschiedlichen Prinzipien Vor- und Nachteile auf, sodass keine der Technologien als ultimativer Durchbruch angesehen werden kann. Vielmehr muss eine Methode, entsprechend der Fragestellung, individuell gewählt und angepasst werden.

Hohe Multiplexfähigkeit unterstützt durch Flexibilität und Schnelligkeit zeichnen die elektronische Nanogen Microarray-Technologie gegenüber den anderen Methoden aus.

Als offenes System erlaubt die Nanogen-Technologie die Entwicklung unterschiedlicher Assayformate oder auch die Konsolidierung von verschiedenen Methoden auf einer Plattform. Dabei ist die Fähigkeit des single site multiplexing, wie beim Capture down-Assay, theoretisch bis 98 SNPs unbegrenzt. Die Multiplexfähigkeit kann durch die Einführung der NC400-Plattform weiter in ein advanced panel multiplexing ausgebaut werden. Da hier der Reportingschritt in die sog. walk away automation integriert ist, können mehrere SNPs auf einem Pad für eine Probe typisiert werden, indem die Reportingschritte sukzessiv erfolgen. So können auf einem Microarray zum Beispiel 80 Proben für 25 SNPs, die auf fünf Pads verteilt sind, unter fünf Reportingschritten typisiert werden (d.h. 2000 SNPs auf einem Chip, z.B. RVA Influenza ASR<sup>13</sup>). Durch die multiple Anwendung im vollautomatischen Format stellt diese Technologie nicht nur eine ideale Plattform für Multiplex-Assays dar, sondern reduziert gleichzeitig den Kostenaufwand.

Die erhöhte Schnelligkeit gegenüber den anderen Methoden, die im Rahmen des SNPforID-Projektes etabliert wurden, ergibt sich schon durch die vollautomatische Bedienung des Loaders und des Readers, die eine sekundenschnelle Beladung, Hybridisierung und Diskriminierung vollziehen. Hinzu kommen zwei wirkungsvolle Faktoren, die komplette Eliminierung des Aufreinigungsschrittes und die schnelle makrounterstützte Auswertung. Die Möglichkeit der Ergebniskontrolle erhöht die Sicherheit der Genotypisierung und trägt zu besserer Genauigkeit bei. Eine Kontrolle kann bei Bedarf nahezu kostenfrei erfolgen, indem Teilschritte wiederholt werden, oder die Protokolle mit zusätzlichen Konditionen erweitert werden.

Der Vorteil der SNaPshot-Methode gegenüber dem elektronischen Microarray liegt darin, dass sie zur Genotypbestimmung nur ein Kapillarelektrophorese-Instrument benötigt. Eine solche Plattform ist in jedem Labor vorhanden, das für forensische Analysen ausgestattet ist. Damit ist für eine Genotypisierung mittels SNaPshot-Kits weder eine teure Geräteneuanschaffung noch eine zeitaufwendige Einarbeitung nötig. Dennoch zeigte die Validierung der SNaPshot-Methode für die 52 SNPs, dass das Protokoll auf die verschiedenen Instrumente angepasst werden muss (Sanchez et al. 2006a, Sanchez et al. 2006b, Sanchez et al. 2008). Die Validierung wurde mittels Doppelanalyse von 700 Proben in vier Laboratorien durchgeführt. Die SNaPshot-Typisierungen von jeweils 1-10ng DNA in 25µl PCR-Volumen ergaben 80% Übereinstimmung zwischen den zwei Analysen. Bei den restlichen 20% musste die Genotypisierung mehr als zweimal durchgeführt werden, bis zwei davon ein gleiches Ergebnis zeigten. Dabei waren dennoch Signale zu beklagen, die durch

---

<sup>13</sup> [www.nanogen.com](http://www.nanogen.com)

den Makro falsch erkannt wurden und manuell editiert werden mussten. Eine solche Editierung ist durch Einführung der statistischen Auswertung bei Nanogen nicht mehr nötig, was die Sicherheit der Microarray-Technologie gegenüber der SNaPshot-Methode erhöht. Die Typisierung von 52 SNPs beweist die Multiplexfähigkeit der SNaPshot-Methode, jedoch nicht ohne Einschränkung. Da es in der Kapillarelektrophorese zur Trennung nach Fragmentlängen kommt, sieht das Designkriterium vor, Analyseprimer mit vier Basen Längenunterschied zu verwenden. Dementsprechend würde die simultane Typisierung von 52 SNPs auch Analyseprimer über 130bp Länge verlangen. Um eine Primer-Dimer-Bildung und Qualitätsverluste zu vermeiden, wurde die 52plex-PCR in zwei SNaPshot-Reaktionen, in 23plex und 29plex, aufgeteilt und getrennt analysiert (Sanchez et al. 2006b, Tabelle 12 in Kapitel 2.2.5.1). Ähnlich haben auch andere Arbeitsgruppen große PCR-Multiplexe in kleinere Multiplexe für die SNaPshot-Analyse geteilt (Inagaki et al. 2004, 39plex-PCR in fünf SNaPshots), (Vallone et al. 2005, 70 SNPs in maximal 6plex PCR und SNaPshot). Das in Großbritannien etablierte Foren-SNP™ Multiplex-Kit (The Forensic Science Service®, UK) (Dixon et al. 2005, Dixon et al. 2006) stellt mit 21 SNPs noch eines der größten Multiplexsysteme dar. Damit sind der realen Multiplexfähigkeit der SNaPshot-Methode Grenzen gesetzt, worunter auch ihre Hochdurchsatzkapazität leidet. Denn mehrere kleine SNaPshot-Reaktionen und deren elektrophoretische Auftrennung bedeuten nicht nur hohe Kosten, sondern auch mehr Zeitaufwand. Vor allem bedeutet das für die SNaPshot-Methode, die ohnehin durch die aufwendige enzymatische Aufreinigung und Minisequenzierung verzögert ist, einen Nachteil gegenüber der Nanogen-Plattform.

Die SNP-Typisierung mittels Microarray am Glasobjektträger kostet in erster Linie ebenfalls wegen der Aufreinigung und der Minisequenzierung mehr Zeit. Hinzukommen noch aufwendige manuelle Hybridisierungs- und Waschschriffe. Der entscheidende Unterschied zwischen der Microarray-Typisierung am Glasobjektträger und einer elektronischen Microarray-Technologie ist die Art und Weise der Hybridisierung. Für die Genotypisierung an Glaträgern werden die fluoreszenzmarkierten DNAs passiv unter Einfluss von verschiedenen Puffern und Temperaturen an die immobilisierten Oligonukleotide hybridisiert. Damit ist die Hybridisierungsgeschwindigkeit und Effizienz vom langsamen Diffusionsprozess abhängig (Edman et al. 1997). Auch nach dem Prinzip des „array of arrays“ dauert die Genotypisierung von 1-12 Proben (unabhängig von der Anzahl der SNPs) über sechs Stunden. Dieser Zeitraum verlängert sich zusätzlich durch die schwierige Auswertung. Allein die Normalisierung der Daten erfolgt in drei Schritten, der Subtraktion des Backgrounds, der Korrektur der spektralen Überlappung und der Allelbalancierung. Erst danach kann die Datenauswertung mittels mehrerer Makros zur Umwandlung der Rohdaten und diverser Software (ImaGene™ 5.5/4.1, BioDiscovery und SNPSnapper National Public Health Institute, Helsinki) erfolgen (Bogus et al. 2006). Bei der Etablierung der Methode für einen vier Laser-Scanner (GeneTACTM LMIV, Genomic Solutions) und die 52 autosomalen SNPs zeigten sich viele unreproduzierbare Fehler. Aus diesem Grund konnte das Verfahren nur für eine Multiplexanalyse von 35 SNPs validiert werden. Die geringe Konkordanz und hohe Fehldetektion liefern Beweise dafür, dass die Microarray-Typisierung am Glasobjektträger keine Alternative zu der elektronischen Microarray-Technologie darstellt. Ein geringer Vorteil der Methode gegenüber einem elektronischen Microarray ist der Kostenfaktor. Ein gespottetes Glas-Microarray ist wesentlich kostengünstiger als ein elektronisches

Microarray, jedoch verlangt die Methode teure Polymerasen, Enzyme und fluoreszenzmarkierte ddNTPs sowie die Anschaffung eines Laser Scanners.

Die Genotypisierung mittels SBE-Minisequenzierung und MALDI-TOF MS als Detektionsmethode zeigt im Hinblick auf die Multiplexfähigkeit einige Einschränkungen bei der genauen Detektion einer A/T Substitution, sowie der Aufreinigung und der begrenzten Detektionsgrenze. Diese Einschränkungen konnten jedoch im Laufe der Etablierung gelöst werden (Mengel-Jorgensen et al. 2005). Bei der SBE-Minisequenzierung wurde das Molekulargewicht von ddUTP mit einem Aminolinker erhöht, sodass A und T getrennt detektierbar waren. Die Aufreinigung wurde verbessert, indem biotinylierte ddNTPs verwendet wurden, um eine sichere Trennung der DNA-Stränge zu erreichen (genannt MATP, monomeric avidin triethylamine purification). Die Erweiterung der Messgrenze in den Bereich des kleineren Molekulargewichts hinein wurde erreicht, indem nach der Minisequenzierung die Primer mittels RNase verkürzt wurden (genannt MARNA). Nach der Etablierung des Einsatzes der veränderten ddNTPs, MATP-Aufreinigung und MARNA-Behandlung eignet sich das MALDI-TOF MS für die Multiplexanalyse von DNA-Fragmenten von 3000 bis 105000Da (Mengel-Jorgensen et al. 2005). Diese Optimierungsmaßnahmen machen die Methode multiplexfähig, gleichzeitig aber auch arbeits- und zeitintensiv. Zusätzlich gestaltet sich die Auswertung von großen Multiplexen und MARNA-verkürzten Proben sehr schwierig. Obwohl die eigentliche Detektion mittels MALDI-TOF MS sehr schnell abläuft, ist diese Methode gegenüber dem elektronischen Microarray zeitaufwendig.

Bei der Validierung der Pyrosequencing-Methode konnte eine 22plex-PCR typisiert werden, indem diese für eine Triplex-Minisequenzierung achtmal verwendet wurde. Dabei zeigten die wachsenden Backgroundsignale, dass die Kapazitätsgrenze der Methode erreicht wurde. Das heißt, die Multiplexfähigkeit der Pyrosequencing-Technologie ist auf maximal 24 SNPs begrenzt (Bender et al. 2006). Positiv bewertet wurde allein die Sensitivität der Methode, denn bereits 30pg DNA waren ausreichend für eine vollständige Typisierung (Harrison et al. 2006). Dennoch kann dieses Ergebnis nicht mit anderen Versuchen verglichen werden, weil hier nur mit einer Pentaplex-PCR experimentiert werden konnte.

Grundlegend wurde das SNPLex-Assay so verändert, dass eine etablierte Multiplex-PCR-Amplifikation der OLA-Reaktion vorangeschaltet wird. Damit ist die Genplex-Methode sensitiv genug, um bereits 1ng DNA genotypisieren zu können (Phillips et al. 2007, Musgrave-Brown et al. 2008). Die Multiplexfähigkeit der Methode ist auf 48 SNPs begrenzt, das reicht aber für die Identifikation eines Unbekannten mit über 99,99%-iger Sicherheit. Maßgeblicher Vorteil der Genplex-Methode gegenüber dem elektronischen Microarray ist, dass sie, wie die SNaPshot-Methode, keine teure Neuanschaffung, sondern nur eine weit verbreitete Kapillarelektrophorese benötigt. Damit stellt das Genplex-Ligationsassay eine multiplexfähige, sichere, kosten- und zeitgünstige Hochdurchsatzmethode dar. Die einzigen nachteiligen Komponenten der Genplex-Methode sind die fehlende Flexibilität und die Abhängigkeit bei dem Sonden- und Assaydesign von der Firma Applied Biosystems. Das sind genau die Bereiche, die die elektronische Microarray-Technologie im Gegensatz dazu besonders auszeichnen.

Die Vorteile des elektronischen Microarrays, die sich bei der vergleichenden Beurteilung in dieser Arbeit aufzeigen, stehen im Einklang mit den Schlussfolgerungen von Gilles et al. 1999. Sie fassen die Vorteile des elektronischen Microarrays gegenüber konventionellen

SNP-Genotypisierungsmethoden, wie RFLP (restriction fragment length polymorphism), Dot-blot Hybridisierung mit ASO (allele-specific oligonucleotide probes) oder SSCP (single-strand conformational polymorphism) in sieben Punkten zusammen:

1. Flexibilität, indem ein Assay beliebig und unmittelbar vor der Analyse konfiguriert werden kann;
2. Geschwindigkeit, durch schnellen elektronischen Transport und Konzentration von Molekülen;
3. Multiplexfähigkeit, durch die individuell kontrollierbaren Pads;
4. Effizienz, indem die einzelnen Schritte in Echtzeit verifiziert und validiert werden können;
5. Laboratory-on-a chip Technologie, wobei zeit- und arbeitsaufwendige Schritte entfallen;
6. Automation, d.h. vollautomatischer Assayablauf;
7. Eignung für Gen-Expressionsanalyse.

Zusammenfassend kann festgehalten werden, dass es dem SNPforID-Konsortium gelungen ist ausführliche praktische Informationen über den forensischen Einsatz der einzelnen SNP-Genotypisierungsmethoden zu liefern. Die vergleichende Gegenüberstellung und Beurteilung der Methoden bestätigen, dass die Vielfalt der Technologien mit ihren unterschiedlich verteilten Vor- und Nachteilen die Bestimmung einer in jeder Hinsicht optimalen Methode zur Bearbeitung forensischer Fragestellungen unmöglich macht. Zum jetzigen technischen Entwicklungsstand muss eine Entscheidung über die Wahl einer Genotypisierungsmethode vielmehr nach dem Schwerpunkt der Fragestellung getroffen werden. Kriterien wie Anzahl der Marker bzw. Assaygröße, Zeit- und Kostenfaktor und zulässige Abweichungen sind dabei zu gewichten.

#### **4.3.4 Ausblick**

In den letzten fünf Jahren rückten SNP-Marker immer stärker in den Fokus der forensischen Molekulargenetik. Das Interesse galt der Einsetzbarkeit dieser Marker für die forensische Forschung und den dazu nötigen technologischen Voraussetzungen (Sobrinho et al. 2005). Trotz dieser Bemühungen kann die SNP-Genotypisierung zum jetzigen Zeitpunkt die in nationalen DNA-Datenbanken fest verankerten STR-Systeme nicht ersetzen (Gill et al. 2004). Aber ein weiterer Zuwachs und der ergänzende Einsatz der SNP-Technologie ist gerade bei speziellen Fragestellungen zu erwarten, wie z.B. weit zurückführender Verwandtschaftsfragen, oder bei schwierigem Ausgangsmaterial (Lessig et al. 2006, Phillips et al. 2008).

Darüber hinaus lieferte die intensive Erforschung der SNP-Marker disziplinübergreifende Erkenntnisse, welche einen neuen Forschungszweig der Forensik eröffneten. Neueste Veröffentlichungen zeigen, dass die Analyse komplexer Merkmale und ihre Anwendung in der forensischen Medizin im Rahmen von Assoziations- und Kopplungsstudien immer mehr in das Interesse der Forscher rückt (Carracedo 2006, Aulchenko et al. 2007). Zusätzlich beschäftigen Gene, die ein phänotypisches Merkmal (physical traits) kodieren (McEvoy et al. 2006), wie z.B. das Melanocortin 1-Rezeptorgen (MC1R), das die rote Pigmentierung des Haares verursacht (Grimes et al. 2001), immer häufiger die Forensik, unter dem Aspekt der Personenkreisbegrenzung bei der Identifizierung eines Unbekannten. Die Vorstellung von mehreren Arbeiten auf dem 22<sup>nd</sup> International Society for Forensic Genetics (ISFG) Kongress 2007 in

Kopenhagen, die SNP basierende, physische Merkmale untersuchen (Branicki et al. 2008, Daniel et al. 2008, Kayser et al. 2008, van Daal und Salini Manoharan 2008), entfachte eine Diskussion in der forensischen Fachgesellschaft über die Vorteile und Gefahren der Analyse solcher Merkmale (Schneider 2008).

Während die Selektion der SNPs durch ausführliche Primärdatenbanken, mit Unterstützung der kontinuierlich entwickelten Referenzdatenbanken (Willuweit und Roewer 2007, Amigo et al. 2008) präzise erfolgen kann, gestaltet sich die Wahl der Genotypisierungsmethode schwierig (Sobrino et al. 2005). Die Assays und Technologien, die im Rahmen des SNPforID-Projektes validiert wurden, können nicht alle Erwartungen an eine forensisch relevante Methode, wie Multiplexfähigkeit, Effizienz und Schnelligkeit bei gleichzeitig hoher Zuverlässigkeit erfüllen. Die elektronische Microarray-Technologie zeigt zunächst gewisse Vorteile gegenüber konventionellen Methoden. Die Vorteile des für die Nanogen-Technologie validierten Capture down-Assays sind in erster Linie die Automatisierung, welche die räumliche Ausrichtung der Moleküle unmittelbar vor der Analyse ermöglicht und dadurch Fehlerquellen minimiert und gleichzeitig die Zeiteffizienz erhöht; zweitens die erfolgreiche Miniaturisierung, indem die Methode wenig Reagenzien und Reaktionen benötigt und dadurch zusätzlich den Zeit- und Arbeitseinsatz und Entsorgungsaufwand reduziert; drittens die Parallelität, die im Multiplexansatz auf einem einzigen Microarray Antworten auf eine Vielzahl von Fragen gibt und die gleichzeitige Sammlung großer Datenmengen ermöglicht.

Obwohl diese Eigenschaften typisch für eine Technologie mit Hochdurchsatzkapazität sind, gibt es noch einige verbesserungswürdige Punkte. Daher sollen im Folgenden Strategien diskutiert werden, um die Anwendung zukunftsfest zu machen. Eine Limitation fällt auf, wenn man bedenkt, dass das validierte Assay nur 32 SNPs simultan analysiert und damit die gewünschte Anzahl des SNPforID-Projekts nicht erreicht wird. Grund dafür war, dass die vorausgewählten 52 SNPs aufgrund flankierender Sequenzen nur in 35 Fällen ein präzises Oligodesign erlaubten. Da das System durchaus die Bearbeitung von mehr SNPs zulässt, kann das oben beschriebene Problem durch eine größere Vorauswahl bestimmter SNPs gelöst werden.

Das validierte Capture down-Assay konnte nicht die gewünschte Konkordanz zu der SNaPshot-Referenzmethode erreichen, damit bleibt die Zuverlässigkeit für die Forensik unerfüllt. Das liegt daran, dass trotz des genauen Oligodesigns die Hybridisierung bzw. Diskriminierung nicht für alle SNPs optimal funktioniert hat. Möglicherweise ist eine thermische Denaturierung der PCR-Produkte vor der Hybridisierung, welche bei Raumtemperatur erfolgt, nicht ausreichend. Auch eine asymmetrische Amplifikation führte nicht zur gewünschten Verbesserung. Gleichermaßen gestaltete sich die Strangtrennung mit Hilfe von Biotin als schwierig, denn nach der Trennung wird der nicht biotinylierte Strang, der sich in Lösung befindet, für die Weiterverwendung gebraucht, wobei schon die kleinste Spur Biotin auf dem Microarray zu Störungen führen kann. Die Hybridisierung könnte erheblich verbessert werden, wenn die PCR-Produkte einzelsträngig vorlägen. Die Gewinnung einzelsträngiger PCR-Produkte könnte mittels Phosphorothioaten an einem PCR-Primer möglich sein. Nikiforov et al. produzierten einzelsträngige PCR-Produkte durch vier Phosphorothioate am 5' Ende eines PCR-Primers (Nikiforov et al. 1994). Nach der PCR findet ein Nukleaseverdau statt, wobei nur der unveränderte Strang hydrolisiert, während die Phosphorothioate auf dem anderen Strang den Verdau verhindern. Ziel dieser Trennung ist



eine Verbesserung der Hybridisierung von PCR-Produkten mit den immobilisierten Capture-Oligos unter der Bedingung, dass sich keine nicht benötigten Stränge als Konkurrenz in der Flüssigkeit befinden.

Weiterhin könnte die Hybridisierung und Diskriminierung der Sonden unter Verwendung sog. Minor Groove Binding (MGB) Moleküle verbessert werden, deren stabilisierende Wirkung auf Hybridisierungsreaktionen oft berichtet wurde (Kutyavin et al. 2000, Afonina et al. 2002, Lukhtanov et al. 2007). Diese natürlich vorkommenden Antibiotika (z.B. Distamycin, oder CC-1065) binden an Duplex DNA, bevorzugt an A/T reichen Sequenzen (minor groove) über Hydrogen-Bindung (Distamycin) oder hydrophobische Interaktion (CC-1065). Zusätzlich erhöhen die MGBs die Schmelztemperatur, wodurch kurze Sonden mit verbesserter mismatch Diskriminierung verwendet werden können. Dieser Effekt könnte eine zuverlässigere Diskriminierung liefern und dadurch die SNP-Genotypisierung unterstützen.

Wird der Kostenfaktor einer Genotypisierung mittels elektronischen Microarrays betrachtet, so stellt dieser zum heutigen Zeitpunkt noch einen erheblichen Nachteil dar. Obwohl die Kosten durch Miniaturisierung reduziert werden konnten, bleibt die Technologie kostspielig, denn schon allein die Neuanschaffung einer Nanogen-Plattform und insbesondere die benötigten Cartridges sind mit großen Ausgaben verbunden. Mit der Einführung des NC400 Instruments der nächsten Generation könnten die Kosten assayabhängig reduziert und dadurch die Effizienz deutlich verbessert werden. Unterstützt von der Neuerung, dass alle Schritte in einem Gerät ohne manuelle Zwischenschritte elektronisch ablaufen, könnten das Amplicon down- und das Capture down-Assay so verändert werden, dass ein Cartridge jeweils höchst ökonomisch ausgenutzt wird. Angenommen, dass für ein Amplicon down-Assay einer Multiplex-PCR aus 50 SNPs nur fünf Pads adressiert würden, bedeutete dies auf einem Cartridge mit 400 Pads könnten 80 verschiedene Personen à 50 SNPs in nur zehn sukzessiven Reportingschritten genotypisiert werden. Entsprechend könnten auf fünf Pads jeweils zehn Capture-Sonden adressiert werden und so ein 50plex-PCR-Produkt im Rahmen eines Capture down-Assays zehnmal sukzessive elektronisch hybridisiert werden. Somit könnten ebenfalls 80 Personen à 50 SNPs in nur zehn sukzessiven Reportingschritten genotypisiert werden. Damit ist die Kapazität der NC400-Plattform für beide Assays erhöht und das Amplicon down-Assay sowie das Capture down-Assay eignen sich zur parallelen Genotypisierung von verschiedenen Individuen für mehrere SNPs, was wiederum die Erwartungen zur Bearbeitung von forensisch relevanten Fragestellungen voll und ganz erfüllt.

## 5 Zusammenfassung

Der zunehmende Bedarf an molekulargenetischen Analysen in der forensischen Routine verlangt immer sensitivere und effizientere Systeme zur Personenidentifizierung. Unter diesem Aspekt bieten Einzelnukleotid-Polymorphismen (SNPs) einige strukturbedingte Vorteile gegenüber den konventionellen Systemen, wie RFLPs oder STRs. So eignen sich die SNPs wegen ihrer niedrigen Mutationsrate am besten zur Lösung von Abstammungsfragen und zur Erforschung von Verwandtschaftsverhältnissen. Weiterhin werden bei einer SNP-Analyse wesentlich kürzere Sequenzen untersucht, was bei degradiertem oder gering konzentriertem Ausgangsmaterial bessere Ergebnisse erwarten lässt. Diese strukturbedingte Eigenschaft drängt zudem die Einführung der Vollautomationsanalyse mit höherer Durchsatzkapazität auf.

Das multizentrische SNPforID-Projekt behandelt vor dem oben beschriebenen Gesamthintergrund Anwendungsmöglichkeiten von Einzelnukleotid-Polymorphismen (SNPs) zur Individualisierung von Personen im Kontext der Zuordnung von biologischen Tatortspuren oder auch bei der Identifizierung unbekannter Toter. Die vorliegende Dissertation entstand im Rahmen des SNPforID-Projektes und hatte die eigenständige Zielsetzung eine hochauflösende Genotypisierungsmethode zu etablieren und zu validieren, die mit hoher Genauigkeit aber geringen Aufwand SNPs im Multiplexformat simultan analysieren kann.

Eine SNP-Selektion und ihre Validierung mittels einer Referenzmethode stellen die notwendigen Grundlagen zur Lösung der zentralen Zielsetzung dar.

Zunächst wurden 29 Y-chromosomale und 52 autosomale SNPs unter der Anforderung ausgewählt, dass sie als Multiplex eine möglichst hohe Individualisierungschance aufweisen. Anschließend folgten die Validierungen beider Multiplex-Systeme und der SNaPshot™-Minisequenzierungsmethode in systematischen Studien unter Beteiligung aller Arbeitsgruppen des Projektes.

Die 29 SNPs des sog. Major Y-chromosome haplogroup typing Kits charakterisieren 31 weltweit verbreitete Haplogruppen. Die systematisch analysierten 1126 Proben aus zwölf Populationen verteilten sich auf 26 dieser Haplogruppen. Die statistische Auswertung der Stichproben der gesamten Y-chromosomalen Validierungsstudie anhand der Analysis of Molecular Variance (AMOVA), welche die Distanz zwischen zwei Haplogruppen ausdrückt, ergab mit 33,7% ( $p < 0,001$ ) eine hohe Varianz zwischen Haplogruppen verschiedener Kontinente und mit nur 6,8% ( $p < 0,001$ ) eine geringe Varianz zwischen Populationen eines Kontinents. Mit 59,5% ( $p < 0,001$ ) wurde jedoch die höchste Varianz innerhalb einer Population errechnet. Unter Berücksichtigung dieser Erkenntnisse kann das Major Y-chromosome haplogroup typing Kit zur Haplogruppenzuordnung herangezogen werden, aber auf keinen Fall können daraus Rückschlüsse auf die geographische Abstammung gezogen werden.

Die Validierung der 52 autosomalen SNPs erfolgte indem forensisch relevante Proben von geringer Quantität und Qualität analysiert und die Ergebnisse mit konventionellen STR-Systemen verglichen wurden. Für die SNP-Genotypisierung der qualitativ schlechten Proben, wie künstlich und natürlich degradiertes DNA, Extrakten aus zersetztem

Muskelgewebe und Haarschaften konnten SNP basiert wesentlich höhere Erfolge erzielt werden. Gleichzeitig weisen die Ergebnisse auf die Anfälligkeit der Methode in Bezug auf unterschiedliche Laborbedingungen und Reagenzienbereitstellung hin, wodurch sich der Einsatz des Assays in einem Routinelabor schwierig gestaltet. Dennoch kann die SNP-Genotypisierung gerade in Fällen, in denen STR-Systeme versagen, eine ergänzende Alternative sein.

Der eigenständige Hauptteil dieser Dissertation beschreibt unter Verwendung der zuvor validierten autosomalen SNPs die Neuentwicklung und Validierung eines Hybridisierungsassays für die elektronische Microarray-Plattform der Firma Nanogen. Dazu wurden im Vorfeld drei verschiedene Assays etabliert, die sich im Funktionsprinzip auf dem Microarray unterscheiden. Davon wurde leistungsorientiert das Capture down-Assay, das die Erwartungen einer forensischen Analyse am besten erfüllte, zur Weiterentwicklung ausgewählt. Der etablierte Assayablauf mit der Immobilisierung der Capture-Sonden auf dem Microarray, selektiver Hybridisierung der PCR-Produkte einer Multiplex und Hybridisierung der Reporter-Oligos erlaubt eine besonders flexible, multiplexfähige und schnelle SNP-Genotypisierung. Nach zahlreichen Optimierungsmaßnahmen hinsichtlich PCR-Produktbehandlung, gerätespezifischer Abläufe und analysespezifischer Oligonukleotiddesigns stand das Capture down-Assay zur simultanen Typisierung von drei Individuen mit je 32 SNPs auf einem Microarray bereit. Anschließend wurde dieses Verfahren anhand von 40 DNA-Proben mit bekannten Genotypen für die 32 SNPs validiert und durch parallele SNaPshot™-Typisierung die Genauigkeit bestimmt.

Das Ergebnis beweist nicht nur die Eignung des validierten Analyseassays und der elektronischen Microarray-Technologie für bestimmte Fragestellungen, sondern zeigt auch deren Vorteile in Bezug auf Schnelligkeit, Flexibilität und Effizienz. Die Automatisierung, welche die räumliche Anordnung der zu untersuchenden Fragmente unmittelbar vor der Analyse ermöglicht, reduziert unnötige Arbeitsschritte und damit die Fehlerhäufigkeit und Kontaminationsgefahr bei verbesserter Zeiteffizienz.

Die Reduzierung der Anzahl der SNPs aus möglichen 52 auf 32 garantierte bei der Validierung mittels 40 Referenzproben eine 94%ige hohe Konkordanz. Einen wesentlichen Beitrag zu dieser Erfolgsrate leistete die selbst entwickelte Datenauswertung. Diese makrobasierte statistische Auswertung behandelt jeden SNP einer Multiplexanalyse einzeln, was die Konkordanz um 12% erhöht im Gegensatz zu der Auswertung mittels der automatischen Onboard Software. Dennoch kann mit einer maximal erreichten Genauigkeit von 94% die Zuverlässigkeit der in der forensischen Genetik aktuell eingesetzten STR-Systeme noch nicht erreicht werden.

Die Rolle des neuen Verfahrens wird damit nicht in einer Ablösung der etablierten Methoden, sondern in einer Ergänzung zur Lösung spezieller Probleme wie z.B. der Untersuchung stark degradierter DNA-Spuren zu finden sein.

## 6 Literaturverzeichnis

- Adams SM, King TE, Bosch E, Jobling MA (2006) The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion. *Forensic Sci Int* 159(1): 14-20.
- Afonina IA, Reed MW, Lusby E, Shishkina IG, Belousov YS (2002) Minor groove binder-conjugated DNA probes for quantitative DNA detection by hybridization-triggered fluorescence. *Biotechniques* 32(4): 940-4, 946-9.
- Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, Lundeberg J (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* 280(1): 103-10.
- Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363(1-2): 83-94.
- Amigo J, Phillips C, Lareu M, Carracedo A (2008) The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. *Int J Legal Med* 122(5): 435-40.
- Amorim A, Pereira L (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int* 150(1): 17-21.
- Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C (2004) A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 75(2): 338-45.
- Arriola E, Lambros MB, Jones C, Dexter T, Mackay A, Tan DS, Tamber N, Fenwick K, Ashworth A, Dowsett M, Reis-Filho JS (2007) Evaluation of Phi29-based whole-genome amplification for microarray-based comparative genomic hybridisation. *Lab Invest* 87(1): 75-83.
- Aulchenko YS, de Koning DJ, Haley C (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 177(1): 577-85.
- Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE (2001) Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* 11(6): 1005-17.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. *Science* 297(5583): 1003-7.
- Balogh MK, Burger J, Bender K, Schneider PM, Alt KW (2003) STR genotyping and mtDNA sequencing of latent fingerprint on paper. *Forensic Sci Int* 137(2-3): 188-95.
- Balogh MK, Børsting C, Sánchez Diz P, Thacker C, Syndercombe-Court D, Carracedo A, Morling N, Schneider PM (2006) Application of whole genome amplification for forensic analysis. *Progr Forensic Genet* 11: 724-727.
- Barber AL, Foran DR (2006) The utility of whole genome amplification for typing compromised forensic samples. *J Forensic Sci* 51(6): 1344-9.
- Bender K, Nehlich C, Harrison C, Musgrave-Brown E, Syndercombe-Court D, Schneider PM (2006) A multiplex SNP typing approach for the DNA pyrosequencing technology. *Progr Forensic Genet* 11: 73-75.
- Bender KF, Maria José; Schneider, Peter M (2004) Preparation of degraded human DNA under controlled conditions. *Progr Forensic Genet* 10.
- Bergen AW, Haque KA, Qi Y, Beerman MB, Garcia-Closas M, Rothman N, Chanock SJ (2005) Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources. *Hum Mutat* 26(3): 262-70.
- Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Coenelly PM, Duceman B, Eisenberg A, Forman L, Kidd KK, Leclair B, Niezgodka S, Parsons TJ, Pugh E, Shaler R, Sherry S, Sozer A, Walsh A (2005) DNA identifications after the 9/11 World Trade Center attack. *Science* 310: 1122-1123.
- Bogus M, Sobrino B, Bender K, Carracedo A, Schneider PM (2006) Rapid microarray-based typing of forensic SNPs. *Progr Forensic Genet* 11: 37-39.
- Børsting C, Sanchez JJ, Morling N (2004) Multiplex PCR, amplicon size and hybridization efficiency on the NanoChip electronic microarray. *Int J Legal Med* 118(2): 75-82.

- Bosch E, Calafell F, Rosser ZH, Norby S, Lynnerup N, Hurles ME, Jobling MA (2003) High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum Genet* 112(4): 353-63.
- Bouakaze C, Keyser C, Amory S, Crubezy E, Ludes B (2007) First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int J Legal Med*.
- Branicki W, Szczerbinska A, Brudnik U, Wolanska-Nowak P, Kupiec T (2008) The OCA2 gene as a marker for eye colour prediction. *Progr Forensic Genet* 12: (in press).
- Bredel M, Bredel C, Juric D, Kim Y, Vogel H, Harsh GR, Recht LD, Pollack JR, Sikic BI (2005) Amplification of whole tumor genomes and gene-by-gene mapping of genomic aberrations from limited sources of fresh-frozen and paraffin-embedded DNA. *J Mol Diagn* 7(2): 171-82.
- Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, Paabo S (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A*.
- Brion M, Quintans B, Zarrabeitia M, Gonzalez-Neira A, Salas A, Lareu V, Tyler-Smith C, Carracedo A (2004) Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. *Gene* 329: 17-25.
- Brion M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2005) Introduction of a single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. *Electrophoresis* 26(23): 4411-20.
- Brown TA (2007) *Genome und Gene*, Spektrum Akademischer Verlag.
- Butler JM, Devaney JM, Marino MA, Vallone PM (2001) Quality control of PCR primers used in multiplex STR amplification reactions. *Forensic Sci Int* 119(1): 87-96.
- Butler JM (2005) *Forensic DNA Typing. Biology, Technology, and Genetics of STR Markers*. 2nd edition.
- Cahill DJ, Nordhoff E (2003) Protein arrays and their role in proteomics. *Adv Biochem Eng Biotechnol* 83: 177-87.
- Carracedo A (2006) Finding genes that underlie cancer using genetic tools. *Clin Transl Oncol* 8(11): 771-2.
- Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B (1999) The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* 20(8): 1682-96.
- Christensen K, Murray JC (2007) What genome-wide association studies can do for medicine. *N Engl J Med* 356(11): 1094-7.
- Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114(2): 127-48.
- Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R (2004) Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 74(5): 1014-22.
- d'Abbadie M, Hofreiter M, Vaisman A, Loakes D, Gasparutto D, Cadet J, Woodgate R, Paabo S, Holliger P (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nat Biotechnol* 25(8): 939-43.
- Daniel R, Sanchez JJ, Nassif NT, Walsh SJ (2008) SNPs associated with physical traits: a valuable tool for the inference of biogeographical ancestry. *Forensic Sci Int: Gen Supplement* 1(1): 538-540.
- Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* 99(8): 5261-6.
- Dhiman N, Bonilla R, O'Kane DJ, Poland GA (2002) Gene expression microarrays: a 21st century tool for directed vaccine design. *Vaccine* 20(1-2): 22-30.
- Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Sci Int* 154(1): 62-77.

- Dixon LA, Dobbins AE, Pulker HK, Butler JM, Vallone PM, et al. (2006) Analysis of artificially degraded DNA using STRs and SNPs--results of a collaborative European (EDNAP) exercise. *Forensic Sci Int* 164(1): 33-44.
- Edman CF, Raymond DE, Wu DJ, Tu E, Sosnowski RG, Butler WF, Nerenberg M, Heller MJ (1997) Electric field directed nucleic acid hybridization on microchips. *Nucleic Acids Res* 25(24): 4907-14.
- Elbein SC, Chu W, Ren Q, Hemphill C, Schay J, Cox NJ, Hanis CL, Hasstedt SJ (2002) Role of calpain-10 gene variants in familial type 2 diabetes in Caucasians. *J Clin Endocrinol Metab* 87(2): 650-4.
- Estivill X, Cheung J, Pujana MA, Nakabayashi K, Scherer SW, Tsui LC (2002) Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. *Hum Mol Genet* 11(17): 1987-95.
- Ewalt KL, Haigis RW, Rooney R, Ackley D, Krihak M (2001) Detection of biological toxins on an active electronic microchip. *Anal Biochem* 289(2): 162-72.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131(2): 479-91.
- Feng LN, M. (1999) Electronic microarray for DNA analysis. *Gene Ther Mol Biol* 4: 183-191.
- Ferrari M, Stenirri S, Bonini P, Cremonesi L (2003) Molecular diagnostics by microelectronic microchips. *Clin Chem Lab Med* 41(4): 462-7.
- Findlay I, Taylor A, Quirke P, Frazier R, Urquhart A (1997) DNA fingerprinting from single cells. *Nature* 389(6651): 555-6.
- Fondevila M, Phillips C, Naveran N, Cerezo M, Rodriguez A, Calvo R, Fernandez L, Carracedo A, Lareu M (2008) Challenging DNA: assessment of a range of genotyping approaches for highly degraded forensic samples. *Forensic Sci Int: Gen Supplement* 1(1): 26-28.
- Geserick G, Wirth I, Strauch H (2001) 100 Jahre Forensische Serologie. *Rechtsmedizin* 11(5): 198-204.
- Gill P (2001a) Application of low copy number DNA profiling. *Croat Med J* 42(3): 229-32.
- Gill P (2001b) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 114(4-5): 204-10.
- Gill P, Werrett DJ, Budowle B, Guerrieri R (2004) An assessment of whether SNPs will replace STRs in national DNA databases--joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM). *Sci Justice* 44(1): 51-3.
- Gilles PN, Wu DJ, Foster CB, Dillon PJ, Chanock SJ (1999) Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nat Biotechnol* 17(4): 365-70.
- Glinsky GV (2006) Integration of HapMap-based SNP pattern analysis and gene expression profiling reveals common SNP profiles for cancer therapy outcome predictor genes. *Cell Cycle* 5(22): 2613-25.
- Goering RC, KLF (2003) Development of a Universal Reporting System for Detecting Single Nucleotide Polymorphisms in the Chromosome of *Staphylococcus Aureus* Utilizing an Electronic Microarray Format.
- Gonzalez M, Bagatolli LA, Echabe I, Arrondo JL, Argarana CE, Cantor CR, Fidelio GD (1997) Interaction of biotin with streptavidin. Thermostability and conformational changes upon binding. *J Biol Chem* 272(17): 11288-94.
- Grimes EA, Noake PJ, Dixon L, Urquhart A (2001) Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci Int* 122(2-3): 124-9.
- Gusella M, Padriani R (2007) G>C SNP of thymidylate synthase with respect to colorectal cancer. *Pharmacogenomics* 8(8): 985-96.
- Gusmao L, Brion M, Gonzalez-Neira A, Lareu M, Carracedo A (1999) Y chromosome specific polymorphisms in forensic analysis. *Leg Med (Tokyo)* 1(2): 55-60.
- Gut IG (2001) Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat* 17(6): 475-92.

- Haff LA, Smirnov IP (1997) Multiplex genotyping of PCR products with MassTag-labeled primers. *Nucleic Acids Res* 25(18): 3749-50.
- Harrison C, Musgrave-Brown E, Bender K, Carracedo A, Morling N, Schneider PM, Syndercombe-Court D (2006) A sensitive issue: Pyrosequencing as a valuable forensic SNP typing platform. *Progr Forensic Genet* 11: 52-54.
- Hellmann A, Rohleder U, Schmitter H, Witting M (2001) STR typing of human telogen hairs - a new approach. *Int J Legal Med* 114(4-5): 269-273.
- Hirschhorn JN, Sklar P, Lindblad-Toh K, Lim YM, Ruiz-Gutierrez M, Bolk S, Langhorst B, Schaffner S, Winchester E, Lander ES (2000) SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. *Proc Natl Acad Sci U S A* 97(22): 12164-9.
- Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, Lasken RS (2003) Unbiased whole-genome amplification directly from clinical samples. *Genome Res* 13(5): 954-64.
- Huang Y, Shirajian J, Schroder A, Yao Z, Summers T, Hodko D, Sosnowski R (2004) Multiple sample amplification and genotyping integrated on a single electronic microarray. *Electrophoresis* 25(18-19): 3106-16.
- Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* 25(3): 333-7.
- Inagaki S, Yamamoto Y, Doi Y, Takata T, Ishikawa T, Imabayashi K, Yoshitome K, Miyaishi S, Ishizu H (2004) A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Sci Int* 144(1): 45-57.
- Jaakson K, Zernant J, Kulm M, Hutchinson A, Tonisson N, Glavac D, Ravnik-Glavac M, Hawlina M, Meltzer MR, Caruso RC, Testa F, Maugeri A, Hoyng CB, Gouras P, Simonelli F, Lewis RA, Lupski JR, Cremers FP, Allikmets R (2003) Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. *Hum Mutat* 22(5): 395-403.
- Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA. *Nature* 316(6023): 76-9.
- Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110(3): 118-24.
- Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci Int* 118(2-3): 158-62.
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4(8): 598-612.
- Kayser M, Liu F, Janssens AC, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van Ijcken WF, den Dunnen JT, Heath S, Zelenika D, Despriet DD, Klaver CC, Vingerling JR, de Jong PT, Hofman A, Aulchenko YS, Uitterlinden AG, Oostra BA, van Duijn CM (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet* 82(2): 411-23.
- Kerlavage A, Bonazzi V, di Tommaso M, Lawrence C, Li P, Mayberry F, Mural R, Nodell M, Yandell M, Zhang J, Thomas P (2002) The Celera Discovery System. *Nucleic Acids Res* 30(1): 129-36.
- King TE, Bosch E, Adams SM, Parkin EJ, Rosser ZH, Jobling MA (2005) Inadvertent diagnosis of male infertility through genealogical DNA testing. *J Med Genet* 42(4): 366-8.
- Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* 29(3): 279-86.
- Kutyavin IV, Afonina IA, Mills A, Gorn VV, Lukhtanov EA, Belousov ES, Singer MJ, Walburger DK, Lokhov SG, Gall AA, Dempcy R, Reed MW, Meyer RB, Hedgpeth J (2000) 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res* 28(2): 655-61.
- Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet* 2: 235-58.
- Landegren U, Kaiser R, Sanders J, Hood L (1998) A ligase-mediated gene detection technique. *Science* 241(4869): 1077-80.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822): 860-921.

- Landsteiner K (1901) Über Agglutinationserscheinungen normalen menschlichen Blutes. *Wien Klin Wochenschrift* 14: 1132-1134.
- Lao O, van Duijn K, Kersbergen P, de Knijff P, Kayser M (2006) Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *Am J Hum Genet* 78(4): 680-90.
- Lao O, L T, Nothnagel M, Krawczak M, Kayser M (2008) Genome-wide SNP scans for finding markers to differentiate European populations. *Progr Forensic Genet* 12: (in press).
- Lareu M, Phillips C, Fondevila M, Carracedo A (2008) Resolution of paternity investigations showing ambiguous STR results using autosomal SNPs as supplementary markers. *Progr Forensic Genet* 12: (in press).
- Lessig R, Wenzel V, Weber M (2006) Bite mark analysis in forensic routine case work. *EXCLI J* 5: 93-102.
- Li J, Harris L, Mamon H, Kulke MH, Liu WH, Zhu P, Mike Makrigiorgos G (2006) Whole genome amplification of plasma-circulating DNA enables expanded screening for allelic imbalance in plasma. *J Mol Diagn* 8(1): 22-30.
- Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 24(4): 381-6.
- Lindroos K, Liljedahl U, Raitio M, Syvanen AC (2001) Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res* 29(13): E69-9.
- Lindroos K, Sigurdsson S, Johansson K, Ronnblom L, Syvanen AC (2002) Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Res* 30(14): e70.
- Lovmar L, Syvanen AC (2006) Multiple displacement amplification to create a long-lasting source of DNA for genetic studies. *Hum Mutat* 27(7): 603-14.
- Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 74(3): 532-44.
- Lukhtanov EA, Lokhov SG, Gorn VV, Podyminogin MA, Mahoney W (2007) Novel DNA probes with low background and high hybridization-triggered fluorescence. *Nucleic Acids Res* 35(5): e30.
- Martin PD, Schmitter H, Schneider PM (2001) A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Sci Int* 119(2): 225-31.
- McEvoy B, Beleza S, Shriver MD (2006) The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Hum Mol Genet* 15 Spec No 2: R176-81.
- Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N (2005) Typing of Multiple Single-Nucleotide Polymorphisms Using Ribonuclease Cleavage of DNA/RNA Chimeric Single-Base Extension Primers and Detection by MALDI-TOF Mass Spectrometry. *Anal Chem* 77: 5229-5235.
- Miller RD, Phillips MS, Jo I, Donaldson MA, Studebaker JF, et al. (2005) High-density single-nucleotide polymorphism maps of the human genome. *Genomics* 86(2): 117-26.
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3): 1215.
- Mousses S, Caplen NJ, Cornelison R, Weaver D, Basik M, Hautaniemi S, Elkahloun AG, Lotufo RA, Choudary A, Dougherty ER, Suh E, Kallioniemi O (2003) RNAi microarray analysis in cultured mammalian cells. *Genome Res* 13(10): 2341-7.
- Musgrave-Brown E, Ballard D, Balogh K, Bender K, Berger B, et al. (2007) Forensic validation of the SNPforID 52-plex assay. *Forensic Science International: Genetics* 1(2): 186-190.
- Musgrave-Brown E, Ballard D, Fondevila-Alvarez M, Harrison C, Phillips C, Prasad Y, Sobrino B, Thacker C, Willhun J, Carracedo A, Syndercombe-Court D, Schneider PM (2008) Forensic validation of the Genplex SNP typing system - results of an interlaboratory study. *Forensic Sci Int: Gen Supplement* 1(1): 389-393.
- Nagy M, Rascon J, Massenkeil G, Ebell W, Roewer L (2006) Evaluation of whole-genome amplification of low-copy-number DNA in chimerism analysis after allogeneic stem cell transplantation using STR marker typing. *Electrophoresis* 27(15): 3028-37.
- Nebert DW (1999) Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin Genet* 56(4): 247-58.



- Nikiforov TT, Rendle RB, Kotewicz ML, Rogers YH (1994) The use of phosphorothioate primers and exonuclease hydrolysis for the preparation of single-stranded PCR products and their detection by solid-phase hybridization. *PCR Methods Appl* 3(5): 285-91.
- Nyren P (2007) The history of pyrosequencing. *Methods Mol Biol* 373: 1-14.
- Oscarson M (2003) Pharmacogenetics of drug metabolising enzymes: importance for personalised medicine. *Clin Chem Lab Med* 41(4): 573-80.
- Panelli S, Damiani G, Espen L, Micheli G, Sgaramella V (2006) Towards the analysis of the genomes of single cells: further characterisation of the multiple displacement amplification. *Gene* 372: 1-7.
- Parsons TJ, Coble MD (2001) Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croat Med J* 42(3): 304-9.
- Pask R, Rance HE, Barratt BJ, Nutland S, Smyth DJ, Sebastian M, Twells RC, Smith A, Lam AC, Smink LJ, Walker NM, Todd JA (2004) Investigating the utility of combining phi29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray genotyping. *BMC Biotechnol* 4: 15.
- Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC (1997) Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res* 7(6): 606-14.
- Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* 10(7): 1031-42.
- Phillips C, Lareu M, Sanchez JJ, Brion M, Sobrino B, Morling N, Schneider PM, Syndercombe-Court D, Carracedo A (2004a) Selecting single nucleotide polymorphisms for forensic applications. *International Congress Series* 1261: 18-20.
- Phillips C, Lareu V, Salas A, Carracedo A (2004b) Nonbinary single-nucleotide polymorphism markers. *Progr Forensic Genet* 10: 27-29.
- Phillips C (2005) Using online databases for developing SNP markers of forensic interest. *Methods Mol Biol* 297: 83-106.
- Phillips C (2007) Online resources for SNP analysis: a review and route map. *Mol Biotechnol* 35(1): 65-97.
- Phillips C, Fang R, Ballard D, Fondevila M, Harrison C, Hyland F, Musgrave-Brown E, Proff C, Ramos-Luis E, Sobrino B, Carracedo A, Furtado MR, Syndercombe-Court D, Schneider PM (2007) Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. *Forensic Sci Int: Gen* 1: 180-185.
- Phillips C, Fondevila M, García-Magariños M, Rodriguez A, Salas A, Carracedo A, Lareu MV (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International: Genetics* 2(3): 198-204.
- Prinz M, Carracedo A, Mayr WR, Morling N, Parsons TJ, Sajantila A, Scheithauer R, Schmitter H, Schneider PM (2007) DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Science International: Genetics* 1(1): 3-12.
- Pusch W, Wurmbach JH, Thiele H, Kostrzewa M (2002) MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* 3(4): 537-48.
- Quintans B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 140(2-3): 251-7.
- Radtkey R, Feng L, Muralhidar M, Duhon M, Canter D, DiPierro D, Fallon S, Tu E, McElfresh K, Nerenberg M, Sosnowski R (2000) Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip. *Nucleic Acids Res* 28(7): E17.
- Raitio M, Lindroos K, Laukkanen M, Pastinen T, Sistonen P, Sajantila A, Syvanen AC (2001) Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. *Genome Res* 11(3): 471-82.
- Relling MV (1989) Polymorphic drug metabolism. *Clin Pharm* 8(12): 852-63.
- Rico F, Moy VT (2007) Energy landscape roughness of the streptavidin-biotin interaction. *J Mol Recognit*.
- Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, et al. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75(1): 128-37.

- Roper A, Reichert W, Mattern R (2007) [Analysis of mitochondrial SNPs in addition to conventional STR-typing in a case of aggravated theft]. *Arch Kriminol* 219(3-4): 98-104.
- Ross P, Hall L, Smirnov I, Haff L (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol* 16(13): 1347-51.
- Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67(6): 1526-43.
- Rouillard JM, Zuker M, Gulari E (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* 31(12): 3057-62.
- Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423(6942): 873-6.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409(6822): 928-33.
- Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, et al. (2007) Genomewide association analysis of coronary artery disease. *N Engl J Med* 357(5): 443-53.
- Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N (2003) Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Sci Int* 137(1): 74-84.
- Sanchez JJ, Brion M, Parson W, Blanco-Verea AJ, Borsting C, Lareu M, Niederstatter H, Oberacher H, Morling N, Carracedo A (2004) Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications. *Forensic Sci Int* 140(2-3): 241-50.
- Sanchez JJ, Hallenberg C, Borsting C, Hernandez A, Morling N (2005) High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. *Eur J Hum Genet* 13(7): 856-66.
- Sanchez JJ, Phillips C, Borsting C, Bogus M, Carracedo A, Syndercombe-Court D, Fondevila M, Harrison C, Morling N, Balogh K, Schneider PM (2006a) Development of a multiplex PCR assay detecting 52 autosomal SNPs. *Progr Forensic Genet* 11: 67-69.
- Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006b) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27(9): 1713-24.
- Sanchez JJ, Børsting C, Balogh K, Berger B, Bogus M, Butler JM, Carracedo A, Syndercombe Court D, Dixon LA, Filipović B, Fondevila M, Gill P, Harrison CD, Hohoff C, Huel R, Ludes B, Parson W, Parsons TJ, Petkovski E, Phillips C, Schmitter H, Schneider PM, Vallone PM, Morling N (2008) Forensic typing of autosomal SNPs with a 29 SNP-multiplex—Results of a collaborative EDNAP exercise. *Forensic Science International: Genetics* 2(3): 176-183.
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235): 467-70.
- Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C (1998) Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci Int* 97(1): 61-70.
- Schneider PM, Martin PD (2001) Criminal DNA databases: the European situation. *Forensic Sci Int* 119(2): 232-8.
- Schneider PM, Balogh MK, Naveran N, Bogus M, Burger J, Bender K, Lareu M, Carracedo A (2004) Whole genome amplification—the solution for a common problem in forensic casework? *Progr Forensic Genet* 10: 24-26.
- Schneider PM (2007) Scientific standards for studies in forensic genetics. *Forensic Sci Int* 165(2-3): 238-43.
- Schneider PM (2008) DNA-based prediction of physical traits – a new dimension for forensic genetics, or a first step towards violation of privacy leading to genetic discrimination? *Progr Forensic Genet* 12: (in press).
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290(5494): 1155-9.

- Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* 74(5): 1023-34.
- Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW (2004) The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics* 1(4): 274-86.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942): 825-37.
- Sobrinho B, Lareu M, Brion M, Carracedo A (2004) SNP genotyping with single base extension-tag microarrays. *Progr Forensic Genet* 10: 331-333.
- Sobrinho B, Brion M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154(2-3): 181-94.
- Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 20(4): 399-407.
- Sosnowski R, Heller MJ, Tu E, Forster AH, Radtkey R (2002) Active microelectronic array system for DNA hybridization, genotyping and pharmacogenomic applications. *Psychiatr Genet* 12(4): 181-92.
- Sosnowski RG, Tu E, Butler WF, O'Connell JP, Heller MJ (1997) Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. *Proc Natl Acad Sci U S A* 94(4): 1119-23.
- Southern E (1979) Gel electrophoresis of restriction fragments. *Methods Enzymol* 68: 152-76.
- Sun G, Kaushal R, Pal P, Wolujewicz M, Smelser D, Cheng H, Lu M, Chakraborty R, Jin L, Deka R (2005) Whole-genome amplification: relative efficiencies of the current methods. *Leg Med (Tokyo)* 7(5): 279-86.
- Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2(12): 930-42.
- Thorisson GA, Stein LD (2003) The SNP Consortium website: past, present and future. *Nucleic Acids Res* 31(1): 124-7.
- Tong AK, Ju J (2002) Single nucleotide polymorphism detection by combinatorial fluorescence energy transfer tags and biotinylated dideoxynucleotides. *Nucleic Acids Res* 30(5): e19.
- Tonisson N, Kurg A, Kaasik K, Lohmussaar E, Metspalu A (2000) Unravelling genetic data by arrayed primer extension. *Clin Chem Lab Med* 38(2): 165-70.
- Tonisson N, Zernant J, Kurg A, Pavel H, Slavin G, Roomere H, Meiel A, Hainaut P, Metspalu A (2002) Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. *Proc Natl Acad Sci U S A* 99(8): 5503-8.
- Tzvetkov MV, Becker C, Kulle B, Nurnberg P, Brockmoller J, Wojnowski L (2005) Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification. *Electrophoresis* 26(3): 710-5.
- Uhlenhuth P (1901) Eine Methode zur Unterscheidung der verschiedenen Blutarten, im Besonderen zum differentialdiagnostischen Nachweise des Menschenblutes. *Dtsch Med Wochenschrift* 6: 82-83.
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26(3): 358-61.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65(Pt 1): 43-62.
- Vallone PM, Butler JM (2004) Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *J Forensic Sci* 49(4): 723-32.
- Vallone PM, Decker AE, Butler JM (2005) Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples. *Forensic Sci Int* 149(2-3): 279-86.

- van Daal A, Salini Manoharan MN (2008) Human Pigmentation Genes. *Progr Forensic Genet* 12: (in press).
- Weidenhammer EM, Kahl BF, Wang L, Duhon M, Jackson JA, Slater M, Xu X (2002) Multiplexed, targeted gene expression profiling and genetic analysis on electronic microarrays. *Clin Chem* 48(11): 1873-82.
- Weir BS, Hill WG (2002) Estimating F-statistics. *Annu Rev Genet* 36: 721-50.
- Werle E, Schneider C, Renner M, Volker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Res* 22(20): 4354-5.
- Westin L, Xu X, Miller C, Wang L, Edman CF, Nerenberg M (2000) Anchored multiplex amplification on a microelectronic chip array. *Nat Biotechnol* 18(2): 199-204.
- Wiegand P, Kleiber M (1997) DNA typing of epithelial cells after strangulation. *Int J Legal Med* 110(4): 181-3.
- Willer CJ, Bonnycastle LL, Conneely KN, Duren WL, Jackson AU, et al. (2007) Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes* 56(1): 256-64.
- Willuweit S, Roewer L (2007) Y chromosome haplotype reference database (YHRD): Update. *Forensic Science International: Genetics* 1(2): 83-87.
- Wise CA, Paris M, Morar B, Wang W, Kalaydjieva L, Bittles AH (2003) A standard protocol for single nucleotide primer extension in the human genome using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 17(11): 1195-202.
- Y-Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12(2): 339-48.
- Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 21(1): 164-75.
- Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet* 71(3): 466-82.
- Zhang L, Wurzner R, Stradmann-Bellinghausen B, Rittner C, Schneider PM (1994) A combined study of human complement C7 IEF and C7 M/N polymorphisms in the Chinese Han population. *Exp Clin Immunogenet* 11(1): 17-22.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13): 3406-15.

## 7 Anhang

### 7.1 Abkürzungsverzeichnis

aCGH	Microarray-based Comperative Genomic Hybridisation
AIMs	Ancestry-Informative Markers
AMOVA	Analysis of Molecular Variance
APEX	Arrayed Primer Extension
Arg	Arginin
ASO	Allele-Specific Oligonucleotide Probes
ASR	Analyte Specific Reagents
Auto1	erste autosomale Multiplex für 23 SNPs
Auto2	zweite autosomale Multiplex für 29 SNPs
AZFc	AZFc Region des Y Chromosoms (q-Arm)
BKA	Bundeskriminalamt
BLAST	Basic Local Alignment Search Tool
BL-WGA	Blunt-End Ligation-Mediated WGA
bp	Basenpaar
BSA	Bovine Serum Albumin
BSE	Base Stacking Energy
CCD	Charged Coupled Device
CMOS	Complementary Metal Oxide Semiconductor
CNVs	Copy Number Variations
CODIS	Combined DNA Index System
cSNPs	coding-SNPs
CTB	Cholera Toxin B
DAD	DNA-Analysedatei
ddNTP	2',3'-Didesoxynucleosid-5'-triphosphat
DNA	Desoxyribonukleinsäure
DNA-IFG	DNA-Identitätsfeststellungsgesetzes
dNTP	2'-Desoxynucleosid-5'-triphosphat
Dp	Diskriminierungsindex
EDNAP	European DNA Profiling Group
ENFSI	European Network of Forensic Science Institutes
ESS-Loci	European Standard Set of Loci
EU	Europäische Union
Exo I	Escheria coli Exonuclease I
FoDNANovG	Forensische DNA Novellierungs Gesetz
FRET	Fluorescence Resonance Energytransfer
His	Histidine
HPLC	High Performance Liquid Chromatography
HSB	High Salt Buffer
Interpol	Internationale kriminalpolizeiliche Organisation
IP	Isoelektrischen Punkt
iscan	initial scan
ISFG	International Society for Forensic Genetics
iSNPs	intergenetic-SNPs

ISSOL	Interpol Standard Set of Loci
LCN	Low Copy Number
LED	Light Emitting Diode
LSB	Low Salt Buffer
Lsg	Arbeitslösung
LSO	Locus Specific Oligo
M	Homozygot-Mutante
MAF	Minor Allele Frequency
Maldi-TOF	Matrix-Assisted Laser Desorption Ionisation Time-Of-Flight
MATP	Monomeric Avidin Triethylamine Purification
MBW	NanoChip® Molecular Biology Workstation
MC1R	Melanocortin 1-Rezeptorgen
MDS	Multidimensional Scaling
MGB	Minor Groove Binding
MW	Heterozygot
NC400	NanoChip® 400
NCBI	National Centre of Biotechnology Information
NDNAD	National DNA Database, UK
nt	Nukleotide
OLA	Oligonucleotid Ligation Assay
PCR	Polymerase Chain Reaction
PD	Power of discrimination
PM	Probability of a match
PMT	Photomultiplier Detector
PPI	Pyrophosphat
pSNPs	perigenic-SNPs
PSV	Paralogous Sequence Variant
RFLP	Restriction Fragment Length Polymorphism
RFUs	Relative Fluorescence Units
SAP	Shrimp Alkaline Phosphatase
SBE	Single Base Extension
SDA	Multiple Displacement Amplification
SEB	Staphylokokkus Enterotoxin B
SNP	Single Nucleotide Polymorphism
SNPforID	High Throughput Analysis of Single Nucleotide Polymorphisms for the Forensic Identification of Persons
SSCP	Single-Strand Conformational Polymorphism
ssDNA	single stranded DNA
StPO	Strafprozessordnung
STR	Short Tandem Repeat
T <sub>m</sub>	Schmelztemperatur
USP9	Ubiquitin-Specific Protease 9
W	Homozygot-Wildtyp
WGA	Whole Genome Amplification
WP	Work Package
YCC	Y-Chromosome Consortium
Φ <sub>ST</sub>	Differentiationsindex



Tabelle 40: Ergebnisse der SNaPshot-Genotypisierung für 29 Y-chromosomale SNPs und 40 Proben aus Thailand. Die unterste Zeile gibt die Haplogruppenzuordnung für jede Probe wieder.

SNP	Hg <sup>1)</sup>	M <sup>2)</sup>	Proben aus Thailand																											
			T1 <sup>3)</sup>	T3	T10	T11	T12	T13	T15	T16	T20	T21	T22 <sup>3)</sup>	T23	T24	T26	T27	T31	T34	T71	T85	T86	T87	T88	T89	T93	T95	T100	T108	T109
SRY	BR	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M168	CR*	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M216	C*	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	G	G	G	G	G	G	G	G	G	G	G
M38	C2	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M217	C3	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M174	D	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M96	E*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M33	E1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M75	E2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
P2	E3*	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M2	E3a	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M35	E3b*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M81	E3b2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M123	E3b3	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
M201	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M170	I	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M304	J	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M9	K*	C	C	G	C	C	G	C	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
M128	N1	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Tat	N3	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M175	O*	A	A	T	A	A	T	A	A	T	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A
M119	O1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
P31	O2	C	T	T	T	T	T	C	T	C	T	C	T	T	T	T	T	T	T	T	C	C	T	T	T	T	C	C	T	T
M122	O3	G	A	G	G	A	G	A	G	A	G	A	G	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M45	P*	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
M207	R*	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M173	R1*	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M17	R1a1	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
P25	R1b	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
			O3	CR	O3	O3	CR	O3	O2	CR	O3	O2	O3	O3	O1	O3	O3	C	C3	O	O2	O2	O3	O3	O3	O1	O2	O2	O3	C3

- 1) Haplogruppenzuordnung.
- 2) Mutationsallel.
- 3) Fand eine PCR-Reamplifikation statt.

Tabelle 40: Fortsetzung.

SNP	Hg <sup>1)</sup>	M <sup>2)</sup>	Proben aus Thailand														
			T112	T121	T124	T128	T136	T139	T142	T147 <sup>3)</sup>	T152	T153	T170	T175			
SRY	BR	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M168	CR*	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M216	C*	A	G	G	G	G	G	G	G	A	G	G	G	G	G	G	
M38	C2	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
M217	C3	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
M174	D	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
M96	E*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
M33	E1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
M75	E2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
P2	E3*	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
M2	E3a	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
M35	E3b*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
M81	E3b2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
M123	E3b3	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
M201	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
M170	I	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
M304	J	C	A	A	A	C	A	A	A	A	A	A	A	A	A	A	
M9	K*	C	C	C	C	G	C	C	C	G	C	C	C	C	C	C	
M128	N1	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
Tat	N3	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
M175	O*	A	A	A	T	T	A	A	A	T	A	A	A	A	A	T	
M119	O1	G	G	T	T	T	T	G	T	T	T	T	T	T	T	T	
P31	O2	C	T	T	T	T	T	C	T	T	T	T	T	T	T	T	
M122	O3	G	A	G	A	A	A	A	G	A	G	A	A	A	A	A	
M45	P*	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
M207	R*	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
M173	R1*	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
M17	R1a1	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
P25	R1b	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
			O1	O3	K	J	O2	O1	O3	C	O3	O2	O2	O2	K		

- 1) Haplogruppenzuordnung.
- 2) Mutationsallel.
- 3) Fand eine PCR-Reamplifikation statt.



Tabelle 41: Ergebnisse der SNaPshot-Genotypisierung für 29 Y-chromosomale SNPs und 40 Proben aus Japan. Die unterste Zeile gibt die Haplogruppenzuordnung für jede Probe wieder.

SNP	Hg <sup>1)</sup>	M <sup>2)</sup>	Proben aus Japan																											
			JP1	JP2	JP3	JP4 <sup>3)</sup>	JP5	JP6 <sup>3)</sup>	JP7 <sup>3)</sup>	JP8 <sup>3)</sup>	JP9 <sup>3)</sup>	JP10 <sup>3)</sup>	JP11	JP12 <sup>3)</sup>	JP13 <sup>3)</sup>	JP14	JP15 <sup>3)</sup>	JP16 <sup>3)</sup>	JP17 <sup>3)</sup>	JP18	JP19	JP20 <sup>3)</sup>	JP21	JP22 <sup>3)</sup>	JP23	JP24 <sup>3)</sup>	JP25	JP26	JP27 <sup>3)</sup>	JP28
SRY	B R	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
M168	CR*	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
M216	C*	A	G	G	G	G	G	G	G	G	A	G	G	G	A	G	G	G	G	G	G	A	G	G	G	G	G	G	G	
M38	C2	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
M217	C3	G	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T		
M174	D	C	T	T	T	TC	T	TC	T	T	T	T	T	TC	T	T	T	T	T	T	T	T	T	T	TC	TC	T	TC		
M96	E*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
M33	E1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
M75	E2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
P2	E3*	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
M2	E3a	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
M35	E3b*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
M81	E3b2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
M123	E3b3	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
M201	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
M170	I	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
M304	J	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
M9	K*	C	C	C	C	G	C	G	C	C	G	C	C	G	C	C	C	C	C	C	C	G	C	C	G	G	C	G		
M128	N1	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
Tat	N3	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
M175	O*	A	A	A	A	T	A	T	A	T	A	T	A	T	A	A	A	A	A	A	A	T	A	A	T	T	T	T		
M119	O1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
P31	O2	C	T	C	T	C	T	T	C	T	T	C	T	T	C	T	C	T	C	T	T	C	C	T	T	T	C	T		
M122	O3	G	A	G	A	A	A	G	A	A	G	A	A	A	G	G	A	A	A	A	G	A	A	A	A	A	A	A		
M45	P*	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
M207	R*	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
M173	R1*	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
M17	R1a1	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
P25	R1b	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
			O2	O3	O2	D	O2	D	O3	O2	C3	O3	O2	D	C	O3	O3	O2	O1	O2	O3	C3	O2	O2	CR	D	D	O2	D	

- 1) Haplogruppenzuordnung.
- 2) Mutationsallel.
- 3) Fand eine PCR-Reamplifikation statt und eine dritte Amplifikation von originaler DNA.

Tabelle 41: Fortsetzung.

SNP	Hg <sup>1)</sup>	M <sup>2)</sup>	Proben aus Japan																
			JP29 <sup>3)</sup>	JP30 <sup>3)</sup>	JP31 <sup>3)</sup>	JP32 <sup>3)</sup>	JP33 <sup>3)</sup>	JP34 <sup>3)</sup>	JP35	JP36	JP37	JP38 <sup>3)</sup>	JP39	JP40					
SRY	B R	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
M168	CR*	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
M216	C*	A	G	G	G	G	G	G	A	G	G	G	G	G	G	G			
M38	C2	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
M217	C3	G	T	T	T	T	T	T	G	T	T	T	T	T	T	T			
M174	D	C	TC	TC	T	T	T	T	TC	T	T	T	TC	TC	T	T			
M96	E*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
M33	E1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
M75	E2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
P2	E3*	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
M2	E3a	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
M35	E3b*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
M81	E3b2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
M123	E3b3	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
M201	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
M170	I	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
M304	J	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
M9	K*	C	G	G	C	C	C	G	C	G	C	C	G	C	G	G			
M128	N1	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
Tat	N3	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
M175	O*	A	T	T	A	A	A	T	T	T	A	A	T	T	T	T			
M119	O1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
P31	O2	C	T	T	C	T	C	T	T	C	T	T	C	C	T	T			
M122	O3	G	A	A	A	G	A	A	A	A	A	A	A	A	A	A			
M45	P*	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
M207	R*	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
M173	R1*	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
M17	R1a1	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
P25	R1b	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
			D	D	O2	O3	O2	K	D	C3	O2	O2	D	D	D	D			

- 1) Haplogruppenzuordnung.
- 2) Mutationsallel.
- 3) Fand eine PCR-Reamplifikation statt und eine dritte Amplifikation von originaler DNA.



Tabelle 42: Fortsetzung.

Table with columns for SNP, Hg1, M2, and Proben aus Deutschland (G60-G8, G\*2-G\*8) and rows for various SNP identifiers like SRY, M168, M216, etc.

- 1) Haplogruppenzuordnung.
2) Mutationsallel.

Tabelle 42: Fortsetzung.

Table with columns for SNP, Hg1, M2, and Proben aus Deutschland (G\*9-G\*36) and rows for various SNP identifiers like SRY, M168, M216, etc.

- 1) Haplogruppenzuordnung.
2) Mutationsallel.

Tabelle 42: Fortsetzung.

SNP	Hg <sup>1)</sup>	M <sup>2)</sup>	Proben aus Deutschland																											
			G*37	G*38	G*39	G*40	G*41	G*42	G*43	G*44	G*45	G*46	G*47	G*48	G*49	G*50	G*51	G*52	G*55	G*56	G*57	G*58	G*59	G*60	G*62	G*63	G*65	G*66	G*67	
SRY	BR	G	G	G	A	A	G	G	G	G	A	G	G	G	G	G	G	A	G	G	G	G	A	A	G	A	A	G	A	G
M168	CR*	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M216	C*	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M38	C2	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M217	C3	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M174	D	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M96	E*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M33	E1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M75	E2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
P2	E3*	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M2	E3a	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M35	E3b*	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M81	E3b2	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M123	E3b3	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
M201	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
M170	I	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M304	J	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M9	K*	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
M128	N1	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Tat	N3	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M175	O*	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M119	O1	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
P31	O2	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
M122	O3	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M45	P*	T	T	T	T	T	C	T	T	C	C	T	T	T	C	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T
M207	R*	G	G	G	G	G	A	A	G	A	A	G	G	G	G	A	G	G	G	G	G	G	A	G	G	G	A	G	A	G
M173	R1*	C	C	C	C	C	A	A	C	A	A	C	C	C	C	A	C	C	C	C	C	C	A	C	C	C	A	C	A	C
M17	R1a1	G	C	C	C	G	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
P25	R1b	A	CA	CA	CA	C	C	C	CA	C	C	C	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA
			R1b	R1b	R1b	R1a1	R1a1	N3	P	R1b	I	I	R1a1	R1b	R1b	R1b	I	R1b	R1b	R1a1	R1b	R1b	I	R1b	R1a1	R1a1	I	R1a1	P	

1) Haplogruppenzuordnung.  
2) Mutationsallel.

Tabelle 42: Fortsetzung.

SNP	Hg <sup>1)</sup>	M <sup>2)</sup>	Proben aus Deutschland														
			G*68	G*69	G*70	G*71	G*72	G*74	G*75	G*76	G*77	G*78	G*79	G*80			
SRY	BR	G	G	G	A	A	G	G	G	G	G	G	G	G	G		
M168	CR*	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
M216	C*	A	G	G	G	G	G	G	G	G	G	G	G	G	G		
M38	C2	G	T	T	T	T	T	T	T	T	T	T	T	T	T		
M217	C3	G	T	T	T	T	T	T	T	T	T	T	T	T	T		
M174	D	C	T	T	T	T	T	T	T	T	T	T	T	T	T		
M96	E*	C	G	G	G	G	G	G	G	C	G	G	C	G	C		
M33	E1	G	T	T	T	T	T	T	T	T	T	T	T	T	T		
M75	E2	A	G	G	G	G	G	G	G	G	G	G	G	G	G		
P2	E3*	A	G	G	G	G	G	G	G	A	G	G	A	G	A		
M2	E3a	C	T	T	T	T	T	T	T	T	T	T	T	T	T		
M35	E3b*	C	G	G	G	G	G	G	G	C	G	G	C	G	C		
M81	E3b2	A	G	G	G	G	G	G	G	G	G	G	G	G	G		
M123	E3b3	T	C	C	C	C	C	C	C	C	C	C	C	C	C		
M201	G	T	G	G	G	G	T	T	G	G	G	G	G	G	G		
M170	I	C	C	A	A	A	A	A	A	A	A	A	A	A	A		
M304	J	C	A	A	A	A	A	A	A	A	A	A	C	C	A		
M9	K*	C	G	C	C	C	G	G	C	C	G	G	G	G	G		
M128	N1	T	A	A	A	A	A	A	A	A	A	A	A	A	A		
Tat	N3	G	A	A	A	A	A	A	A	A	A	A	A	A	A		
M175	O*	A	T	T	T	T	T	T	T	T	T	T	T	T	T		
M119	O1	G	T	T	T	T	T	T	T	T	T	T	T	T	T		
P31	O2	C	T	T	T	T	T	T	T	T	T	T	T	T	T		
M122	O3	G	A	A	A	A	A	A	A	A	A	A	A	A	A		
M45	P*	T	C	T	T	T	C	C	T	T	C	C	C	C	C		
M207	R*	G	A	G	G	G	A	A	G	G	A	A	A	A	A		
M173	R1*	C	A	C	C	C	A	A	C	C	A	A	A	A	A		
M17	R1a1	G	C	C	G	G	C	C	C	C	C	C	C	C	C		
P25	R1b	A	C	CA	C	C	C	C	CA	CA	C	C	C	C	C		
			I	R1b	R1a1	R1a1	G	G	R1b	R1b	E3b	J	J	E3b3			

1) Haplogruppenzuordnung.  
2) Mutationsallel.

## 7.3 Vorveröffentlichungen

*Electrophoresis* 2005, 26, 4411–4420

4411

María Brión<sup>1</sup>  
 Juan J. Sanchez<sup>2</sup>  
 Kinga Balogh<sup>3</sup>  
 Catherine Thacker<sup>4</sup>  
 Alejandro Blanco-Verea<sup>1</sup>  
 Claus Børsting<sup>2</sup>  
 Beate Stradmann-Bellinghausen<sup>3</sup>  
 Magdalena Bogus<sup>3</sup>  
 Denise Syndercombe-Court<sup>4</sup>  
 Peter M. Schneider<sup>3</sup>  
 Angel Carracedo<sup>1</sup>  
 Niels Morling<sup>2</sup>

<sup>1</sup>Institute of Legal Medicine,  
 National Genotyping Center  
 (CEGEN), University of  
 Santiago de Compostela, Spain

<sup>2</sup>Department of Forensic Genetics,  
 Institute of Forensic Medicine,  
 University of Copenhagen,  
 Denmark

<sup>3</sup>Institute of Legal Medicine,  
 University of Mainz, Germany

<sup>4</sup>Barts and the London,  
 Queen Mary's School of  
 Medicine and Dentistry,  
 London, UK

### Introduction of an single nucleotide polymorphism-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages

The European Consortium “High-throughput analysis of single nucleotide polymorphisms for the forensic identification of persons – SNPforID”, has performed a selection of candidate Y-chromosome single nucleotide polymorphisms (SNPs) for making inferences on the geographic origin of an unknown sample. From more than 200 SNPs compiled in the phylogenetic tree published by the Y-Chromosome Consortium, and looking at the population studies previously published, a package of 29 SNPs has been selected for the identification of major population haplogroups. A “Major Y-chromosome haplogroup typing kit” has been developed, which allows the multiplex amplification of all 29 SNPs in a single reaction. Allele genotyping was performed with a single base extension reaction (minisequencing) detected by CE. The validation of the multiplex was performed in a total of 1126 unrelated males distributed among 12 worldwide populations. The approach takes advantage of the specific geographic distribution of the Y-chromosome haplogroups and demonstrates the utility of binary polymorphisms to infer the origin of a male lineage.

**Keywords:** Capillary electrophoresis / Geographic origin / Multiplex / Single base extension / Y-SNPs  
 DOI 10.1002/elps.200500293

### 1 Introduction

The utility of Y-chromosome-specific markers has been already well established in many different areas. Their genetic diversity can be used to provide information on male-specific patterns of migration in the past, and the origin and diversity of specific populations [1–4]. In addition, in the forensic field they are shown to be useful for screening samples from sexual assault cases as well as to provide evidence on the male lineage in deficiency paternity cases [5–7].

Because of the lack of recombination in this genomic region, the polymorphic markers described cannot be considered as independent and therefore haplotypes have to be constructed. Desirable characteristics of markers are a high level of polymorphism in the popula-

tion, the lowest possible incidence of recurrent mutations, and a maximum of simplicity, allowing their detection in multiplex reactions and in small fragments of DNA. Polymorphisms such as microsatellites, historically used in the forensic field, show high levels of polymorphism but they are prone to recurrent mutational events that may generate haplotypes identical by state [8]. In contrast, single nucleotide polymorphisms (SNPs) which may include small insertions and deletions are amenable to classic phylogenetic analysis; however, they are not very polymorphic. This disadvantage is compensated by the simplicity of the markers, allowing the analysis of a high number of SNPs in a single multiplex reaction, using small quantities of DNA [9, 10].

In the case of autosomal SNPs used in a forensic context, the required number of SNPs to give comparable levels of information to STRs has already been determined [11]. However, as the Y-chromosome exhibits extensive genetic differentiation, the selection of SNPs is far more complicated. This differentiation across geographical distance [12] results in markedly different haplogroup profiles distribution, depending on the region of the world studied [1, 13].

In 2002, the Y-Chromosome Consortium (YCC) [14] has compiled 245 binary polymorphisms in a parsimonious tree defining 153 haplogroups. Subsequent publications

**Correspondence:** Dr. María Brión, Institute of Legal Medicine, National Genotyping Center (CEGEN), University of Santiago de Compostela, San Francisco s/n. E-15782, Santiago de Compostela, Spain

**E-mail:** brioniml@usc.es

**Fax:** +34-98-158-0336

**Abbreviations:** AMOVA, molecular analysis of the variance; SBE, single base extension; SNP, single nucleotide polymorphism; YCC, Y-Chromosome Consortium

[15–17] have increased the number of SNPs, but all of them are classified into the 18 major clades (A–R) defined by the YCC. It has been described that most of the haplogroups belonging to each of these major clades tend to be confined to specific continents [15], so this geographical distribution can be used to make inferences on the possible geographic origin of any sample of interest. In a forensic scenario this may serve as a tool to predict the origin of the paternal lineage of any stain contributor.

A large number of SNP genotyping methods are now available [18, 19], and usually the choice of the appropriate method depends on the number of SNPs and the number of individuals that need to be typed. The work presented here was planned to be used for laboratories involved in forensic DNA analysis, so the selected genotyping method was a multiplex PCR followed by single base extension (SBE) reaction using the SNaPshot multiplex kit. The main reason for this selection was the possibility of genotyping SNPs without investment in new technologies, simply using a CE instrument, which is available in most of these laboratories for multiplex STR typing by fragment analysis.

The main objectives of the present study were, first, the selection of candidate Y-SNPs to provide clues on the possible geographic origin of a sample. The second objective was the development of multiplex strategies, allowing the analysis of as many SNPs as possible in a single reaction. Finally, the last goal was the validation of the strategy in a sample of more than one thousand unrelated males distributed all around the world, to check the accuracy and reproducibility of the multiplex, as well as the potential to differentiate between the major population groups.

## 2 Materials and methods

### 2.1 Samples and DNA purification

Nine human DNA samples from eight males and one female and one dog male were distributed among the participating labs, in order to compare the results of the new multiplex assay. The dog sample was used to check the possibility of amplification because of homologous regions between the human and dog Y chromosome.

A total of 1 126 unrelated males from Denmark (150), Greenland (90), Turkey (51), Thailand (84), China (55), Japan (40), Germany (150), Galicia (130), Mozambique (130), Somalia (105), Argentina (81 racially mixed individuals), and Colombia (60 ethnically interbred individuals) were typed. The samples were distributed equally among three of the participating labs (Copenhagen,

Mainz, and Santiago de Compostela to be typed for the 29 SNPs. Blood on FTA cards, DNA purified by phenol/chloroform extraction or the QIAamp DNA blood mini kit (Qiagen) was used. DNA concentration was determined by real-time PCR using the Quantifiler™ Human DNA Quantification Kit and the ABI 7300 instrument (Applied Biosystems, Foster City, CA, USA) or using the NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, Rockland, DE). The protocols were approved by the local ethical committees (Copenhagen: KF-01-037/03, Santiago USC-FM 030104).

### 2.2 Selection of Y-chromosome markers

From population studies previously published [1, 20–22] a package of 29 SNPs was selected for the identification of major population haplogroups present all around the world. The 29 SNPs and the haplogroup tree defined with them are represented in Fig. 1. Twelve SNPs define major clades and the others subdivide some of these clades.

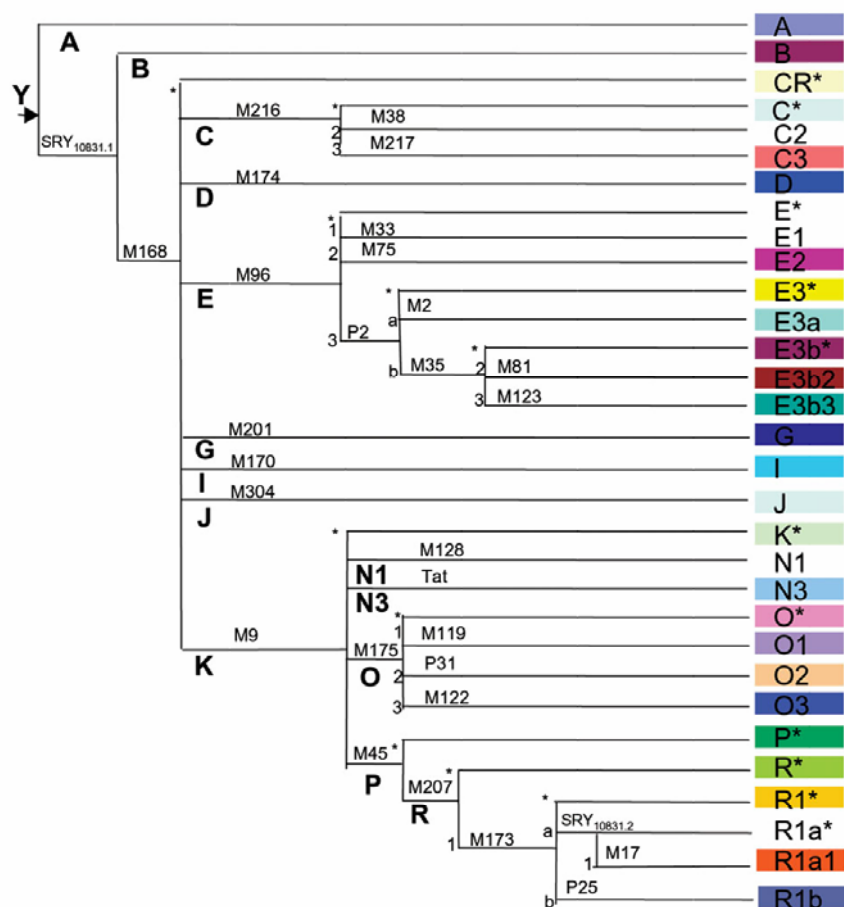
### 2.3 Primer design

PCR primers (Table 1) and SBE primers (Table 2) were designed according to the guidelines outlined in Sanchez *et al.* [9]. HPLC-purified oligonucleotides were purchased from DNA Technology A/S (Aarhus, Denmark). The quality of the oligonucleotides was investigated by MALDI-TOF analysis.

### 2.4 PCR conditions and purification of PCR products

The Major Y Haplogroup multiplex PCR was amplified in a 25 µL reaction volume containing 1–10 ng DNA, 1 × PCR buffer, 8 mM MgCl<sub>2</sub>, 700 µM of each dNTP, 0.26–7.15 µM of each primer, and 2 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems) using the following cycling program: denaturation at 94°C for 5 min followed by 35 cycles at 95°C for 30 s, 60°C for 30 s, and 65°C for 30 s, followed by 7 min at 65°C. All PCR amplifications were performed in GeneAmp PCR system 9600 or 9700 (Applied Biosystems) or Eppendorf Mastercycler gradient instruments. PCR products were purified on a MinElute™ PCR purification spin column (Qiagen) and the DNA eluted in 20 µL of Milli-Q water. Alternatively, the PCR products were purified by adding 1 µL ExoSAP-IT™ (Amersham Biosciences), or by adding 0.75 U SAP (Amersham Biosciences) and 0.225 U Exol (USB Corporation) to 2.5 µL PCR product. The mix was incubated at 37°C for 15 min or 37°C for 1 h, respectively. The SAP and Exol enzymes were inactivated by incubation at 75°C for 15 min.





**Figure 1.** Phylogenetic tree defined with the binary Y-chromosomal polymorphisms analyzed. Marker names are indicated above the lines and lineage names are shown below the lines, but the length of each branch has no significance. Only the haplogroups with a colored square appear in the population samples analyzed.

## 2.5 SBE reaction and purification of the SBE products

The SBE reactions were performed in a total volume of 8  $\mu$ L with 0.3  $\mu$ L purified PCR product, 4  $\mu$ L SNaPshot™ reaction mix (Applied Biosystems), and 1  $\mu$ L SBE primer mix (Table 2). The SBE primer mix was stored in 160 mM ammonium sulfate (Sigma-Aldrich). The SBE reaction was performed using the following program: 30 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 30 s. After the SBE reaction, 1 U of SAP (Amersham Biosciences) was added and the mix was incubated at 37°C for 45 min. SAP was inactivated by incubation at 75°C for 15 min.

## 2.6 Detection and analysis of the SBE products

Two microliters SAP-purified SBE products were mixed with 20  $\mu$ L Hi-Di formamide and analyzed by CE using an AB 3100 or AB 3100 genetic analyzer with 36-cm capillary arrays (Applied Biosystems), POP-4 polymer (Applied Biosys-

tems), and 22 s injections at 2000 V. Analysis was done using GeneScan Analysis 3.7 with a minimum peak height of 100 relative fluorescence units (RFUs) for all dyes. Automated allele calls were made using Genotyper 3.7 macros.

## 2.7 Sequencing analysis

The PCR-amplified fragments corresponding to the locus M174 were purified on a MinElute PCR purification spin column (Qiagen) following the manufacturer. The DNA was eluted in 10  $\mu$ L of Milli-Q water. Both strands of the PCR product from four males and one female were sequenced by DNA Technology A/S (Arhus, Denmark).

## 2.8 Statistical analysis

Binary marker haplogroup frequencies were calculated in the 12 populations analyzed. Arlequin 2.0 software [23] was used to calculate genetic distances, as pairwise

**Table 1.** Y-chromosome markers and primer sequences for the multiplex PCR amplification

Locus	GenBank or DbSNPs accession	Mutation	PCR primers (5' → 3')		μM	Amplicon size, bp
			Forward primer	Reverse primer		
M2	rs3893	A > G	acggaaggagtctaaaattcagg	aaaatacagctcccccttatcct	1.63	128
M9 <sup>b)</sup>	rs3900	C > G	aggaccctgaaatacagaactg	aaatattcaacattcacaaggaa	6.50	186
M17 <sup>b)</sup>	rs3908	4G > 3G	cctggtcataacactggaatc	agctgaccacaaactgatgtaga	1.95	170
M33 <sup>b)</sup>	AC009977	A > C	tgaccgcataggctgagaca	ttgaagccccaagagagac	3.58	160
M35 <sup>b)</sup>	AC009977	G > C	aggcatggtcccttctat	tccatgcagacttcggagt	3.90	96
M38	AC009977	T > G	agatctgttgactagtacccta	gagctggcacatctgtcataata	1.30	177
M45 <sup>b)</sup>	Rs2032631	G > A	gagagaggatatacaaaattggcagt	tgacagtggcaccacaaaggct	0.39	138
Tat <sup>b)</sup>	AC002531	T > C	tatatggactctgagttagactgtga	ggtgccgtaaaagtgtgaataatc	0.52	115
M75	AC010889 <sup>c)</sup>	G > A	ctgtcaaaagccaaacagatttc	ttgaacagaggcatttgtgaat	5.20	120
M81 <sup>b)</sup>	rs2032640	C > T	catctttaacaaaa gaggtaaaatttgtcc	cattgtttacatggcctataat attcagt	2.60	179
M96 <sup>b)</sup>	AC010889 <sup>c)</sup>	G > C	tgccctctcacagagcactt	ccaccactttgttgcttg	7.15	143
M119	AC010889 <sup>c)</sup>	A > C	gatgtagaaaaatgtattgggtattcc	gaaatgccaggtaaattgactcac	0.78	188
M122 <sup>a)</sup>	AC010889 <sup>c)</sup>	T > C	gttgccctttgaaatgaataaatc	gaatttgacatttctaggtattcagg	0.65	175
M123 <sup>a)</sup>	AC010889 <sup>c)</sup>	G > A	gttgccctttgaaatgaataaatc	gaatttgacatttctaggtattcagg	0.65	175
M128	AC010889 <sup>c)</sup>	-2 bp	gcaaaagagaactgcctcta taaaatc	gcagaattatctacctcttcaaactg	2.60	139
M168	Rs2032595	C > T	gtggagtatgtgtggagggtga	cctctatcagaccatggtaattca	2.60	145
M170 <sup>b)</sup>	rs2032597	A > C	cagctcttattaagttatgtttca tattctgtg	gtcctcattttacagtgagacacaac	1.63	119
M173 <sup>b)</sup>	rs2032624	A > C	tttcttacaattcaaggcatttag	ctgaaaacaaaactggcttatca	0.98	81
M174	rs2032602	T > C	gcttctctgaataccttctggagt	tcttgcaaggaaaagtgtgc	0.46	94
M175 <sup>b)</sup>	rs2032678	-5 bp	gatttaactctctgaatcaggcacat	tttactgatacctttgttctgttcattc	0.65	79
M201	rs2032636	G > T	tcagatctaataatccagatc aactgag	gttcaaatcccatatccagca	0.65	93
M207	AC006376	A > G	ggggcaaatgtaagtcaagc	tcactcaacctctgttgaa	0.33	83
M216	AC010877	C > T	ccaatggaaattttataaccaca	tgacactgtagttatgtatacctgttg	1.63	160
M217	rs2032668	A > C	tctgttctgagatcattctaactg	ctgctgtggcttcatcaaaata	1.95	145
M304	AC007678	A > C	caaagaaaagcaggagattgttaa	aaacgtcttataccaaaatcaccag	0.65	90
P2	AC010137 <sup>c)</sup>	C > T	agccatctttccttaagagaggt	gcttgagtctcttaggttttga	0.26	104
P25 <sup>b)</sup>	rs150173	C > CA	tggaccatcacctgggtaaagt	ggcagataaggtgtcacatcacat	0.26	109
P31	AC002992	T > C	aacaggtagggtgatttggttaca	gacccgatatcgtgccatt	5.85	122
SRY <sub>10831</sub> <sup>b)</sup>	rs2534636 <sup>c)</sup>	A > G	tcacatagctcctgaaccatta	ccacataggtgaacctgaaaatg	1.63	150

a) Markers on the same DNA fragment.

b) PCR primers were published previously [24].

c) These accession numbers contain opposite sequences compared to the information reported in YCC 2002.

values of  $F_{ST}$ , and to perform an molecular analysis of the variance (AMOVA) analysis. A multidimensional scaling (MDS) analysis was performed on the genetic distances using SPSS version 11.5 software package.

### 3 Results

#### 3.1 Y-chromosome SNP selection

From a forensic point of view, and because the Y-chromosome cannot be used to individualize samples, our interest was focused on selecting SNPs to make an inference of the geographic origin of any male

sample. For this purpose, from more than 250 SNPs compiled in the phylogenetic tree published by the YCC, a package of 29 SNPs has been selected for the identification of major population haplogroups.

In the first step SNPs were selected according with their position in the parsimony tree and according to the haplogroup determined by them and their frequencies among populations. In the second step, some of the SNPs previously selected were changed due to problems in multiplex amplification or because new SNPs likely to be more useful were discovered during the interval between the two steps.



**Table 2.** SBE primer sequences for the detection of 29 Y-chromosome markers

Locus	Poly (dNTP) (5' → 3')	Sequence of 5'-tail (5' → 3')	Target-specific sequence (5' → 3')	a)	μM	Primer size, nt
M174	None	None	taccttctggagtgcc	F	0.06	17
M45	None	caa	ctcagaaggagcttttgc	R	0.14	22
M2	None	gacaa	ctttatcctccacagatctca	R	0.20	26
Tat	None	None	gctctgaaatattaaattaaacaac	R	0.08	26
M170	None	T	ctattttacttaaaaatcattgttc	F	0.05	30
M217	None	acaa	ttatgtattttccttctgaagagt	R	0.07	30
P25	None	tcgtgaaagtctgacaa	tgccctgaaacctgcctg	F	0.41	34
M201	None	aagtctgacaa	taataatccagatcaactgagg	F	0.02	34
M38	None	gtgaaagtctgac	cagtactattatggaaccaact	F	0.10	38
M304	None	gaaagtctgacaa	tgtcaattgaaagtaactgtga	F	0.14	38
M123	None	acgtcgtgaaagtctgacaa	atttctaggtattcaggcgtg	R	0.68	42
M207	None	cgtgaaagtctgacaa	caaatgtaagtaagcaagaattta	F	0.24	42
M35	None	gccacgtcgtgaaagtctgacaa	cccaattttccttgggacactg	F	0.54	46
M128	None	cgtgaaagtctgacaa	cctcttcaactgaaatgaaataactg	F	0.11	46
P31	None	gccacgtcgtgaaagtctgacaa	tacataataaggttttttgggtg	F	0.08	50
M216	None	ggtgccacgtcgtgaaagtctgacaa	gctagtattgtatacctgtgaat	R	0.08	50
M17	None	aactaggtgccacgtcgtgaaagtctgatct	ccaaaattcacttaaaaaacc	R	0.07	54
M175	None	ctaaactaggtgccacgtcgtgaaagtctgacaa	cacatgccttctcacttctc	F	0.24	54
M173	None	actaaactaggtgccacgtcgtgaaagtctgacaa	tacaattcaagggcatttagaac	F	0.07	58
M119	None	gactaaactaggtgccacgtcgtgaaagtctgacaa	tattccaattcagcatacaggc	R	0.20	58
M96	None	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	ggaaaacaggctctcataata	F	0.36	62
M122	DC + 5(dTdC) + dT	ctaaactaggtgccacgtcgtgaaagtctgacaa	tcagattttcccctgagagc	R	0.10	66
M75	4(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	aaagacaattatcaaacacatcc	F	0.02	70
SRY <sub>10831</sub>	5(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	ttgtatctgactttttcacacagt	F	0.06	74
M168	DC + 6 (dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	ctattgttttaattctcagctagc	R	0.03	78
M81	7(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	cttggtttgtgtgagtatactctatgac	R	0.08	82
M9	9(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	catgtctaaattaaagaaaaataagag	R	0.88	86
P2	16(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	gtgccctaggaggagaa	R	0.10	90
M33	DC + 12 (dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	cagttacaaaagtataatgtctgagat	R	0.27	94

a) Detection orientation has been probed relative to the YCC information reported in YCC 2002.

The final SNPs selected are represented in Fig. 1, where the haplogroups defined by them and their phylogenetic relationship are also shown.

### 3.2 Y-chromosome SNP genotyping

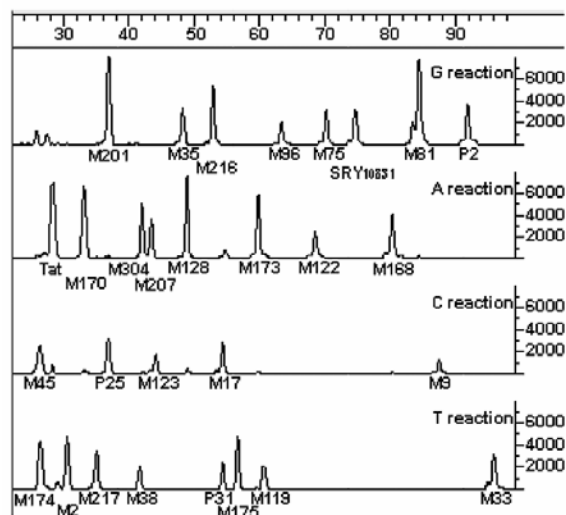
Once the SNPs were selected, a multiplex assay was developed using the guidelines for PCR multiplex design previously described in Sanchez *et al.* [9, 24], allowing the

simultaneous amplification of all 29 loci in a single reaction. The PCR multiplex design was performed by considering (i) the quality of flanking sequence of the candidate SNP, (ii) the predicted interactions between primers, and (iii) homologies to other amplicons in the multiplex and to other sequences in the human genome. It was also important to ensure that the PCR fragments were as short as possible and of uniform size, in order to obtain efficient amplification of all loci. The final result was a stable PCR multiplex reaction with 28 amplicons ranging in size from

79 to 188 bp (the two SNP loci M122 and M123 are located on the same amplicon). The SBE primers were also designed according to the guidelines described in Sanchez *et al.* [9, 24]. The target specific sequences of the SBE primers ranged from 17 to 29 nucleotides and the noncomplementary tails from 0 to 72 nucleotides. Tails were added in order to separate the SBE products efficiently in the subsequent electrophoresis. The SBE products in the final multiplex design ranged from 18 to 95 nucleotides in size. All PCR and SBE primers were first tested in singleplex reactions and later in multiplex reactions. Balanced peak heights in the final SBE reaction were obtained by altering the primer concentrations used in the PCR and SBE multiplexes. The final PCR and SBE primer concentrations are given in Tables 1 and 2, respectively.

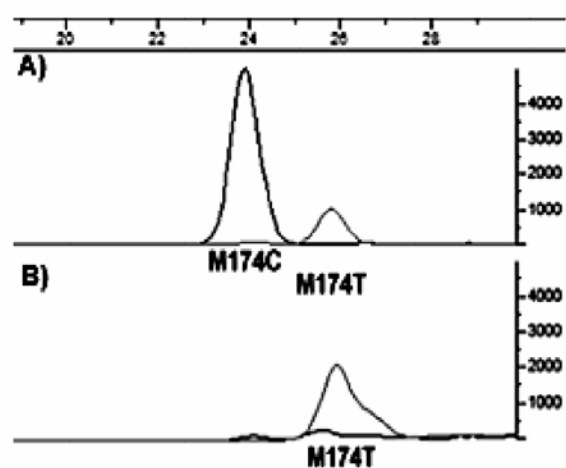
In Fig. 2, a representative electropherogram of the 29-plex is shown. For the short SBE primers a lack of correlation was observed between the observed and expected sizes. This may be due to the sequence composition of the SBE primers and the relative contribution of the respective fluorochrome of the labeled ddNTP. The difference between the observed and the expected size was not observed for the longer SBE primers.

The quality of the electropherograms is completely dependent on the dosage and quality on the initial DNA template. The optimum was found with 1–10 ng; however, good results and complete genotyping were achieved with only 0.25 ng. Fresh samples and old samples frozen for more than 3 years were analyzed and complete profiles were obtained.



**Figure 2.** Electropherogram with 29 Y-chromosome SNP profile from a male donor belonging to the Hg K\*(xN1, N3, OP).

One of the 29 selected SNPs, P25, is the result of a duplication event [25]. When the P25 marker has its ancestral allele, the duplication is not detected, but with the derived allele, the duplication is observed. Similarly, two M174 alleles were detected in 12 Japanese samples and one Chinese sample (example shown in Fig. 3) containing the derived allele of M174, whereas all other samples contained the ancestral allele alone. The sequence of the M174 locus was BLASTed against the database and we realized that this SNP is located in the ubiquitin-specific protease 9 gene (USP9Y), which has a homolog on the X-chromosome (USP9X). Looking at the great homology between the X- and Y-chromosome sequences, with only four-base difference, and since we had observed a peak comparable in size to the SBE products of the ancestral M174 allele in our female control samples, we suspected that the homologous region on the X-chromosome may be amplified in the PCR reaction. We amplified M174 in singleplex from four of the Japanese males and one Danish female, and sequenced the PCR products. A fragment homologous to the M174 locus was amplified from the female sample and the fragment contained the ancestral M174 allele. In addition, at least two fragments were amplified from the Japanese male samples. One fragment contained the derived allele and one fragment contained the ancestral allele. We made mixtures of male and female DNA with a dominant amount of female DNA (0.5, 5, 10, 20, and 46 ng) and a fixed amount of male DNA (0.25 ng). Two different male samples and a female sample were used. A complete 29-plex profile was obtained for all mixtures; however, the signal from the ancestral M174 allele increased proportionally to the amount of female DNA in the mixture (data not shown). Therefore, we conclude that the X-chromosome USP9X



**Figure 3.** GeneScan analysis of SBE reactions performed on M174 PCR products from a male donor belonging to the Y-chromosome haplogroup D (A) and a female donor (B).

locus was also amplified in the multiplex PCR. An alternative M174 primer design is being carried out to avoid possible preferential amplifications in case of mixtures.

### 3.3 Intercomparison exercise

An intercomparison exercise was performed with ten reference samples analyzed by all the participating labs. Haplogroups determined for eight male samples were concordant in all the labs. No results were obtained with the dog DNA, whereas a single extension product similar in size to the ancestral allele of the Y-chromosome M174 marker was detected for the female control sample.

### 3.4 Worldwide population study

A population study was performed to obtain allele and haplogroup frequencies for the major population groups. Among the 31 haplogroups defined by the 29 SNPs analyzed, 26 of them were found in a total of 1126 individuals distributed among four continents, America, Asia, Africa, and Europe. Haplogroup frequencies and diversities are represented in Table 3. The most frequent haplogroups are E3b\*(xE3b2, E3b3) in Somalia and E3a in Mozambique, which represent more than 70% of the samples of these populations, R1b in Galicia and Colombia, with more than 50% of the samples in each population, and O3 in China also representing more than 50% of the samples.

**Table 3.** Haplogroup frequencies

Haplogroup	Colombia (n = 60)	Argentina (n = 81)	Greenland (n = 90)	Thailand (n = 84)	China (n = 55)	Japan (n = 40)	Mozambique (n = 130)	Somalia (n = 105)	Germany (n = 150)	Denmark (n = 150)	Galicia (n = 130)	Turkey (n = 51)
A							0.008					
B							0.177	0.010				
CR*(xCE,G,IK)				0.036		0.025		0.029			0.008	
C*(xC2,C3)				0.036		0.025						
C2												
C3			0.011	0.036	0.055	0.075						
D					0.018	0.300						
E*(xE1,E2,E3)												
E1												
E2							0.069					
E3*(xE3a,E3b)									0.020		0.008	
E3a	0.083						0.715	0.010	0.020			
E3b*(xE3b2,E3b3)	0.067		0.011					0.743	0.067	0.020	0.046	0.020
E3b2	0.033						0.008	0.029			0.046	
E3b3	0.033	0.012							0.013	0.007	0.008	0.059
G	0.017	0.074						0.010	0.020	0.007	0.015	0.098
I	0.033	0.062	0.222						0.193	0.373	0.131	0.078
J	0.067	0.099	0.022	0.012				0.019	0.067	0.027	0.146	0.333
K*(xN1,N3,OP)	0.017	0.012		0.048	0.055	0.025		0.143	0.007		0.015	0.039
N1												
N3		0.012							0.013			
O*(xO1,O2,O3)				0.024								
O1				0.107	0.164	0.025						
O2				0.250	0.164	0.350						
O3				0.429	0.545	0.175						
P*(xR)	0.050	0.346	0.533						0.020	0.013		0.020
R*(xR1)	0.050											0.020
R1*(xR1a,R1b)										0.020	0.008	
R1a*(xR1a1)												
R1a1	0.017		0.056	0.024				0.010	0.133	0.173	0.008	0.137
R1b	0.533	0.383	0.144				0.023		0.427	0.360	0.562	0.196
HgD <sup>a)</sup>	0.702	0.723	0.649	0.744	0.654	0.768	0.455	0.429	0.757	0.704	0.646	0.826

a) Haplogroup diversity.



Figure 4 represents the observed haplogroup frequencies in the 12 populations in relation to the world map. At first glance the differences among populations are obvious; however, an ANOVA analysis was performed to look at this in greater depth, grouping the samples by continents. The results showed a high percentage of variation among continents, 33.7% ( $p < 0.001$ ) and a low percentage of variation among populations within continents, 6.8% ( $p < 0.001$ ). Most of the variance was attributable to differences within populations, 59.5% ( $p < 0.001$ ). These differences among groups of populations are also evident when a Principal component analysis was performed with the haplogroup frequencies (data not shown) and also when the first distances were multidimensionally scaled (Fig. 5). Both Asian and African populations are clearly grouped in a cluster in both analyses. As the two South American populations analyzed here have a high degree of historical coancestry with Europeans they tend to cluster together with European populations.

#### 4 Discussion

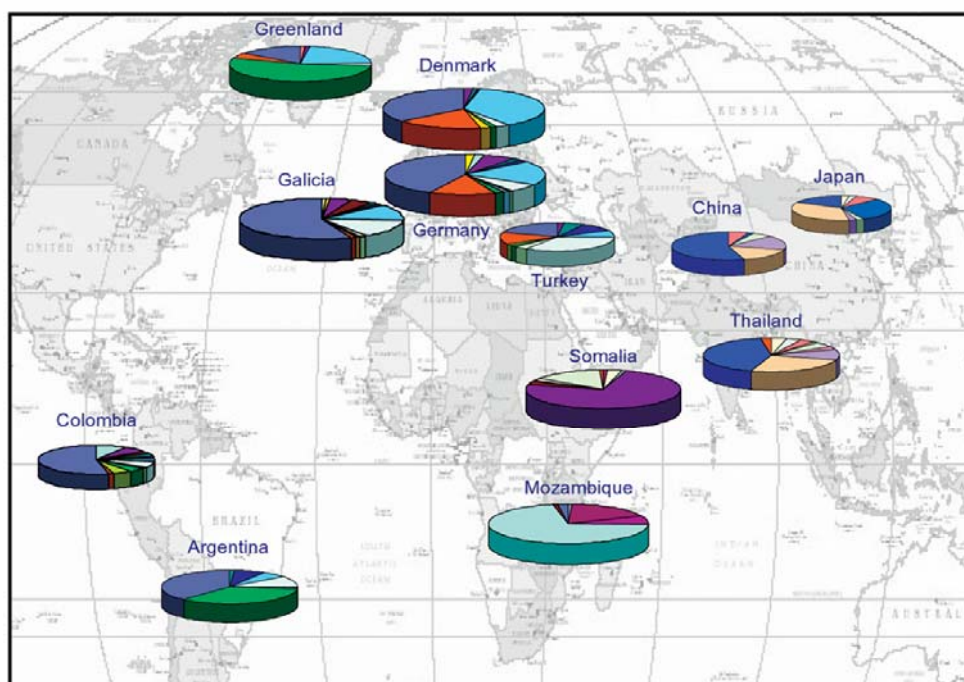
The objective of the SNPforID consortium (<http://www.snpforid.org/>) is the selection of SNPs suitable for forensic typing which should form the core set of markers

for future forensic high-throughput typing technologies. For the Y-chromosome, the interest was focused on making inferences on the geographic origin of an unknown sample. To evaluate this really interesting feature of the Y-SNPs, we selected the 29 SNPs used in this study as a first step of future designs more specific for each major haplogroup.

#### 4.1 Technological remarks

Once the SNPs had been selected, the second objective of the consortium was to find an appropriate technology for typing the SNPs in any interested laboratory. Nowadays, a large number of SNP genotyping methods are available; however, some of them require high investments and new developments, not always accessible in most of the routine laboratories. The selected method in this case was an SBE reaction, using the SNaPshot multiplex kit of Applied Biosystems, with the only requirement being access to an automatic sequencer.

From a technical point of view, we can conclude that the Major Y-chromosome haplogroup typing kit is a reliable technology, since the results for the control samples sent to all the labs were concordant; is a robust system, since



**Figure 4.** Haplogroup frequencies. Diagrams for each population are located over their position in the world map. Size of diagrams is proportional to the sample size, and haplogroup colors are the same as those in Fig. 1.

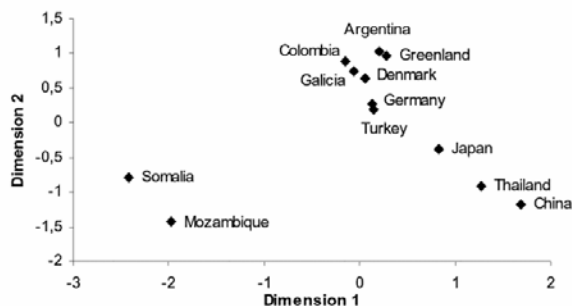


Figure 5. MDS of the  $F_{ST}$  distances. Stress value 0.07.

good results were obtained from diverse samples with different quantities and qualities; and as it was described in many studies [4, 25, 26] the SBE method used is an accurate technique, without the necessity of investment in new technologies. Our strategy of multiplex PCR and primer extension reaction can, however, easily be adapted to other platforms, such as MALDI-TOF-MS [27].

#### 4.2 Population validation

Among the 31 haplogroups defined using the 29 binary markers, 26 were found in the populations analyzed. Setting aside the South American samples (Colombians and Argentines) and the Greenlanders, because they are considered as mixed populations, six haplogroups appeared only in Asians, three haplogroups only in Africans, and seven haplogroups only in Europeans. We have named these haplogroups “private haplogroups.”

Looking more deeply at the South-east Asian data, their private haplogroups were C\*, D, O\*, and its immediate sub-clades, O1, O2, and O3. As haplogroup C3 is also typical of these Asian populations [28], we decided to consider it as a private haplogroup, in spite of it being detected in one individual from Greenland because of the admixed origin of this population. These Asian haplogroups include 88.2% of the Thai samples, 94.6% of the Chinese samples, and 95% of the Japanese samples; so if we consider these populations to be representative of the South-east Asians as a whole, we could say that approximately 93% of people with a south-east Asian patrilineage could be detected anywhere with our Major Y-chromosome haplogroup typing kit.

Among the African samples, only three private haplogroups were observed, A, B, and E3; in addition, haplogroup E3a has also been described as a haplogroup with a sub-saharan origin [29], so we have considered it as private for Africans. The three German individuals as well as the single individual from Colombia most probably

have a sub-saharan origin. The private haplogroups in this case include 96.9% of the samples from Mozambique, but only the 2% of the Somali samples; so it is clear that these haplogroups are not representative of all African populations, but the sub-saharan ones. Most of the samples from Somalia (74%) belong to haplogroup E3b, which was described to have a sub-saharan origin, expanding into North Africa and the Near East, and then into southern Europe [30, 31]. The Somali samples were included in a group of 201 Somali immigrants to Denmark in whom the frequency of E3b1 was 77.6% [31].

Europeans were the most abundant group of samples in this study, and presented seven private haplogroups, E3\*(xE3a, E3b, E3b3, I, N3, P\*(xR), R\*(xR1), and R1\*(R1a, R1b). However, haplogroup R1b described as an European haplogroup [1] also appeared in three samples from Mozambique, and can probably be explained by European admixture. If we set aside these three samples and consider R1b as a private haplogroup, they represent 68.6% of individuals in the German population sample, 77.3% in Denmark, 71.7% in Galicia, but only 37.3% of individuals in Turkey. The highest frequency of the seven haplogroups, 89.9%, was found in Greenland. This population despite the fact of being historically influenced by European migration [32], shows a high frequency of haplogroup P\*(xR) because the Native American haplogroup Q was not differentiated with the selected set of markers. On average, setting aside the Greenlanders, we could say that 62.1% of the people with a European origin could be detected anywhere with our Major Y-chromosome haplogroup typing kit. When other haplotypes are considered which do not have a private distribution, then a clear prediction is not possible. Nevertheless, one might consider making a relative prediction based on a likelihood ratio. In this case, two alternative hypotheses need to be formulated which can be calculated on the basis of the relative proportions among the different populations exhibiting a given haplogroup.

#### 4.3 Concluding remarks

The possibility to make inferences on the geographical origin of an unknown sample contributor has an enormous value in the investigative phase of a criminal procedure. However, it cannot be used as scientific evidence in the trial, where the individual DNA profile of the accused person is directly compared with the respective profiles of the samples collected at the scene of the crime.

Looking at the results obtained with the population samples selected, the Major Y-chromosome haplogroup typing kit seems to have utility in making inferences about the possible geographic origin of any sample of interest. With



this first panel of SNPs it can be possible to discriminate only between the major human groups present all around the world. To provide discrimination between more closely related populations, a new selection of markers will be required. Depending on the composition of the populations, different multiplexes will be necessary in combination to assign a sample to a specific haplogroup to give a more accurate prediction on the geographic origin of a sample. It is evident, however, that a Y haplotype-based prediction will always be limited to the elucidation of the patrilineage and thus could be quite misleading regarding the phenotype of a given individual. It cannot be excluded that a person exhibiting a European phenotype may carry a sub-saharan or Asian Y-chromosome, which has been passed on over many generations. To overcome this pitfall, a more complex approach is required which has to include autosomal SNPs with population-specific distributions [33], as well as the assignment to a mitochondrial DNA haplogroup [34] to include information as well about the maternal lineage. When all these three independent sources of genetic variation are considered equally by combining them into a single likelihood calculation, an accurate prediction of the geographic origin of an unknown sample contributor may be possible.

Furthermore, it has to be kept in mind that if we want to take advantage of the specific properties of Y-chromosome inheritance, we should always check that the markers we are analyzing do not have homologous loci on the X-chromosome, as it is the case for the M174 polymorphism analyzed here and in previous studies [21, 35].

*We want to thank Amparo Acosta from the University of Cauca (Colombia) for providing the Colombian samples. The technical assistance of Annemette Holbo Birk, Meli Rodríguez, and Raquel Calvo is highly appreciated. This work was partially supported by the European Community (GROWTH programme, SNPforID, contract G6RD-CT-2002-00844), the "Ministerio de Ciencia y Tecnología" (BMC2003-09822) and Ellen and Aage Andersen's Foundation.*

Received April 15, 2005

Revised July 24, 2005

Accepted August 2, 2005

## 5 References

- [1] Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., *et al.*, *Am. J. Hum. Genet.* 2000, 67, 1526–1543.
- [2] Arredi, B., Poloni, E. S., Paracchini, S., Zerjal, T., *et al.*, *Am. J. Hum. Genet.* 2004, 75, 338–345.
- [3] Semino, O., Magri, C., Benuzzi, G., Lin, A. A., *et al.*, *Am. J. Hum. Genet.* 2004, 74, 1023–1034.
- [4] Brion, M., Quintans, B., Zarrabeitia, M., Gonzalez-Neira, A., *et al.*, *Gene* 2004, 329, 17–25.
- [5] Gusmao, L., Brion, M., Gonzalez-Neira, A., Lareu, M., Carracedo, A., *Leg. Med.* 1999, 1, 55–60.
- [6] Jobling, M. A., Pandya, A., Tyler-Smith, C., *Int. J. Legal Med.* 1997, 110, 118–124.
- [7] Jobling, M. A., *Forensic Sci. Int.* 2001, 118, 158–162.
- [8] Pereira, L., Prata, M. J., Amorim, A., *Forensic Sci. Int.* 2002, 130, 147–155.
- [9] Sanchez, J., Borsting, C., Morling, N., *Methods Mol. Biol.* 2004, 297, 209–228.
- [10] Brion, M., *Methods Mol. Biol.* 2004, 297, 229–242.
- [11] Gill, P., *Int. J. Legal Med.* 2001, 114, 204–210.
- [12] Seielstad, M. T., Minch, E., Cavalli-Sforza, L. L., *Nat. Genet.* 1998, 20, 278–280.
- [13] Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., *et al.*, *Am. J. Hum. Genet.* 1999, 64(3), 817–831.
- [14] Y-Chromosome Consortium, *Genome Res.* 2002, 12, 339–348.
- [15] Jobling, M. A., Tyler-Smith, C., *Nat. Rev. Genet.* 2003, 4, 598–612.
- [16] Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., *et al.*, *Am. J. Hum. Genet.* 2004, 75, 128–137.
- [17] Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., *et al.*, *Hum. Genet.* 2004, 114, 127–148.
- [18] Chen, X., Sullivan, P. F., *Pharmacogenomics J.* 2003, 3, 77–96.
- [19] Sobrino, B., Carracedo, A., *Methods Mol. Biol.* 2004, 297, 107–126.
- [20] Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., *et al.*, *Nat. Genet.* 2000, 26, 358–361.
- [21] Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., *et al.*, *Ann. Hum. Genet.* 2001, 65, 43–62.
- [22] Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., *et al.*, *Science* 2000, 290, 1155–1159.
- [23] Schneider, S., Roessli, D., Excoffier, L., *Genetics and Biometry Laboratory*, University of Geneva, Switzerland 2000.
- [24] Sanchez, J. J., Borsting, C., Hallenberg, C., Buchard, A., *et al.*, *Forensic Sci. Int.* 2003, 137, 74–84.
- [25] Sanchez, J. J., Brion, M., Parson, W., Blanco-Verea, A. J., *et al.*, *Forensic Sci. Int.* 2004, 140, 241–250.
- [26] Quintans, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., *et al.*, *Forensic Sci. Int.* 2004, 140, 251–257.
- [27] Mengel-Jørgensen, J., Sanchez, J. J., Borsting, C., Kirpekar, F., Morling, N., *Anal. Chem.* 2004, 76, 6039–6945.
- [28] Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Tyler-Smith, C., *Am. J. Hum. Genet.* 2002, 71, 466–482.
- [29] Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., *et al.*, *Am. J. Hum. Genet.* 2004, 74, 532–544.
- [30] Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., *et al.*, *Am. J. Hum. Genet.* 2004, 74, 1014–1022.
- [31] Sanchez, J. J., Hallenberg, C., Borsting, C., Hernandez, A., Morling, N., *Eur. J. Hum. Genet.* 2005, 13, 856–866.
- [32] Bosch, E., Calafell, F., Rosser, Z. H., Norby, S., *et al.*, *Hum. Genet.* 2003, 112, 353–363.
- [33] Phillips, C., Lareu, M., Salas, A., Fondevila, M., *et al.*, *Progr. in Forensic Genetics 10, Int. Congress Series* 2004, 1261, 233–235.
- [34] Quintans, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V., Carracedo, A., *Forensic Sci. Int.* 2004, 140, 251–357.
- [35] Karafet, T., Xu, L., Du, R., Wang, W., *et al.*, *Am. J. Hum. Genet.* 2001, 69, 615–628.

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Forensic Science International: Genetics I (2007) 186–190

[www.elsevier.com/locate/fsig](http://www.elsevier.com/locate/fsig)

## Forensic validation of the SNPforID 52-plex assay

Esther Musgrave-Brown<sup>a,\*</sup>, David Ballard<sup>a</sup>, Kinga Balogh<sup>b</sup>, Klaus Bender<sup>b</sup>, Burkhard Berger<sup>c</sup>, Magdalena Bogus<sup>b</sup>, Claus Børsting<sup>d</sup>, María Brion<sup>e</sup>, Manuel Fondevila<sup>e</sup>, Cheryl Harrison<sup>a</sup>, Ceylan Oguzturun<sup>a</sup>, Walther Parson<sup>c</sup>, Chris Phillips<sup>e</sup>, Carsten Proff<sup>f</sup>, Eva Ramos-Luis<sup>e</sup>, Juan J. Sanchez<sup>d</sup>, Paula Sánchez Diz<sup>e</sup>, Bea Sobrino Rey<sup>e</sup>, Beate Stradmann-Bellinghausen<sup>b</sup>, Catherine Thacker<sup>a</sup>, Angel Carracedo<sup>e</sup>, Niels Morling<sup>d</sup>, Richard Scheithauer<sup>c</sup>, Peter M. Schneider<sup>f</sup>, Denise Syndercombe Court<sup>a</sup>

<sup>a</sup> Centre for Haematology, ICMS, Barts & the London, Queen Mary's School of Medicine & Dentistry, London, United Kingdom<sup>b</sup> Institute of Legal Medicine, Johannes Gutenberg University, Mainz, Germany<sup>c</sup> Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria<sup>d</sup> Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark<sup>e</sup> Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain<sup>f</sup> Institute of Legal Medicine, University of Cologne, Cologne, Germany

Received 24 January 2007; accepted 27 January 2007

### Abstract

The advantages of single nucleotide polymorphism (SNP) typing in forensic genetics are well known and include a wider choice of high-throughput typing platforms, lower mutation rates, and improved analysis of degraded samples. However, if SNPs are to become a realistic supplement to current short tandem repeat (STR) typing methods, they must be shown to successfully and reliably analyse the challenging samples commonly encountered in casework situations.

The European SNPforID consortium, supported by the EU GROWTH programme, has developed a multiplex of 52 SNPs for forensic analysis, with the amplification of all 52 loci in a single reaction followed by two single base extension (SBE) reactions which are detected with capillary electrophoresis.

In order to validate this assay, a variety of DNA extracts were chosen to represent problems such as low copy number and degradation that are commonly seen in forensic casework. A total of 40 extracts were used in the study, each of which was sent to two of the five participating laboratories for typing in duplicate or triplicate. Laboratories were instructed to carry out their analyses as if they were dealing with normal casework samples. Results were reported back to the coordinating laboratory and compared with those obtained from traditional STR typing of the same extracts using Powerplex<sup>®</sup> 16 (Promega). These results indicate that, although the ability to successfully type good quality, low copy number extracts is lower, the 52-plex SNP assay performed better than STR typing on degraded samples, and also on samples that were both degraded and of limited quantity, suggesting that SNP analysis can provide advantages over STR analysis in forensically relevant circumstances. However, there were also additional problems arising from contamination and primer quality issues and these are discussed.

© 2007 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** SNP; Degraded; Single base extension; Multiplex

### 1. Introduction

The advantages of analysing single nucleotide polymorphisms (SNPs) over short tandem repeats (STRs) in forensic casework have been the subject of much debate in recent years

[1–3]. These include the ability to design PCR primers to generate shorter amplicons than is possible with STR analysis, thereby making SNPs more likely to successfully type degraded DNA. Such degraded material is difficult to analyse with STRs [4,5] but is frequently encountered in forensic casework. Other advantages include a lower mutation rate than STRs, making them suitable for paternity analysis and complex relationship testing, and an absence of stutter peaks, which simplifies interpretation. SNP analysis also provides more flexibility in

\* Corresponding author. Tel.: +44 207 882 2272; fax: +44 207 882 2183.  
E-mail address: [e.musgrave-brown@qmul.ac.uk](mailto:e.musgrave-brown@qmul.ac.uk) (E. Musgrave-Brown).



the choice of typing platform than STR analysis, which may allow a greater degree of multiplexing than STRs and facilitate data-basing along with a greater degree of automation [6]. Because the majority of SNPs are biallelic, their discrimination power is much lower than that of STRs, and it is estimated that at least 50 SNPs are required to provide the same power as the routinely used STRs [7].

The SNPforID consortium ([www.snpforid.org](http://www.snpforid.org)) is supported by the EU GROWTH programme and was set up in 2002 with the primary aim of developing a SNP multiplex that could be used in forensic casework. The major product of this collaboration was the production of an assay to amplify 52 SNPs in a single PCR, with amplicons ranging in size from 59 to 115 bp. Following this reaction, several SNP detection methods could be used to analyse the products, including pyrosequencing, microarray hybridization, oligo ligation assays, and MALDI-TOF mass spectrometry. However, the forensic validation of the assay described here was exclusively performed using single base extension (SBE) reactions, as the capillary electrophoresis equipment necessary to detect these products is available in most forensic laboratories. The selection of the SNPs and development of the assay is described in more detail in Sanchez et al. [2].

This paper describes an inter-laboratory study to validate the 52-plex on a variety of challenging samples designed to mimic those encountered in forensic casework. The aim of the study was to test whether SNP typing is a realistic supplement to the current suite of assays available to the forensic investigator.

## 2. Methods

### 2.1. Study set-up

A total of 40 extracts were chosen for study, and each was aliquotted in triplicate. Two of the aliquots were sent out to two of the laboratories for SNP analysis, while the remaining one was tested in London using Powerplex<sup>®</sup> 16 (Promega). Some of the participating laboratories had been responsible for the design and testing of the multiplex, whereas others were less experienced.

Extracts were divided up so that no two laboratories received the same sample set. By sending out the aliquots in duplicate, we were able to investigate how the laboratories that had extensive experience of using the 52-plex on good quality samples were able to cope with poor quality samples and to compare their performance with laboratories that had less experience. In all cases, reference genotypes were determined by running good quality extracts at the London laboratory.

### 2.2. Challenging DNA extracts

Extracts were chosen for inclusion into this study in order to mimic material encountered in forensic casework. All were quantified in duplicate using Quantifiler<sup>™</sup> (Applied Biosystems). They can be split into the four categories below:

(1) Low copy number extracts: control DNA from cell lines 9947a and 9948 (Promega) was obtained to test assay

sensitivity. This was sent out to two of the laboratories at its original concentration ( $\sim 10$  ng/ $\mu$ L) and participants were asked to dilute this so that 1 ng, 500 pg, 200 pg, 100 pg, 50 pg, and 20 pg of DNA was added to the initial PCR.

(2) Degraded extracts: these were produced by sonication with the Ultrasonic Cell-Disruptor Sonifier [8]. Products were then separated on a 1% agarose gel and bands corresponding to fragment sizes of  $\sim 1000$ ,  $\sim 600$ ,  $\sim 400$  and  $\sim 300$  bp were extracted by electroelution. Following ethanol precipitation, the products were eluted in HPLC.

(3) Degraded and low copy number extracts: these were produced by extracting mouth swabs and blood-soaked cotton wool that had been left in warm, humid conditions for several months. The extracts were run on an agarose gel to check for degradation before being quantified. Concentrations varied from  $<0.01$  to 0.35 ng/ $\mu$ L.

### 2.3. SNP analysis

Each laboratory was asked to complete the study following their standard laboratory procedures for forensic samples. Information on the concentration of the extracts was provided, but extract volumes were limited so laboratories had to decide upon the best way to proceed with the analysis. Participants were permitted to quantify or concentrate their samples, but were requested to inform the coordinator if they had done so. It was requested that each sample be typed at least in duplicate and preferably in triplicate, and that participants would follow guidelines for amplifying and typing their extracts that were previously set out at a SNPforID workshop held in Copenhagen in 2004. All relevant information on reaction set up, primer concentrations, etc. can be found in Sanchez et al. [2]. Heterozygote peak ratio thresholds were specified in the Copenhagen guidelines and any heterozygote balances that fell beyond this range were reported with a '?' after the smaller peak (for example 'C,T?' or 'C?,T'). Where no peaks could be detected, the genotype was simply reported as 'N'.

### 2.4. Powerplex<sup>®</sup> 16 STR analysis

The third aliquot of each extract was amplified in triplicate using the Powerplex<sup>®</sup> 16 kit (Promega) and typed according to standard laboratory guidelines (minimum peak height of 50 rfu) using GeneScan<sup>®</sup> or GeneMapper<sup>™</sup> software (Applied Biosystems). Extracts were treated as real casework samples.

## 3. Results and discussion

### 3.1. Evaluation of results

A total of 390 genotypes were returned to the coordinator, with two or three repeats for each sample reported from each participant. One hundred and thirty-three consensus genotypes were then determined for each sample in each laboratory as outlined in Table 1. Each locus of the consensus genotype was then compared with the reference genotype and assigned two scores between 0 and 2, which correspond to the number of



Table 1  
Examples of how consensus genotypes were determined and scores assigned

Reference genotype	Observed genotype			Consensus genotype	Correct score	Drop-in score
	Replicate 1	Replicate 2	Replicate 3			
C	C	C	C	C	2	0
C	C	C	C/T	C	2	0
C	C/T	C/T	C	C/T	1	1
C	T	C/T	T	T	0	2
C/T	C	C/T	C/T	C/T	2	0
C/T	C	T	N	N	0	0

correct alleles and the number of incorrect (drop-in) alleles reported (Table 1). These scores were then used to calculate the proportion of total alleles correctly typed and the proportion of drop-in alleles. Powerplex<sup>®</sup> 16 results were analysed in the same way.

### 3.2. Assay sensitivity

In each of the two laboratories running these extracts, the 52-plex performed less effectively than Powerplex<sup>®</sup> 16 for every starting quantity of DNA, both in terms of the number of alleles correctly assigned (Fig. 1a) and for the proportion of alleles dropping in (Fig. 1b). In addition, one of the laboratories performed consistently worse than the other for both of these measures. These results conflict with those reported for previous consortium work [2] where full profiles were obtained

from only 500 pg of DNA, whereas in this study even 1 ng of DNA failed to produce optimal results.

The consistent differences between the participating laboratories and the lower than expected performance of the 52-plex when compared with previous work may suggest that training and primer stock quality are extremely important in this type of assay. In fact, both participating laboratories commented that they had experienced problems with the quality of their primer stocks.

The relatively poor performance of the 52-plex in typing limited material may be due to problems arising as a result of interactions during the PCR. It is likely that a PCR that amplifies such a high number of loci will encounter increased primer–primer interactions when compared with assays that amplify fewer loci, and these problems will be especially likely in circumstances where template is limited. Similar opportunities for such interactions also arise during the SBE reaction. This may also explain why there is an increased incidence of drop-in in the 52-plex (Fig. 1b).

A higher proportion of drop-in was observed in the 52-plex for all three sets of extracts and there are several possible explanations for this. The first is that drop-in peaks were only assigned if they were observed more than once. If drop-in occurs on a purely random basis, then in the case of the 52-plex, the chances of seeing the same allele twice for any given locus is higher because there are only two alleles in the population. However, for Powerplex<sup>®</sup> 16, the number of alleles at a particular locus is much higher, so the chances of observing the same drop-in allele more than once are much lower. Problems may also have arisen as a result of differences between facilities in the different laboratories, and this is discussed in the next section.

### 3.3. Degraded DNA extracts

Both Powerplex<sup>®</sup> 16 and the 52-plex produced good results when the average DNA fragment length was large, but the performance of the STR typing quickly decreased with average DNA fragment size (Fig. 2a), as was also demonstrated in a previous study [9]. This is to be expected given the difference in amplicon size between the two assays (the range of amplicon sizes for the SNPforID 52-plex is 59–115 bases, whereas that for Powerplex<sup>®</sup> 16 is 107–473 bases) and demonstrates one of the major potential advantages of SNP analysis in forensic casework. However, drop-in is more frequent in the 52-plex

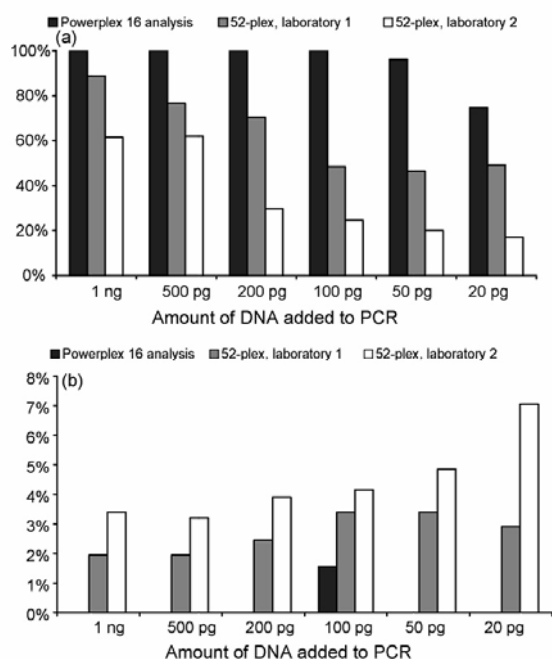


Fig. 1. Results of STR and SNP typing of varying starting amounts of DNA. (a) The proportion of total alleles typed correctly and (b) the proportion of total alleles dropping in.

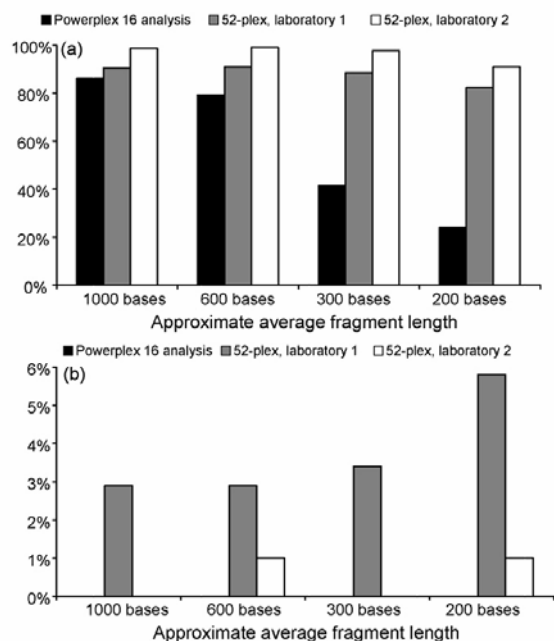


Fig. 2. Results of STR and SNP typing of artificially degraded DNA. (a) The proportion of total alleles typed correctly and (b) the proportion of total alleles dropping in.

than in Powerplex<sup>®</sup> 16, particularly in one of the laboratories (Fig. 2b). It may be that this is due to differences between the laboratory facilities. All of the Powerplex<sup>®</sup> 16 amplifications were set up in London in a specialist low copy number facility, but this was not necessarily the case for the laboratories carrying out the SNP typing. The apparent difference between laboratories 1 and 2 in the amount of drop-in (Fig. 2b) suggests that the problem is due to an issue at that particular laboratory, rather than an issue with the 52-plex itself.

3.4. Low copy number and degraded extracts

These extracts were variable in quality (concentration and degree of degradation) and, unsurprisingly, produced variable results. Again, some of the laboratories reported consistent problems with their primer pools and as a result have been removed from the graphs below (note that because of the way in which the samples were distributed, four laboratories initially typed these samples, so the two best performing participants are displayed).

These results indicate that SNP typing was more successful than STR typing in 21 of 26 cases (Fig. 3a), but that drop-in is still more of a problem with the 52-plex (Fig. 3b).

3.5. Variation between laboratories

In order to compare typing proficiency between laboratories, the average proportion of loci successfully typed was calculated for each sample, and the difference between this average and

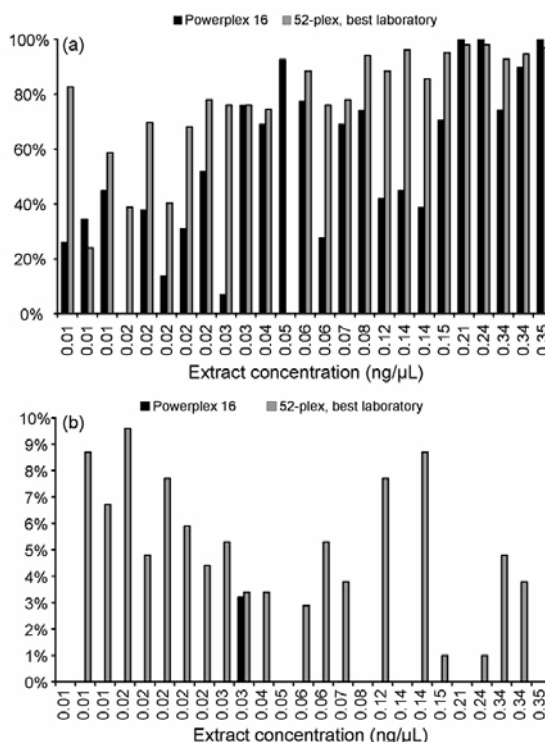


Fig. 3. Results of STR and SNP typing of degraded and low copy number DNA. (a) The proportion of total alleles typed correctly and (b) the proportion of total alleles dropping in.

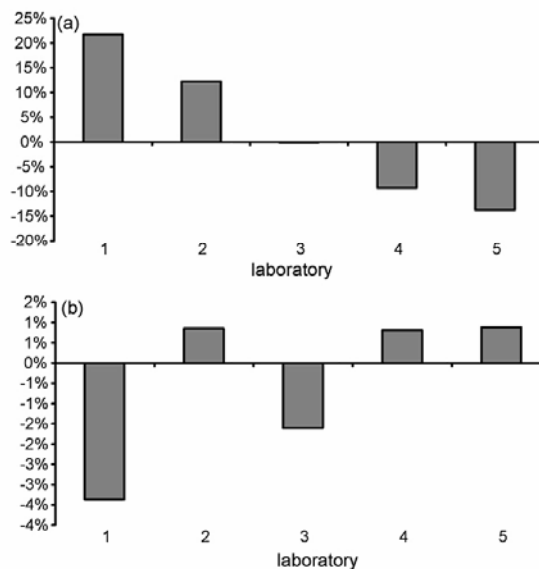


Fig. 4. Differences between the laboratories in the average number of loci successfully typed (a), and the average amount of drop-in (b).

the individual proportion for each laboratory was determined. The average difference across all samples was then calculated for each laboratory, and the results are displayed below. Laboratories that typed, on average, a higher proportion of loci successfully will therefore have a positive score, whereas those scoring lower on average will have a negative score. These results are summarized in Fig. 4a. The results for drop-in were analysed in the same way and are represented in Fig. 4b. A negative score indicates that the laboratory reported less drop-in than average, and a positive score indicates more drop-in than average.

When Fig. 4a and b are compared, it is clear that the laboratories that produced the least successful typing results also reported the greatest degree of drop-in, suggesting that the quality of results achieved with the 52-plex is dependent upon the laboratory in which it is run.

#### 4. Conclusions

The results obtained in this study demonstrate that, for the majority of the challenging samples tested, the 52-plex assay successfully typed a higher proportion of the loci than STR analysis. Although, results for good quality, low copy number extracts (typed in the sensitivity study) were poorer than from STR analysis, this sort of extract is less likely to be encountered in forensic casework than one which is degraded or both low copy number and degraded. In these cases, SNP typing performs well in comparison with STR typing, and it is therefore likely that such an assay could become a realistic supplement to such methods. However, the problems encountered with drop-in highlight the importance of appropriate

laboratory facilities, consistent reagent quality, as well as a more standardized approach for allele calling, when analysing such samples. Thus, it is more likely that forensic SNP typing will not be seen as a replacement, but rather an alternative when STRs fail to give results for challenging samples. We see a continuing interest in the forensic application of SNPs for human identification, and we expect that the autosomal SNP panel proposed by the consortium will serve as a starting point for the development of commercial typing kits.

#### References

- [1] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, *Nat. Rev. Genet.* 5 (10) (2004) 739–751.
- [2] J.J. Sanchez, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis* 27 (9) (2006) 1713–1724.
- [3] L.A. Dixon, et al., Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (EDNAP) exercise, *Forensic Sci. Int.* 164 (1) (2006) 33–44.
- [4] M. Graw, H.J. Weisser, S. Lutz, DNA typing of human remains found in damp environments, *Forensic Sci. Int.* 113 (1–3) (2000) 91–95.
- [5] E.M. Golenberg, A. Bickel, P. Weihs, Effect of highly fragmented DNA on PCR, *Nucl. Acids Res.* 24 (24) (1996) 5026–5033.
- [6] B. Sobrino, M. Brion, A. Carracedo, SNPs in forensic genetics: a review on SNP typing methodologies, *Forensic Sci. Int.* 154 (2–3) (2005) 181–194.
- [7] P. Gill, An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, *Int. J. Legal Med.* 114 (4–5) (2001) 204–210.
- [8] K. Bender, M.J. Farfan, P.M. Schneider, Preparation of degraded human DNA under controlled conditions, *Forensic Sci. Int.* 139 (2/3) (2004) 135–140.
- [9] P.M. Schneider, et al., STR analysis of artificially degraded DNA—results of a collaborative European exercise, *Forensic Sci. Int.* 139 (2–3) (2004) 123–134.

---

## Eigene Publikationen

**Balogh MK**, Burger J, Bender K, Schneider PM, Alt KW (2003) Fingerprint from Fingerprints. Progress in Forensic Genetics 9, Int. Congress Series: 953-957.

**Balogh MK**, Burger J, Bender K, Schneider PM, Alt KW (2003) STR genotyping and mtDNA sequencing of latent fingerprint on Paper. Forensic Science International 137(2-3): 188-195.

Schneider PM, **Balogh MK**, Naveran N, Bogus M, Bender K, Lareu M, Carracedo A (2004) Whole genome amplification – the solution for a common problem in forensic casework? Progress in Forensic Genetics 10, Int. Congress Series 1261: 24-26.

Brión M, Sanchez JJ, **Balogh K**, Thacker C, Blanco-Verea A, Børsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2005) Introduction of an SNP-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages. Electrophoresis 26: 4411-4420.

Brión M, Sanchez JJ, **Balogh K**, Thacker C, Blanco-Verea A, Børsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006) Analysis of 29 Y-chromosome SNPs in a single multiplex useful to predict the geographic origin of male lineage. Progress in Forensic Genetics 11, Int. Congress Series 1288: 13-15.

Sanchez JJ, Phillips C, Børsting C, **Balogh K**, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27: 1713-1724.

Thacker CR, **Balogh K**, Børsting C, Sánchez Diz P, Carracedo A, Morling N, Schneider PM, Syndercombe-Court D (2006) The effect of whole genome amplification on samples originating from more than one donor. Progress in Forensic Genetics 11, Int. Congress Series 1288: 722-724.

**Balogh K**, Bender K, Schneider PM and the SNPforID Consortium (2006) Application of Nanogen Microarray Technology for Forensic SNP Analysis. Progress in Forensic Genetics 11, Int. Congress Series 1288: 43-45.

**Balogh K**, Børsting C, Sánchez Diz P, Thacker C, Syndercombe-Court D, Carracedo A, Morling N, Schneider PM, SNPforID Consortium (2006) Application of Whole Genome Amplification for Forensic Analysis. Progress in Forensic Genetics 11, Int. Congress Series 1288: 724-727.

Musgrave-Brown E, Ballard D, **Balogh K**, Børsting C, Sánchez Diz P, Carracedo A, Morling N, Schneider PM, Syndercombe-Court D (2007) Forensic validation of the SNPforID 52-plex assay. Forensic Science International: Genetics 1: 186-190.

Sanchez JJ, Phillips C, Børsting C, **Balogh K**, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2007) Forensic typing of autosomal SNPs with a 29 SNP-multiplex – results of a collaborative EDNAP exercise Forensic Science International: Genetics 2(3): 176-183.