**Diagnostische Analyse des hochkomplexen OPN1LW/OPN1MW-Genclusters mittels Langstrecken-Sequenzierung und MLPA**

**Abstract**

Pathogene Varianten im OPN1LW/OPN1MW-Gencluster sind ursächlich für eine Reihe von leichten bis schweren Sehstörungen mit Farbsehstörungen. Die weit verbreitete Next-Generation-Sequenzierung (NGS) mit kurzen Reads ist für die Analyse des OPN1LW/OPN1MW-Genclusters ungeeignet, und viele Patienten mit pathogenen Varianten bleiben unterdiagnostiziert. Ein diagnostischer Gentest für das OPN1LW/OPN1MW-Gencluster wurde entwickelt, der aus einer Kopienzahl-Analyse mittels multiplex Ligation-dependent Probe Amplification (MLPA) und einer Sequenzanalyse mittels langstreckiger zirkularer Konsensus-Sequenzierung besteht. Die Leistung wurde an 50 klinischen Proben bestimmt, die zur genetischen Bestätigung der klinischen Diagnose (n = 43) oder zur Trägerstatusanalyse (n = 7) eingesandt wurden. Es wurde eine breite Palette pathogener Haplotypen nachgewiesen, darunter Deletionen, Hybridgene, Einzelvarianten und Kombinationen von Varianten. Der entwickelte Gentest für das OPN1LW/OPN1MW-Gencluster ist ein diagnostischer Test, der sowohl strukturelle als auch Nukleotidvarianten mit einer einfachen Analyse erkennen kann und die diagnostische Versorgung von Patienten mit Sehbehinderung verbessert.

**Introduction**

Die Farb- und Lichtempfindlichkeit des menschlichen Sehens wird durch die Zapfen in der Netzhaut reguliert. Menschen haben drei Arten von Zapfen-Opsinen mit unterschiedlichen relativen spektralen Empfindlichkeiten; die kurz-, mittel- und langwellenempfindlichen Opsine haben eine optimale spektrale Empfindlichkeit für das blaue, grüne und rote Lichtspektrum. Jeder Zapfen in der Netzhaut exprimiert nur eine Art von Zapfen-Opsin und ist daher nur für eines der drei Spektren empfindlich.

Die mittel- und langwellenempfindlichen Opsine werden durch OPN1MW [MIM: 300821] bzw. OPN1LW [MIM: 300822] kodiert. OPN1LW und OPN1MW liegen in enger Nachbarschaft auf dem X-Chromosom und teilen mehr als 98 % Sequenzhomologie. Das sogenannte OPN1LW/OPN1MW-Gencluster besteht in der Regel aus einem OPN1LW- und bis zu fünf OPN1MW-Genen, die kopfseitig aneinander gereiht sind (Abb. 1a). Etwa 3,5 Kilobasen (kb) stromaufwärts des OPN1LW befindet sich die Locus-Kontrollregion (LCR), die die Expression entweder des ersten oder zweiten Opsin-Gens stromabwärts der LCR sicherstellt.

Varianten im OPN1LW/OPN1MW-Gencluster verursachen eine Reihe von milden bis schweren X-chromosomalen Farbsehstörungen. Rot- oder Grünsehschwäche (Protanopie [MIM: 303900] bzw. Deuteranopie [MIM: 303800]) mit normaler Sehschärfe und normalem Zapfen-Elektroretinogramm (ERG) betrifft etwa 1 von 12 Männern und 1 von 200 Frauen. Die Bornholm-Augenkrankheit (BED [MIM: 300843]), eine schwerwiegendere Sehbehinderung, ist mit starker Kurzsichtigkeit, leicht bis mäßig reduzierter Sehschärfe und reduziertem Rot- und Grünzapfen-ERG verbunden, zusätzlich zu Protanopie oder Deuteranopie. Patienten mit blauer Zapfenmonochromasie (BCM [MIM: 303700]) nehmen nur das blaue Lichtspektrum wahr, kombiniert mit angeborenem Nystagmus, Photophobie, starker Kurzsichtigkeit, fehlendem Rot- und Grünzapfen-ERG und stark reduzierter Sehschärfe.

Es wurden verschiedene Arten pathogener Varianten im OPN1LW/OPN1MW-Gencluster nachgewiesen. Aufgrund der hohen Homologie und der engen Nachbarschaft von OPN1LW und OPN1MW treten häufig Umlagerungen auf. Nicht-homologe Rekombination kann zur Bildung von Hybrid-Opsin-Genen führen, die ein funktionelles Opsin-Protein mit optimaler spektraler Empfindlichkeit für das rote oder grüne Lichtspektrum kodieren. Eine weitere häufige strukturelle Variante ist eine Deletion der LCR, die BCM verursacht. Pathogene Einzel-Nukleotid-Varianten wurden ebenfalls beschrieben, z.B. die c.607C > T p.(Cys203Arg)-Variante, die häufig bei Patienten mit BCM nachgewiesen wird. Darüber hinaus umfasst Exon 3 in beiden, OPN1LW und OPN1MW, eine Region mit acht Nukleotidvarianten, die sieben Aminosäuren kodieren (c.453G > A p.(Arg151Arg), c.457C > A p.(Leu153Met), c.465G > A p.(Val155Val), c.[511G > A;513G > T] p.(Val171Ile), c.521C > T p.(Ala174Val), c.532 A > G p.(Ile178Val), c.538G > T p.(Ser180Ala)), die in der Bevölkerung häufig auftreten. Spezifische seltene Kombinationen dieser Varianten sind jedoch pathogen, da sie eine fehlerhafte Spleißung verursachen. Beispielsweise ist die LIAVA-Kombination in OPN1LW und die MVVVA-Kombination in OPN1MW ursächlich für BED.

Derzeit sind nur wenige diagnostische Gentests verfügbar, um das OPN1LW/OPN1MW-Gencluster zu analysieren. Die meisten Tests zielen nicht ausschließlich auf das Cluster ab, sondern die OPN1LW/OPN1MW-Gene werden als Teil umfassenderer Gen-Panels getestet und durch kurze Reads der NGS analysiert. Kurze Reads der Sequenzierung generieren Reads von 100-400 Basenpaaren (bp) Länge, die zu kurz sind, um zwischen den hoch homologen OPN1LW- und OPN1MW-Genen zu unterscheiden. Darüber hinaus kombinieren keine der verfügbaren Tests Kopienzahl- und Sequenzanalyse, was erforderlich ist, um das vollständige Spektrum kausaler Varianten zu erkennen. Aufgrund der Unverfügbarkeit vollständiger genetischer Tests des OPN1LW/OPN1MW-Opsin-Genclusters bleiben viele Patienten mit pathogenen Varianten unterdiagnostiziert. Wahrscheinlich trägt die genetische Unterdiagnose zur mangelnden Sensibilisierung der Kliniker bei, retinalen Erkrankungen, die durch dysfunktionale Opsin-Gene verursacht werden, zu erkennen.

Das Ziel dieser Studie ist es, die diagnostische Versorgung von Patienten mit Sehbehinderungen und Farbsehstörungen zu erleichtern und zu verbessern, indem ein genetischer Assay zur Analyse des OPN1LW/OPN1MW-Genclusters entwickelt wird, der die Kopienzahl-Analyse mittels MLPA und die Sequenzanalyse mittels langstreckiger zirkularer Konsensus-Sequenzierung kombiniert.

**Ergebnisse**

**Leistung des Opsin-MLPA an Kontrollproben**

MLPA ist eine Technik zur Quantifizierung der Kopienzahl vorgewählter Ziele, in diesem Fall genomischer Ziele innerhalb des Opsin-Locus. Jedes Ziel erzeugt ein fluoreszierendes Signal, und der Vergleich dieser Signale, durch Berechnung eines Verhältnisses zwischen einer Probe und einer einzelnen oder einer Gruppe von Referenzproben mit bekannten Kopienzahlen, wird verwendet, um die Kopienzahl der Probe zu bestimmen. Da es noch keine öffentlich zugängliche Referenzprobe mit beschriebenen OPN1LW- und OPN1MW-Kopienzahlen gibt, wurde zunächst die neu entwickelte X080 Opsin MLPA an zehn männlichen und zehn weiblichen Proben durchgeführt, deren Farbsehstatus unbekannt war. Bei allen Sonden wurden fluoreszierende Signale in der erwarteten Fragmentlänge detektiert, und die Verhältnisse der Kontrollsonden zwischen allen Proben lagen wie erwartet zwischen 0,8 und 1,2. Anschließend wurde die Kopienzahl der LCR- und OPN1LW/OPN1MW-Gene in allen Proben bestimmt, indem vier männliche Proben mit vergleichbar niedrigen fluoreszierenden Signalen für alle Ziele im Opsin-Locus als Referenz genommen wurden, unter der Annahme, dass diese eine LCR-, eine OPN1LW- und eine OPN1MW-Genkopie hatten. Eine langstreckige Sequenzanalyse des OPN1LW/OPN1MW-Genclusters, die an den vier männlichen Kontrollproben durchgeführt wurde, bestätigte, dass die Ziele der MLPA-Sonden korrekt identifiziert wurden; zudem waren in diesen vier männlichen Proben keine heterozygoten Varianten vorhanden, was ebenfalls darauf hinweist, dass diese Männer jeweils eine OPN1LW- und eine OPN1MW-Genkopie haben. Anschließend wurden die Kopienzahlen in den restlichen Proben unter Verwendung dieser vier männlichen Proben als Referenz bestimmt. In den verbleibenden sechs männlichen Proben wurden eine LCR- und drei oder vier OPN1LW/OPN1MW-Genkopien detektiert, und in den zehn weiblichen Proben wurden zwei LCR- und vier bis acht OPN1LW/OPN1MW-Genkopien detektiert.

#### Kopienzahl-Analyse in klinischen Proben

Die Opsin-MLPA wurde an 50 klinischen Proben durchgeführt (43 männliche und 7 weibliche Proben). In zwei der 43 männlichen klinischen Proben waren beide Sondensignale der LCR nicht nachweisbar, was auf eine Deletion der LCR hinweist. Von den verbleibenden 41 männlichen Proben hatten zehn eine einzige OPN1LW/OPN1MW-Genkopie, 16 hatten zwei und 15 hatten drei oder mehr OPN1LW/OPN1MW-Genkopien. Die sieben weiblichen Proben hatten zwei LCR und vier bis sechs OPN1LW/OPN1MW-Genkopien.

#### Spezifität der langstreckigen Sequenzierungs-Amplicons

Da auch kausale Nukleotidvarianten im Opsin-Locus beschrieben wurden, wurden vier LR-Amplicons entworfen, um das OPN1LW/OPN1MW-Gencluster abzudecken (Abb. 1a). Die Spezifität dieser Amplicons wurde an vier männlichen Kontrollproben mit je einem OPN1LW und einem OPN1MW bestimmt. Alle vier LR-PCR-Reaktionen ergaben Amplicons der erwarteten Länge (Abb. 2). Zudem wurden die Sequenzierungs-Reads der Amplicons für das erste und das zweite (und nachfolgende) Opsin-Gen nur auf den Opsin-Locus ausgerichtet, wenn sie auf das gesamte Genom abgebildet wurden (zusätzliche Abb. 3). Obwohl das Amplicon für das letzte Opsin-Gen ein einzelnes Band ergab, wurden neben der Ausrichtung der meisten Reads auf den Opsin-Locus auch einige Reads auf Chromosom 1 ausgerichtet. Das Amplicon für die LCR zeigte ein zusätzliches unspezifisches Produkt von ~6 kb (Abb. 2), dessen Reads auf eine intronische Region von IGF1R auf Chromosom 15 ausgerichtet wurden.

#### Variantenanalyse in klinischen Proben

Eine Sequenzierungsstrategie zur Bestimmung, welche Sequenzierungsreaktionen pro klinischer Probe erforderlich sind, wird in Abb. 3 im Detail erklärt, und die 50 klinischen Proben wurden entsprechend sequenziert. Details zu den Mapping-Statistiken der für jede klinische Probe sequenzierten Amplicons sind in Tabelle 3 im Anhang aufgeführt. In zwei Proben (USN04466 und USN15024) wurde das LCR-Amplicon durchgeführt, da die MLPA das Fehlen der LCR ergab. Es wurde kein PCR-Produkt amplifiziert, wahrscheinlich umfassen die Deletionen in diesen Proben auch die genomische Sequenz, an die der Vorwärts- und/oder Rückwärtsprimer des LCR-Amplicons bindet. In 48 Proben wurde das erste Gen-Amplicon sequenziert, und in allen 48 Proben war das erste Gen im Cluster ein (Hybrid-)OPN1LW (Tabelle 1 und 2 im Anhang). In 39 Proben wurden Varianten im (Hybrid-)OPN1LW nachgewiesen, darunter eine Variante unbekannter Signifikanz (VUS) im Promotorbereich (c.-33\_-13delinsATCAC), Protein-trunkierende Varianten (c.802\_804delinsG p.(Arg268fs) und c.852C > A p.(Tyr284\*)), die pathogene c.607T > C p.(Cys203Arg) Variante, missense VUS (c.347C > G p.(Ser116Cys) und c. 1013G > T p.(Gly338Val)), pathogene Kombinationen von Varianten LIAVA, LIVVA, LVAVA, MIAVA und MVAVA sowie die MVVVA-Kombination unbekannter Signifikanz (Tabelle 1 und 2 im

Anhang).

Die Sequenzierung des zweiten (und nachfolgenden) Gen-Amplicons wurde in 38 klinischen Proben durchgeführt. Da ein bis vier Gene gleichzeitig sequenziert wurden, reichten die detektierten Variantenprozentsätze von ~25% bis 100%. Anhand des Variantenprozentsatzes wurde abgeleitet, ob eine, mehrere oder alle Genkopien die Variante(n) aufwiesen. In allen 38 Proben war das zweite (und nachfolgende) Gen im Cluster (ein) (Hybrid-)OPN1MW (Tabelle 1 und 2 im Anhang). In 29 Proben wurden Nukleotidvarianten nachgewiesen, darunter die pathogene c.607T > C p.(Cys203Arg) Variante, die missense VUS c.38G > A p.(Arg13His) und c.659T > C p.(Met220Thr), pathogene Kombinationen von Varianten LVAVA, MIAVA und MVAVA sowie die MVVVA-Kombination unbekannter Signifikanz (Tabelle 1 und 2 im Anhang).

Die Sequenzierung des letzten Gen-Amplicons wurde in fünf männlichen Proben mit drei Opsin-Genen durchgeführt, in denen die letzten beiden Gene nicht identisch waren. Da der Rückwärtsprimer dieses Amplicons außerhalb der duplizierten Region liegt und in allen Proben nur eines der beiden zuvor detektierten (Hybrid-)OPN1MW-Gene sequenziert wurde (Tabelle 1 im Anhang), nehmen wir an, dass das Amplicon spezifisch für das letzte Opsin-Gen im Cluster ist. In zwei Proben war das letzte Gen die OPN1MW-Genkopie ohne die pathogene Variante, in einer Probe ein Hybrid-OPN1MW und in den letzten beiden Proben ein OPN1MW mit der MVAVA-Kombination (Tabelle 1 im Anhang).

In zwei Proben (USN01345 und USN20083), in denen keine kausalen Varianten nachgewiesen wurden, und in zwei Proben (USN15983 und USN03197), bei denen unklar war, ob kausale Varianten detektiert wurden, wurde die Sequenzierung der LCR durchgeführt, um eine kleine Deletion der LCR auszuschließen, die durch MLPA übersehen worden sein könnte. In diesen Proben wurden keine Deletionen oder seltene Varianten in der LCR nachgewiesen.

#### Bestimmung des OPN1LW/OPN1MW-Genclusters in klinischen Proben

Die Ergebnisse der MLPA und der langstreckigen Sequenzierung wurden kombiniert, um die genetische Zusammensetzung des OPN1LW/OPN1MW-Genclusters in den 50 klinischen Proben zu bestimmen. In 43 Proben bestand die klinische Fragestellung darin, die klinische Diagnose von Protanopie, BED, BCM oder Zapfendystrophie genetisch zu bestätigen. In 39 der 43 Patientenproben wurden kausale Varianten nachgewiesen (Abb. 4a), in einem Fall blieb die Kausalität aufgrund des Nachweises einer VUS sowohl im OPN1LW als auch im OPN1MW unklar (Abb. 4b), in einem Fall wurde eine genetische Ursache für Protanopie nachgewiesen, obwohl der Patient mit BED überwiesen wurde (Abb. 4b) und in zwei Proben wurde ein gutartiges OPN1LW/OPN1MW-Allel identifiziert (Abb. 4c).

In sieben Proben bestand die klinische Fragestellung darin, den Trägerstatus zu bestimmen. Interessanterweise wurde in drei Proben, die auf den Trägerstatus für BCM getestet wurden, weil sie jeweils einen Sohn mit BCM ohne familiäre Vorbelastung hatten, ein Genrekombinationsereignis im Keimbahn- oder frühen Embryonalstadium nachgewiesen, das zu einem Hybrid-Allel mit der pathogenen c.607T > C p.(Cys203Arg)-Variante in ihrem betroffenen Sohn führte (Abb. 5a, Tabelle 2 im Anhang). Außerdem hatte in USN19477 ein noch nicht definiertes genomisches Ereignis im Keimbahn- oder frühen Embryonalstadium stattgefunden, da in dieser Probe die LIAVA-Sequenz, die im einzigen Opsin-Gen in ihrem Sohn vorhanden war, in keiner ihrer Opsin-Genkopien nachgewiesen wurde (Tabelle 2 im Anhang). USN18152 mit zwei betroffenen Kindern war Trägerin von BED, während ihre Schwester ein gutartiges Allel von ihrer Mutter geerbt hatte (Abb. 5b, Tabelle 2 im Anhang). Die verbleibende Probe (USN09034) hatte wahrscheinlich ein gutartiges Allel von ihrer Mutter geerbt, die eine obligate Trägerin war (Tabelle 2 im Anhang).

### Diskussion

Der entwickelte Assay für das hochkomplexe genetische OPN1LW/OPN1MW-Cluster ist ein diagnostischer Gentest, der eine Vielzahl pathogener Varianten im Cluster durch die Kombination von Kopienzahl-Analyse und langstreckiger Sequenzierung nachweisen kann. Die Analyse ist unkompliziert, da MLPA eine schnelle und effiziente Methode zur Kopienzahl-Analyse darstellt und der langstreckige Sequenzierungsansatz es ermöglicht, die genetische Zusammensetzung des gesamten OPN1LW- oder OPN1MW-Gens in einzelnen Sequenzierungs-Reads zu bestimmen. Der entwickelte Assay verbessert die diagnostische Versorgung von Patienten mit Kegelfunktionsstörungen und kann Ärzten bei der Identifizierung zukünftiger Patienten helfen, bei denen eine genetische Analyse des OPN1LW/OPN1MW-Genclusters erforderlich ist.

Die Kopienzahl-Analyse des OPN1LW/OPN1MW-Genclusters wurde mittels eines neu entwickelten MLPA-Tests durchgeführt. Kontrollproben wiesen alle Kopienzahlen im Bereich von Personen mit normalem Farbsehen auf, während in klinischen Proben abweichende OPN1LW/OPN1MW-Genkopienzahlen detektiert wurden. Anschließend wurden langstreckige Sequenzierungsreaktionen für die Variantenanalyse des OPN1LW/OPN1MW-Clusters entwickelt. Die Zuordnung der Sequenzierungs-Reads wurde auf eine künstliche genomische Referenzsequenz durchgeführt, die aus einem OPN1LW und einem OPN1MW bestand, da die genomischen Referenzsequenzen von GRCh37 und GRCh38 für die Zuordnung der Sequenzierungs-Reads nicht geeignet waren. Sowohl GRCh37 als auch GRCh38 enthalten mehrere OPN1MW-Genkopien (GRCh37: 2 OPN1MW-Genkopien, GRCh38: 3 OPN1MW-Genkopien), und Sequenzierungs-Reads des zweiten (und nachfolgenden) Gen-Amplicons werden zufällig einer der OPN1MW-Genkopien zugewiesen, was zu falschen Variantenprozentzahlen führt.

Da nur die ersten beiden Opsin-Gene stromabwärts der LCR in Protein übersetzt werden, ist die genetische Zusammensetzung nur der ersten beiden Gene im Cluster klinisch relevant. Selbst mit dem verwendeten langstreckigen Ansatz, der über 16 kb umfassende Amplicons verwendet, konnte kein spezifisches Amplicon für das zweite Gen im Cluster entwickelt werden. Die meisten klinischen Proben haben jedoch nur zwei Opsin-Gene in ihrem Cluster, und in diesen Proben wird das zweite (und nachfolgende) Gen-Amplicon nur Ergebnisse für das zweite Gen liefern, da keine nachfolgenden Gene vorhanden sind (Abb. 1b). Ein zusätzliches LR-PCR wurde entwickelt, um die genetische Zusammensetzung des zweiten Gens in Proben mit drei Opsin-Genen zu bestimmen (Abb. 1c). Obwohl der entwickelte diagnostische Assay die genaue Zusammensetzung des Clusters nicht bestimmen kann, wenn vier oder mehr Opsin-Genkopien vorhanden sind und die genetische Zusammensetzung der zweiten und nachfolgenden Genkopien nicht identisch ist (Abb. 1d), konnte der Assay die Zusammensetzung der ersten beiden Gene im OPN1LW/OPN1MW-Gencluster in allen 43 getesteten männlichen klinischen Proben bestimmen, was darauf hinweist, dass dieser Assay in der überwiegenden Mehrheit der Patienten für die Analyse kausaler Haplotypen im OPN1LW/OPN1MW-Gencluster ausreichend ist.

### Methoden

#### DNA-Proben aus dem klinischen Archiv

Ein diagnostisches Labor kann (anonymisierte) archivierte klinische Proben verwenden, um neuartige diagnostische Tests zu validieren und zu implementieren. Die abgeleiteten klinisch relevanten Varianten können geteilt werden, aber ohne ausdrückliche Zustimmung zur Datenweitergabe auf individueller Patientenebene können FASTQ, BAM und VCFs nicht offengelegt werden. Diese Methoden entsprechen auch den relevanten Richtlinien und Vorschriften und wurden vom Ethikrat des Radboud Universitair Medisch Centrum genehmigt (2020-7142) und der Deklaration von Helsinki. Alle klinischen Proben (n=50) wurden zwischen 2015 und 2020 von der Abteilung für Humangenetik des Radboud Universitair Medisch Centrum erhalten. In 43 männlichen Proben wurde der genetische Test zur genetischen Bestätigung der Diagnose von Protanopie (n=1), BED (n=32), BCM (n=8) oder Zapfendystrophie (n=2) angefordert.

Die Art der Sehbehinderung wurde von erfahrenen Augenärzten basierend auf entsprechenden klinischen Untersuchungen diagnostiziert. In sieben weiblichen Proben wurde der genetische Test zur Trägerstatusanalyse angefordert, aufgrund von mehreren betroffenen Kindern (n=1), einem betroffenen Kind ohne positive Familienanamnese (n=4) oder einem betroffenen oder Träger-Geschwister (n=2). Die genomische DNA-Isolation aus peripherem Blut wurde auf automatisierte Weise wie zuvor beschrieben durchgeführt. Kurz gesagt, die genomische DNA-Isolation wurde automatisch unter Verwendung eines HamiltonMicrolab STARautoload-Systems mit einem integrierten Chemagen MSM I-Separationsmodul (Hamilton Robotics GmbH) durchgeführt. Die DNA-Isolation erfolgte mit dem Chemagic DNA blood kit special (PerkinElmer) gemäß den Anweisungen des Herstellers. Für die Zwecke dieser Studie wurden die Proben anonymisiert und mit einer eindeutigen Studiennummer (USN) codiert.

#### MLPA

Ein MLPA-Test spezifisch für das OPN1LW/OPN1MW-Gencluster wurde zusammen mit MRC Holland entwickelt. Das SALSA MLPA Probemix X080 Opsin besteht aus 9 Kontrollsonden und 16 Sonden, die auf das OPN1LW/OPN1MW-Gencluster abzielen, alle auf dem X-Chromosom. Zwei Sonden zielen auf die LCR ab. Sechs Sonden zielen auf OPN1LW (NM\_020061.5): c.-77A in Exon 1, c.300A in Exon 2, c.457C in Exon 3, c.706A in Exon 4 und c.830 A und c.926A in Exon 5. Sechs Sonden zielen auf OPN1MW (NM\_000153.2): c.-62A in Exon 1, c.300G in Exon 2, c.457A in Exon 3, c.706 G in Exon 4 und c.830T und c.926T in Exon 5. Zwei Sonden zielen sowohl auf OPN1LW als auch auf OPN1MW: c.607T in Exon 4 und c.1083A in Exon 6 (Abb. 1a). Die Sondensequenzen sind in Tabelle 4 im Anhang angegeben. Bei den Sonden, die c.457 und c.706 von sowohl OPN1LW als auch OPN1MW anvisieren, wurden innerhalb von 10 Nukleotiden der Zielsequenz Einzel-Nukleotid-Varianten beschrieben. Dementsprechend besteht die Möglichkeit, dass das Vorhandensein einer solchen Variante die Leistung der Sonde beeinflusst.

Die MLPA-Reaktion wurde gemäß dem Protokoll des Herstellers (MRC Holland) auf einem Thermocycler (Geneamp PCR system 9700 (ThermoFisher) oder Veriti 96 well thermal cycler (Applied Biosystems)) durchgeführt. Kurz gesagt, wurden 50–100 ng DNA denaturiert und mit 1,5 µl SALSA MLPA-Puffer und 1,5 µl Sondenmix gemischt. Nach 16 bis 20 Stunden Hybridisierung bei 60 °C wurden 3 µl SALSA Ligase-Puffer A, 3 µl SALSA Ligase-Puffer B, 1 µl SALSA Ligase-65 und 25 µl H2O hinzugefügt und 15 Minuten bei 54 °C inkubiert. Schließlich wurde eine Polymerase-Kettenreaktion (PCR) durch Zugabe von 2 µl SALSA PCR-Primer-Mix, 0,5 µl SALSA-Polymerase und 7,5 µl H2O durchgeführt. Das PCR-Protokoll umfasste: 35 Zyklen von 30 Sekunden bei 95 °C, 30 Sekunden bei 60 °C und 1 Minute bei 72 °C, gefolgt von einem abschließenden Elongationsschritt von 20 Minuten bei 72 °C. Eine Mischung aus 1 µl MLPA-Probe und 8,8 µl Formamid (Hi-Di, Applied Biosystems) und 0,2 µl Größenstandard (GeneScan 500-Liz, Applied Biosystems) wurde auf einem Fragmentanalysator (Modell 3130, Applied Biosystems) analysiert. Genemarker (V2.6.7, Softgenetics) wurde für die Datenanalyse verwendet.

#### Langstrecken-PCR

Vier Langstrecken-PCRs (LR-PCRs) wurden entwickelt, um das OPN1LW/OPN1MW-Gencluster zu umfassen (Abb. 1a). Ein Amplicon von 14.374 bp wurde spezifisch für die LCR-Region entwickelt (Primer 099-692 GCAAAGGCTCTTCCTTTGTG und 099-693 AGGTGAAGGCAGGGTAAGGT). Ein Amplicon von 15.634 bp umfasste spezifisch das erste Opsin-Gen (meist ein (Hybrid-)OPN1LW) im Cluster (Primer 044-982 GAGGCGAGGCTACGGAGT und 044-984 GCAGTGAAAGCCTCTGTGACT). Ein Amplicon von 14.034 bp umfasste das zweite und gegebenenfalls nachfolgende Opsin-Genkopien (meist (Hybrid-)OPN1MW) im Cluster (Primer 080-273 AGCTGGGAGTACAGGTATTTG und 044-984). Schließlich wurde ein Amplicon von 16.589 bp entwickelt, um einen Teil des letzten Opsin-Gens (meist ein (Hybrid-)OPN1MW) im Cluster zu identifizieren (Primer 099-820 AGGTGTAGAGCCCTAGCAAAC und 099-821 TCTCATTCATAAATTGCTGGTA).

Die PCR wurde wie folgt durchgeführt: 100 ng DNA, 12,5 µl LongAmp Hot Start Taq 2x Master Mix (Bioke), 2 µl 10 µM Vorwärts- und Rückwärtsprimer, 10,5 µl H2O. Das PCR-Protokoll für LCR, erstes Gen und zweite und nachfolgende Gene umfasste: 94 °C für 1 Minute, gefolgt von 30 Zyklen von 30 Sekunden bei 94 °C und 14 Minuten bei 65 °C, gefolgt von einem abschließenden Elongationsschritt von 10 Minuten bei 65 °C. Das PCR-Protokoll für das letzte Gen umfasste: 94 °C für 1 Minute, gefolgt von 30 Zyklen von 30 Sekunden bei 94 °C, 1 Minute bei 56 °C und 15 Minuten bei 65 °C, gefolgt von einem abschließenden Elongationsschritt von 10 Minuten bei 65 °C. Alle Amplicons wurden auf Agarosegel oder DNA ScreenTape Analysis (TapeStation, Agilent) überprüft.

#### Bibliotheksvorbereitung

LR-Amplicons wurden mit AMPure PB-Perlen (Pacific Biosciences) gereinigt, unter Verwendung eines Perlenverhältnisses von 1,5x. Die Bibliotheksvorbereitung wurde gemäß dem Protokoll "Procedure and Checklist—Preparing SMRTbell Libraries using PacBio Barcoded Adapters for Multiplex SMRT Sequencing" (Pacific Biosciences, Part Number 100-538-700-02) durchgeführt. Kurz gesagt, wurden die Probenkonzentrationen mit dem Qubit 3.0 (ThermoFisher Scientific) gemessen. Gleiche molare Mengen an Amplicons, die sich auf eine Gesamteingabe von 200 ng bis 1 µg DNA summieren, wurden verwendet, um eine einstufige Endreparatur und Adapterligation (unter Verwendung von Barcode-Haarpin-Adaptern) durchzuführen, gefolgt von der gleichmolarer Pooling. Der Pool wurde mit AMPure PB-Perlen gereinigt, gefolgt von DNA-Schadensreparatur, Exonuklease-Verdauung und zwei weiteren Runden AMPure PB-Perlenreinigung.

#### Erzeugung von Polymerase-gebundenen SMRTbell-Komplexen und SMRT-Sequenzierung

Die Erzeugung von Polymerase-gebundenen SMRTbell-Komplexen wurde mit der Option "Sample Setup" in SMRTLink (Pacific Biosciences) durchgeführt. Kurz gesagt, wurden Sequenzierungsprimer konditioniert und an die SMRTbell-Bibliothek angelegt, gefolgt von Verdünnung und Bindung der Sequenzierungs-Polymerase. Der Polymerase-gebundene Komplex wurde mit AMPure PB-Perlen gereinigt, und die Konzentration wurde mittels Qubit gemessen. Eine interne Kontrollprobe wurde verdünnt und dem Polymerase-gebundenen Komplex hinzugefügt, zusammen mit DTT, Sequel-Zusatz und Komplexverdünnungspuffer.

Die Sequenzierung wurde mit der Option "Run Design" in SMRTLink durchgeführt. Bibliotheken wurden mit Diffusionsladung mit einer Plattenkonzentration von 4,5 pM geladen. Alle Läufe wurden mit einer Filmlaufzeit von 20 Stunden pro SMRTcell sequenziert und beinhalteten eine Vorverlängerung. Die Sequenzierung wurde auf einem Sequel I-System (Pacific Biosciences) mit ICS-Version 6.0 durchgeführt. Die Amplicons des OPN1

LW/OPN1MW-Genclusters wurden zusammen mit anderen Arten von Langstrecken-Amplicons sequenziert. Sequenzierungsläufe wurden zweimal pro Woche durchgeführt, unabhängig von der Anzahl der verfügbaren Proben, mit maximal 24 verschiedenen Barcodes pro Sequenzierungslauf, ein Barcode kann mehr als ein Amplicon enthalten. Dementsprechend variiert die Abdeckung zwischen den verschiedenen Proben. Die geringe Anzahl von Barcodes garantiert jedoch eine ausreichende Abdeckung für jede Analyse. Die Mindestabdeckung für die Datenanalyse betrug 20x.

#### Datenanalyse und Variantenerkennung

Nach der Sequenzierung wurden die Rohdaten unter Verwendung von CCS-Mapping in SMRTLink analysiert. Die Analyse wurde mit Standardeinstellungen und geringfügigen Änderungen durchgeführt: Es wurden nur Reads, die länger als 10 kb waren, abgebildet, und für das CCS-Mapping wurde als Referenz ein angepasstes GRCh37 mit nur einer Kopie von OPN1MW ausgewählt. Die zusätzliche Einstellung "PlaceGapConsistently" wurde zu den Algorithmusoptionen hinzugefügt, um eine Linksanpassung zu ermöglichen, und "consolidate bam" wurde auf "ON" gesetzt (Standard: OFF). Die erzeugten BAM-Dateien wurden in SeqNext (JSI Medical Systems) hochgeladen, und die Variantenerkennung wurde mit Standardeinstellungen auf OPN1LW (NM\_020061.6) und OPN1MW (NM\_000513.2) durchgeführt.

#### Optische Genomkartierung

Die optische Genomkartierung von Kontrollproben S3.1 bis S3.4 wurde wie zuvor beschrieben durchgeführt, mit geringfügigen Änderungen. Kurz gesagt, wurde ultrahochmolekulare DNA aus 650 µl peripherem Blut (EDTA) unter Verwendung des SP Blood and Cell Culture DNA Isolation Kit gemäß den Anweisungen des Herstellers (Bionano Genomics) isoliert. Pro Probe wurden 750 ng DNA mit dem DLS (Direct Label and Stain) DNA Labeling Kit (Bionano Genomics) markiert, und die markierte DNA wurde auf dem Saphyr-System (Bionano Genomics) unter Verwendung von ICS-Version 5.2 abgebildet. Die annotierte de novo-Assembly-Pipeline wurde mit der Bionano Solve-Software 3.6.1 ausgeführt. Daten des Genome in a Bottle-Samples NA12878 werden mit freundlicher Genehmigung von Bionano Genomics bereitgestellt. Qualitätsstatistiken der Proben sind in Tabelle 5 im Anhang angegeben.