

# Keimbahn-CNV- Analyse mit dem Illumina DNA Prep with Exome 2.5 Enrichment- Workflow

Effektive Erkennung von  
Kopienzahlvarianten mithilfe  
einer Normalgruppe



## Einleitung

Kopienzahlvarianten (CNVs, Copy Number Variations) sind eine wichtige Quelle genetischer Vielfalt beim Menschen.<sup>1</sup> Diese Klasse von Genvarianten zeichnet sich dadurch aus, dass die Anzahl der Kopien eines Gens oder einer genomischen Region zwischen Individuen variiert und große Genduplikationen oder -deletionen umfasst. Die Analyse von CNVs kann Forschern dabei helfen, Varianten zu bestimmen, die mit komplexen Merkmalen oder einer Anfälligkeit für Erkrankungen wie Krebs, Autoimmunerkrankungen, Erbkrankheiten usw. assoziiert sind.<sup>1-4</sup>

Mit NGS-Verfahren (Next Generation Sequencing, Sequenzierung der nächsten Generation) lassen sich, anders als beim Einsatz von Microarrays für den Nachweis von CNVs, weitere Einzelheiten zu strukturellen Varianten erkennen. Darüber können mit der NGS spezifische Varianten unter 50 kb bestimmt werden, was mit Arrays nicht möglich ist. Illumina DNA Prep with Exome 2.5 Enrichment gehört zu einer schnellen und zuverlässigen High-Performance-Lösung für die Exomsequenzierung (WES, Whole Exome Sequencing), die für die NGS-CNV-Analyse geeignete Bibliotheken generiert. In Kombination mit Sequenziersystemen und Analysepipelines von Illumina kann dieses Bibliotheksvorbereitungskit zur Bestimmung von Einzelnukleotid-Varianten (SNVs, Single-Nucleotide Variants), kleinen Insertionen und Deletionen (Indels) sowie großen strukturellen CNVs verwendet werden.

Im vorliegenden Anwendungshinweis wird ein optimierter Workflow für die Exomsequenzierung bei der Keimbahn-CNV-Analyse vorgestellt, der von der Bibliotheksvorbereitung bis hin zu den Erkenntnissen reicht (Abbildung 1). Die WES-Lösung vereint Illumina DNA Prep with Exome 2.5 Enrichment, die bewährte Illumina-NGS sowie die hochpräzise DRAGEN™-Sekundäranalyse zur Bestimmung von CNVs.

Anwendern steht zudem Emedgene™ für die Varianteninterpretation zur Verfügung, eine Forschungsplattform auf Basis erklärbarer künstlicher Intelligenz (XAI, Explainable Artificial Intelligence). Die Emedgene-Software ermöglicht die Analyse und Auswertung von Varianten sowie die Kuratierung und die Automatisierung von Standardverfahren (SOP, Standard Operating Procedure). Zudem umfasst die Software optimierte Optionen für die Erstellung von Forschungsberichten.

## Methoden

### Proben

Für den Keimbahn-Nachweis von CNVs im Exom müssen Anwender eine Baseline-Referenzdatei anhand eines Panels mit mutmaßlichen Normalproben für den Vergleich während der Analyse entwickeln. Diese Normalgruppenreferenz dient der Normalisierung der Target-Reads und ermöglicht den genauen Nachweis von CNVs bei der Untersuchung von Fallproben. Die Proben für die Normalgruppe müssen unter den gleichen Bedingungen wie die Fallproben vorbereitet und sequenziert werden.

Zur Erstellung der Normalgruppe wurden 54 Proben (Coriell Institute for Medical Research) ohne bekannte pathogene CNVs (Tabelle 1) ausgewählt, um die Vielfalt der natürlichen Variation zwischen Human-Superpopulationen abzubilden (basierend auf dem 1000 Genomes Project, Tabelle 2).<sup>5</sup> Zum Nachweis der Performance des WES-Workflows für den CNV-Nachweis wurden 17 Fallproben mit einer zuvor nachgewiesenen CNV innerhalb eines Gens oder einer Region von Interesse gemeinsam mit der Normalgruppe analysiert.

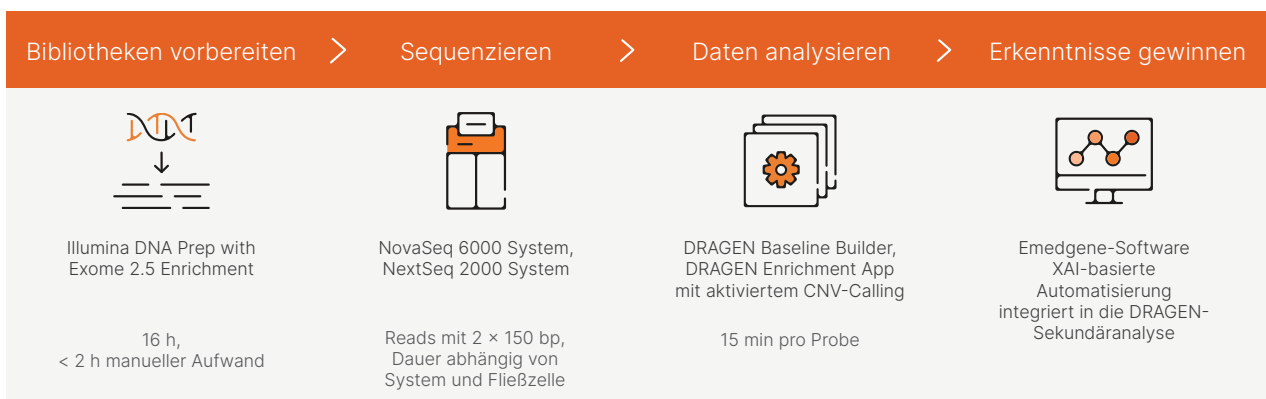


Abbildung 1: Umfassender WES-Workflow für das CNV-Calling: Durch die Kombination von Illumina DNA Prep with Exome 2.5 Enrichment, Illumina-NGS-Plattformen, schneller Datenanalyse mit der DRAGEN-Sekundäranalyse und der Variantenauswertung mit der Emedgene-Software profitieren Sie von einem anwenderfreundlichen, rationalisierten Workflow für die genaue Exomsequenzierung zur Untersuchung von CNVs.

Tabelle 1: Die Proben-IDs des Coriell Institute sind in der Normalgruppe enthalten.

HG00096	HG01393	HG01950	HG03006	NA10830	NA12877	NA18502	NA18970	NA21112
HG00190	HG01441	HG01985	HG03870	NA10831	NA12878	NA18508	NA19681	NA24143
HG00262	HG01551	HG01990	HG03882	NA10835	NA12889	NA18622	NA19720	NA24149
HG00626	HG01599	HG02013	HG03898	NA10838	NA12890	NA18637	NA20509	NA24631
HG00628	HG01896	HG02348	HG04090	NA10839	NA12891	NA18942	NA20875	NA24694
HG01392	HG01914	HG02521	HG04214	NA12249	NA12892	NA18957	NA21098	NA24695

### Bibliotheksvorbereitung

Bibliotheken für die Exomsequenzierung wurden gemäß dem Illumina DNA Prep with Exome 2.5 Enrichment-Protokoll (Dokument-Nr. 100000048041 v07) anhand von 50 ng gDNA aus Coriell Institute-Proben in doppelter Ausführung vorbereitet, sofern nicht anders angegeben. IDT for Illumina DNA/RNA UD Indexes Sets A-D, Tagmentation (Illumina, Katalog-Nr. 20027213, 20027214, 20042666 und 20042667) wurden zur eindeutigen Indizierung über alle vorbereiteten Bibliotheken hinweg verwendet. Die Voranreicherungsbibliotheken wurden nach Masse (Soll von 250 ng pro Probe) bei 12-Plex pro Pool gepoolt und über Nacht mit dem Twist Bioscience for Illumina Exome 2.5 Panel hybridisiert, das als Bestandteil der Illumina DNA Prep with Exome 2.5 Enrichment-Kits (Illumina, Katalog-Nr. 20077595 und 20077596) erhältlich ist. Die Bibliotheken wurden mit und ohne Spike-in des Twist Bioscience for Illumina Mitochondrial Panel (Illumina, Katalog-Nr. 20093180) vorbereitet. Das Spike-in des Mitochondrial Panel erfolgte bei der im Referenzhandbuch zu Illumina DNA Prep with Exome 2.5 Enrichment (Dokument-Nr. 100000157112) angegebenen Verdünnung im Verhältnis 1:100.

### Sequenzierung

Die angereicherten Bibliotheken wurden auf dem NovaSeq™ 6000 System und dem NextSeq™ 2000 System unter Verwendung von Standardreagenzien für die Sequenzierung durch Synthese (SBS, Sequencing by Synthesis) und Paired-End-Läufen mit 150 bp sequenziert. Zur Gewährleistung einer hohen Coverage-Tiefe bei allen Proben sowie zur Ermöglichung eines Downsamplings im Rahmen der Analyse wurden die maximal möglichen Datenausgaben verwendet. Die Normalgruppe und die Fallproben wurden mit dem NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) (Illumina, Katalog-Nr. 20028312) und NextSeq 2000 P3 Reagents (300 cycles) (Illumina, Katalog-Nr. 20040561) sequenziert. Reagenzien-Kits für das NovaSeq 6000 System enthalten zur einfachen Anwendung drei gebrauchsfertige, mit allen erforderlichen Reagenzien vorbefüllte Kartuschen.

Tabelle 2: Zusammenfassung der Attribute der Normalgruppe

Geschlecht	Anzahl der Proben
Männlich	27
Weiblich	27
Superpopulation <sup>5</sup>	
Europäisch (EUR)	18
Amerikanisch, gemischt (AMR)	8
Afrikanisch (ARF)	7
Südasiatisch (SAS)	9
Ostasiatisch (EAS)	12

Tabelle 3: Sequenzierungsdurchsatz und -Coverage auf unterschiedlichen Plattformen für Normalgruppen- und CNV-Proben

Plattform	Proben pro Lauf	Durchschnittliche Coverage je Probe
NovaSeq 6000 System	72	ca. 540-fach
NextSeq 2000 System	12	ca. 280-fach

Das NextSeq 2000 System ermöglicht Flexibilität anhand unterschiedlicher Fließzellenkonfigurationen und bietet die Möglichkeit zur geräteinternen DRAGEN-Analyse. Die Anzahl der Proben pro Lauf und die durchschnittliche Coverage, die je Probe über alle Läufe hinweg erreicht wird, werden in [Tabelle 3](#) dargestellt.

## Datenanalyse

Die Normalgruppe mit den einzelnen Target-Zählungen wurde mit der in BaseSpace™ Sequence Hub verfügbaren DRAGEN Baseline Builder App v4.2 anhand des hg38-Nongraph-Referenzgenoms erstellt. Die Sequenzierungsdaten für CNV-Fallproben wurden einem Downsampling auf 80 Mio Reads (ca. 120- bis 140-fache durchschnittliche On-Target-Coverage-Tiefe) und 50 Mio. Reads (ca. 70- bis 90-fache durchschnittlichen On-Target-Coverage-Tiefe) unterzogen und mit der DRAGEN Enrichment App v4.2 mit aktiviertem CNV-Nachweis analysiert, die ebenfalls in BaseSpace Sequence Hub verfügbar ist. Für jede Plattform wurden die Proben anhand der entsprechenden Normalgruppe normalisiert. Die DRAGEN Enrichment App führt eine genaue, effiziente Sekundäranalyse für das umfassende Varianten-Calling (SNV, Indels, CNV und strukturelle Varianten (SV) usw.) sowie andere Anwendungen aus. Das Varianten-Calling mit DRAGEN für Illumina DNA Prep with Exome 2.5 Enrichment kann auch auf dem NextSeq 2000 System erfolgen oder vollständig in den Forschungsworkflow zur Varianteninterpretation der Emedgene-Software integriert werden.

## Ergebnisse

### Variantenübergreifender CNV-Nachweis

Die Analyse mit der DRAGEN-CNV-Pipeline in der DRAGEN Enrichment App zeigte auf sämtlichen Plattformen einen robusten CNV-Nachweis bei den in dieser Studie untersuchten Varianten bei nur 50 Mio. Reads oder ca. 70–90-facher durchschnittlicher Coverage-Tiefe. Der CNV-Nachweis wird anhand folgender Formel bestimmt: (Anzahl der Regionen mit Calling)/(Regionen mit Überschneidung zu Exom-Panel und Referenz). CNVs wurden als nachgewiesen behandelt, wenn diese Kennzahl über 98 % betrug (Tabelle 4).

Mehrere der getesteten Varianten finden sich im *DMD*-Gen, das wichtig für die Erforschung genetischer Erkrankungen ist (Tabelle 4).<sup>6</sup> Mit Integrated Genomics Viewer lassen sich die erwarteten *DMD*-CNV-Ereignisse für die sieben getesteten betroffenen Proben visualisieren (Abbildung 2). Hierbei zeigt sich die Fähigkeit zum CNV-Nachweis in diesem Gen mit dem Illumina Exome 2.5-Workflow. Die BED-Datei des Exome 2.5-Panels gibt die vom Panel abgedeckten Regionen an (Abbildung 2, oben). Die Ergebnisse sind von unterschiedlichen Faktoren abhängig. Es wird empfohlen, die Sequenzierung für die Erstprüfung eines Laborworkflows mit einer mehr als 150-fachen mittleren Coverage durchzuführen, um eine ausreichende Coverage für die gängigsten Varianten sicherzustellen.

## Zusätzliche Analyse mit Emedgene-Software möglich

Die Emedgene-Software rationalisiert und skaliert die Varianteninterpretation, wodurch sich die Dauer pro Forschungsfall um 50–75 % verkürzt.<sup>7</sup> Die zahlreichen Funktionen ermöglichen eine benutzerdefinierte fundierte Varianteninterpretation, darunter XAI für transparente, evidenzgestützte, automatisierte Rankings potenziell verursachender Varianten für Proben, ein laufend aktualisiertes Annotations- und Evidenzdiagramm, Variantenvisualisierung, Variantenkuratierung, benutzerdefinierte Automatisierung und vieles mehr. Die Emedgene-Software wurde im Hinblick auf eine effiziente und intuitive Benutzererfahrung entwickelt.

Zur Optimierung der Fallprüfung stellt die KI-Variantenpriorisierung bzw. „Shortlist“-Funktion Varianten (einschließlich CNVs) zusammen, die bei einem Fall am wahrscheinlichsten zu einem Ergebnis führen. Diese Funktion umfasst weiterführende Evidenz und ermöglicht Analysten erhebliche Zeiteinsparungen. In einer separaten Studie<sup>8</sup> an 51 Singletons, bei denen zuvor eine CNV-Variante zum Ergebnis geführt hatte, wurde die entsprechende Variante in einer Shortlist mit 10 Varianten in 92 % der Fälle identifiziert. Bei 6 % (n = 3) war die zum Ergebnis führende Variante in der Kandidatenliste vorhanden. Während der Fallprüfung ermöglicht die Emedgene-Software das Aufheben des Tagging der durch die Shortlist ausgewählten Varianten sowie das manuelle Tagging von Varianten, die nicht durch die Shortlist ausgewählt wurden. Die automatisierte \*Klassifikation nach ACMG umfasst ebenfalls CNVs.

### Zugriff auf Normalgruppendateien

Die Normalgruppendateien, die mit Proben des Coriell Institute generiert wurden, stellen Forschern einen bekannten Datensatz für den Vergleich mit Fallproben zur Verfügung, ohne dass Zeit und Ressourcen für die Erfassung, Sequenzierung und Analyse dieses Probensatzes aufgewendet werden müssen. Diese Standarddateien eignen sich am besten für Ersttests des Illumina DNA Prep with Exome 2.5 Enrichment-Workflows für die CNV-Analyse. Labore werden dazu angehalten, eine laborprotokollspezifische Normalgruppe zu entwickeln. Wenn diese Dateien für den Ersttest verwendet werden, müssen die Bibliotheken für Fallproben gemäß dem gleichen Protokoll vorbereitet werden, das für die Vorbereitung der Normalgruppe verwendet wurde. Durch Abweichungen beim Protokoll können sich Unterschiede bei den Coverage-Profilen ergeben, was die Wirksamkeit der Standardnormalgruppe verringert. Dateien und zusätzliche Informationen finden Sie auf der Support-Website von Illumina.

\* ACMG, American College of Medical Genetics and Genomics.

Tabelle 4: CNV-Nachweis sortiert nach ungefährender Größe des Ereignisses basierend auf erwarteten Ereigniskoordinaten

Coriell-ID	Chromosom	Gen- (betroffene Exons) oder Chromosomenposition	Erwartetes Ereignis	Ungefähre Größe des Ereignisses (kb) <sup>a</sup>	Erkannt bei 50 Mio. Reads?
NA04315	X	<i>DMD</i> (44)	Verlust	0,15	✓
NA05115	X	<i>DMD</i> (45)	Verlust	0,18	✓
NA05117	X	<i>DMD</i> (45)	Verlust	0,18	✓
NA23599	X	<i>MECP2</i> (3–4)	Verlust	2,25	✓
NA18949	17	<i>BRCA1</i> (15–16)	Verlust	3,59	✓
NA21939	15	<i>FBN1</i> (42–43)	Verlust	3,70	✓
HG03857	16	<i>PALB2</i> (5–7)	Verlust	4,23	✓
HG00343	22	<i>CHEK2</i> (9–10)	Verlust	4,25	✓
NA23127	X	<i>DMD</i> (27–28)	Zunahme	7,47	✓
NA04520	16	<i>TSC2</i> (1–15)	Verlust	16,28	✓
NA04327	X	<i>DMD</i> (5–7)	Zunahme	22,70	✓
NA10283	X	<i>DMD</i> (72–79)	Verlust	54,38	✓
HG00500	2	<i>SPAST</i> (4–17)	Zunahme	58,84	✓
NA23675	X	<i>MECP2</i> (alle)	Zunahme	76,14	✓
NA23087	X	<i>DMD</i> (2–30)	Zunahme	608,45	✓
NA06870	18	18p11.32-18p11.1	Zunahme	15.390,21	✓
NA20125	10	10q23.1-10q26.3	Zunahme	52.877,99	✓

a. Basierend auf den erwarteten Ereigniskoordinaten der betroffenen Exon- oder Chromosomenposition.

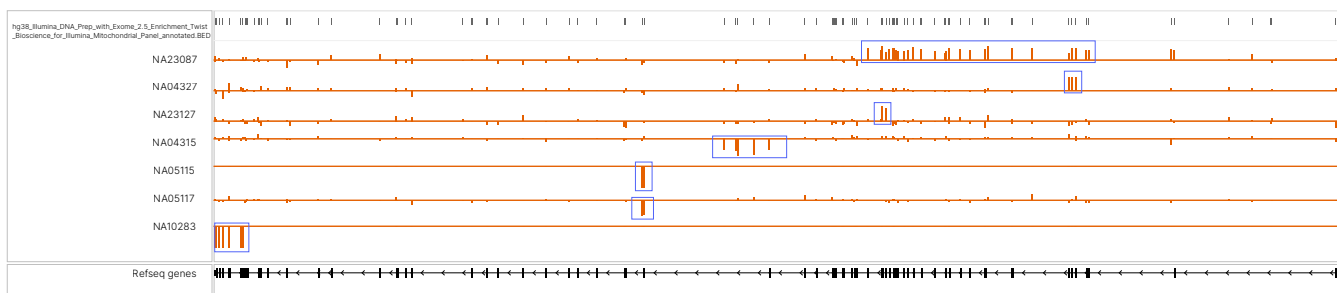


Abbildung 2: Visualisierung der in sieben Proben erkannten CNVs, die erwartete Ereignisse im *DMD*-Gen enthalten: Dateien mit der BigWig-Darstellung des Tangente-Normale-Signals (\*.tn.bw) werden im Rahmen der Analyse mit der DRAGEN Enrichment App ausgegeben und zeigen Zunahme- und Verlustregionen basierend auf den Zielregionen im Twist Bioscience for Illumina Exome 2.5-Anreicherungspanel an. Die Tracks werden automatisch skaliert als Balkendiagramm angezeigt.

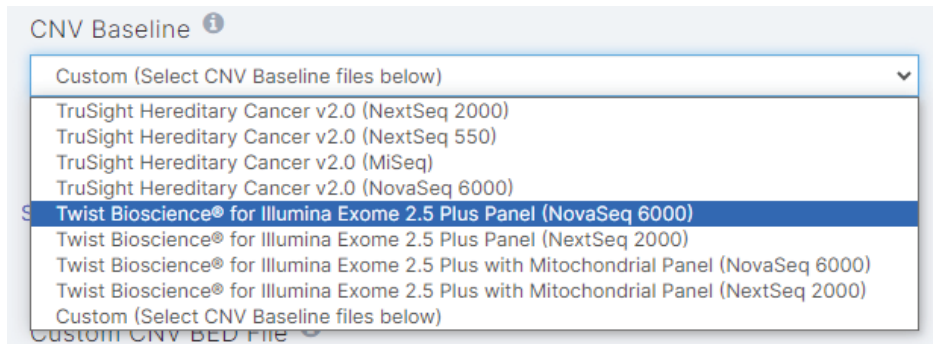


Abbildung 3: Das Normalgruppen-Drop-down-Menü in der DRAGEN Enrichment App v4.3 in BaseSpace Sequence Hub.

Bei Verwendung von BaseSpace Sequence Hub bietet die DRAGEN Enrichment App v4.3 Drop-down-Optionen zur Durchführung einer CNV-Erstanalyse mit dem Twist Bioscience for Illumina Exome 2.5-Anreicherungspanel für mehrere Referenzen (hg19, hg38 und hs37d5, sowohl Graph als auch Nongraph) auf mehreren Plattformen (NextSeq 2000 System und NovaSeq 6000 System) anhand mehrerer Bibliotheken (Twist Bioscience for Illumina Exome 2.5-Anreicherungspanel mit und ohne Twist Bioscience for Illumina Mitochondrial Panel-Spike-in) (Abbildung 3). Fallproben, die mit diesem Analyseworkflow verwendet werden, müssen wie im Abschnitt „Methoden“ des vorliegenden Anwendungshinweises beschrieben vorbereitet werden.

## Zusammenfassung

Beim NGS-Exom-Workflow zur Analyse von CNVs kommen Illumina DNA Prep with Exome 2.5 Enrichment für die Bibliotheksvorbereitung, die Sequenzierung auf dem NovaSeq 6000 System oder dem NextSeq 2000 System sowie DRAGEN-Anwendungen für die Datenanalyse zum Einsatz. Die Ergebnisse zeigen die Anwendung einer Normalgruppe als Referenz für die CNV-Analyse von Fallproben unter Verwendung des erläuterten WES-Workflows. Die hohe CNV-Recall-Quote der DRAGEN-Sekundäranalysesoftware korreliert stark mit anderen Methoden. Dieser Workflow kann einschließlich der Verwendung der Normalgruppe zur CNV-Analyse auch an andere Illumina-Sequenzierungsplattformen angepasst werden. Die Emedgene-Software unterstützt Labore bei der Durchführung zusätzlicher Analysen wie Varianteninterpretation und Erstellung von Forschungsberichten.

## Weitere Informationen

[Referenzdaten der Normalgruppe](#)

[Analyse von Kopienzahlvarianten](#)

[Illumina DNA Prep with Exome 2.5 Enrichment](#)

[Sequenzierungsplattformen von Illumina](#)

[DRAGEN-Sekundäranalyse](#)

[DRAGEN Germline-Anwendung](#)

[Emedgene-Software](#)

## Quellen

1. Zhang F, Gu W, Hurles ME, Lupski JR. [Copy number variation in human health, disease, and evolution](#). *Annu Rev Genomics Hum Genet.* 2009;10:451-481. doi:10.1146/annurev.genom.9.081307.164217
2. Mehawej C, Maalouf JE, Abdelkhalik M, Mahfouz P, Chouery E, Megarbane A. [CNV Analysis through Exome Sequencing Reveals a Large Duplication Involved in Sex Reversal, Neurodevelopmental Delay, Epilepsy and Optic Atrophy](#). *Genes (Basel).* 2024;15(7):901. doi:10.3390/genes15070901
3. Fang X, Ma M, Rong W, et al. [Exome sequencing confirms the clinical diagnosis of both joubert syndrome and klinefelter syndrome with keratoconus in a han Chinese family](#). *Front Genet.* 2024;15:1417584. doi:10.3389/fgene.2024.1417584
4. Zeng Y, Ding H, Wang X, et al. [High positive predictive value of CNVs detected by clinical exome sequencing in suspected genetic diseases](#). *J Transl Med.* 2024;22(1):644. doi:10.1186/s12967-024-05468-1
5. Harrison PW, Amode MR, Austine-Orimoloye O, et al. [Ensembl 2024](#). *Nucleic Acids Res.* 2024;52(D1):D891-D899. doi:10.1093/nar/gkad1049
6. Kozareva V, Stroff C, Silver M, Freidin JF, Delaney NF. [Clinical analysis of germline copy number variation in DMD using a non-conjugate hierarchical Bayesian model](#). *BMC Med Genomics.* 2018;11(1):91. doi:10.1186/s12920-018-0404-4
7. Greenwood Genetic Center. [GGCC reduces turn around time on genomic analysis by 75% with Emedgene's AI platform](#). [ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform](https://ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform). Veröffentlicht am 12. September 2019. Aufgerufen am 14. August 2024.
8. Archivdaten. Illumina, Inc. 2024.



1 800 8094566 (USA, gebührenfrei) | +1 858 2024566 (Tel. außerhalb der USA)  
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Alle Rechte vorbehalten. Alle Marken sind Eigentum von Illumina, Inc. bzw. der jeweiligen Inhaber.  
Spezifische Informationen zu Marken finden Sie unter [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).  
M-GL-01453 DEU v1.0