

# Molekulare Karyotypisierung in der genetischen Diagnostik

## Technologien und Anwendungen

Die klassische Untersuchung der Chromosomen (Karyotypisierung) erfolgt mikroskopisch an Zellen, die während der Metaphase der Zellteilung fixiert worden sind. Diese konventionelle Karyotypisierung kann in Abhängigkeit von der chromosomalen Region eine maximale Auflösung von etwa 5 Megabasen (Mb) erreichen. Die chromosomalen Anomalien der bekannten Mikrodeletions- und Mikroduplikations Syndrome weisen definitionsgemäß eine geringere Größe auf und sind mikroskopisch nicht erkennbar. Klinische Diagnosen dieser Syndrome können bei einem bestimmten Verdacht durch eine genetische Analyse, z. B. durch eine MLPA- („multiplex ligation-dependent probe amplification“) oder FISH-Untersuchung („fluorescence in situ hybridization“) überprüft werden. Die Treffsicherheit der Diagnostik ist jedoch aufgrund der heterogenen phänotypischen Ausprägung gleicher Syndrome oder bei nichtsyndromaler Entwicklungsverzögerung vermindert [5]. Außerdem ist eine genomweite Analyse mit lokusspezifischen Methoden nicht sinnvoll. Seit der Einführung der molekularen Karyotypisierung können nun genomweit Deletionen und Duplikationen (einer bestimmten Minimalgröße) zuverlässig detektiert werden (s. Artikel Rauch et al., Siebert et al., Gillesen-Kaesbach et al., Held et al.). Eine Vielzahl neuer Mikrodeletions- und Mikroduplikationssyndrome sind durch

die molekulare Karyotypisierung entdeckt worden. Die Detektionsrate von diagnostisch relevanten Aberrationen konnte bei klinisch nicht identifizierbaren Fällen von etwa 3–5% (mit den klassischen Methoden) auf etwa 15–20% gesteigert werden [8].

### Molekulare Karyotypisierung

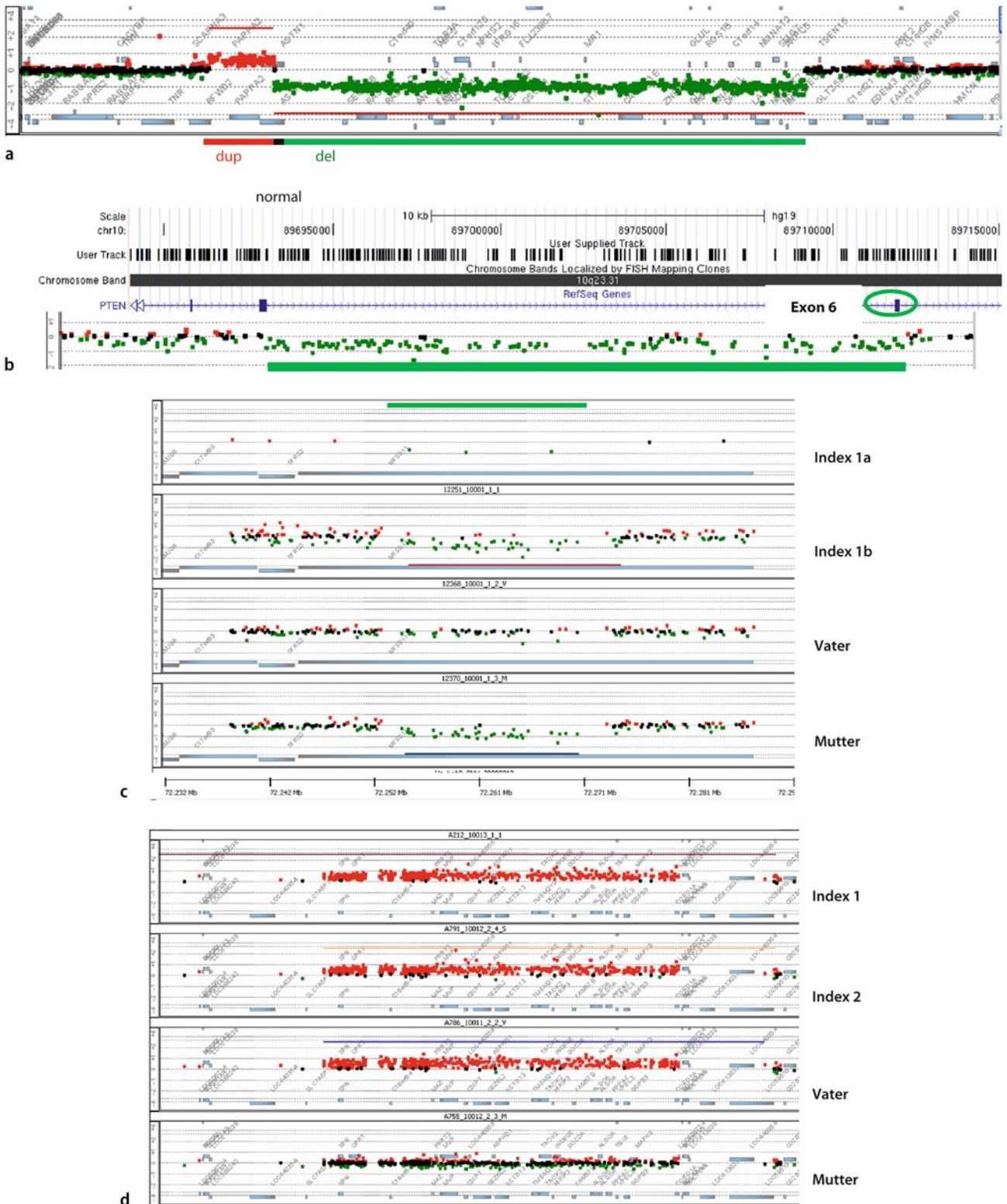
Grundsätzlich kann man zwei verschiedene Typen der molekularen Karyotypisierung unterscheiden. Zum einen den Oligonukleotid-Array, der nur Daten zu Kopienzahlvarianten („copy number variants“, CNVs) erhebt, und zum anderen den SNP-Array (SNP: „single nucleotide polymorphism“), der neben Gewinnen und Verlusten auch Auskunft über die Allelverteilung bekannter SNPs liefert. Mittlerweile bieten fast alle Anbieter auch Kombinationen von beiden an. Die molekulare Karyotypisierung wird heutzutage für die Diagnostik bei geistiger Entwicklungsstörung (EBM Kapitel 11.4, GOP 11.500) und zur Charakterisierung von Tumoren eingesetzt.

### Oligonukleotid-Array-CGH

Bei der Oligonukleotid-Array-CGH („comparative genomic hybridization“) werden sog. Arrays (Matrizen) eingesetzt, die auf einen Objektträger oder ummantelte Chips aufgebracht sind. Diese Ar-

rays bestehen aus vielen einzelnen Punkten. Jeder dieser Punkte enthält wiederum eine feste Menge der gleichen einzelsträngigen Oligonukleotide, die jeweils einen definierten Abschnitt im Humangenom repräsentieren. An diese Sonden binden während der Hybridisierung Abschnitte fluoreszenzmarkierter DNA, die ein Gemisch aus Test- und Referenz-DNA darstellt. Zuvor werden beide mit einem jeweils spezifischen Farbstoff markiert. Thermodynamische Prozesse veranlassen die einzelsträngige DNA, wie sie für die Hybridisierung eingesetzt wird, wieder Doppelstränge zu bilden. Dabei kommt es zu einer kompetitiven Bindung der Test- und Referenz-DNA an die Sonden. Sind für einen Abschnitt die gleiche Menge an Test- und Referenzprobe vorhanden, so findet man nach der Hybridisierung die gleiche Menge der beiden Fluoreszenzfarbstoffe an diesem Punkt des Arrays. Liegt für einen Abschnitt des Genoms nicht das gleiche Verhältnis von Test- und Referenzmaterial vor, wird mehr Farbstoff von der im Überschuss vorhandenen DNA an dem korrespondierenden Punkt des Arrays zu finden sein.

Die Verteilung der gebundenen Fluoreszenzfarbstoffe wird mit einem hochauflösenden Scanner gemessen. Nach einer Normalisierung der Messwerte werden die  $\log_2$ -Verhältnisse der Farbintensitäten berechnet. Liegen an einem Punkt die gleichen Intensitätsmesswerte für bei-



**Abb. 1** ▲ Graphische Darstellung der Ergebnisse der molekularen Karyotypisierung mittels der  $\log_2$ -Verhältnisse der Farbintensitäten (normal: schwarze Punkte, Deletion: grüne Punkte, Duplikation: rote Punkte) auf Agilent 244 k bzw. 400 k und Custom-Arrays. **a** Exemplarische Darstellung einer heterozygoten Duplikation und einer heterozygoten Deletion, wie sie in einer Person mit nichtsyndromaler geistiger Entwicklungsverzögerung identifiziert wurden. Zwischen beiden Bereichen liegt ein „normaler“ Abschnitt. **b** Custom-Array-CGH: Identifikation einer heterozygoten Deletion von Exon 6 im *PTEN*-Gen. Bei dem Patienten war klinisch ein Cowden-Syndrom diagnostiziert worden. Die Sequenzierung von *PTEN* war unauffällig, die MLPA-Daten konnten nicht eindeutig interpretiert werden. **c** Custom-Array-CGH: Analyse der familiären Segregation einer heterozygoten Deletion (13–21 kb) auf 17q25.2, bei der die Aberration aus der ursprünglichen Analyse (Index 1a, 244 k Array) nur an 3 Sonden sichtbar war. **d** Custom-Array-CGH: Familiäre, heterozygote Duplikation 16p11.2 („Autismus-Locus“) bei Index 1 und 2 mit Entwicklungsstörung sowie beim nichtbetroffenen Vater, analysiert per Custom-Array

de Farben vor, so beträgt das  $\log_2$ -Verhältnis null. Liegt eine heterozygote Deletion in der Test-DNA vor, so beträgt dieser Wert  $-1$ . Liegt eine heterozygote Duplikation vor, beträgt er  $0,585$  (■ **Abb. 1a**). Es handelt sich dabei um ideale Werte, die in der Realität variieren (können). Für jedes Chromosom werden diese Werte graphisch dargestellt (■ **Abb. 1a**). Zur automatischen Identifikation der Aberrationen werden verschiedene Algorithmen eingesetzt. Eine visuelle Kontrolle der errechneten Chromosomenprofile gibt zusätzliche Sicherheit.

Die Auflösung der Array-CGH ist abhängig von der Anzahl der eingesetzten Sonden, woraus sich die Größe der kleinsten detektierbaren Kopienzahlvarianten ableiten lässt (etwa 5–10 kb bei 1 M Oligo-Arrays und bis unter 1 kb bei Custom-Oligo-Arrays). Die Methode detektiert jedoch keine balancierten Translokationen, Inversionen, Punktmutationen und niedriggradige Mosaik.

### SNP-Arrays

Im Unterschied zur molekularen Karyotypisierung durch Oligonukleotid-Array-CGH werden bei SNP-Arrays zur genomweiten Detektion verschiedener Mutationstypen wie CNVs, uniparentale Disomie (UPD) und Homozygotieregionen neben den Intensitätsdaten zusätzlich Genotypdaten der analysierten SNPs erhoben. Hier wird nur eine genomische DNA, die Patienten-DNA, hybridisiert und die erhobenen Messwerte mit zuvor analysierten Kontrollexperimenten verglichen. Die SNP-Genotypen werden durch verschiedene Methoden (Illumina: „single base extension“, Affymetrix: „single-/multiple-base mismatch hybridization“) analysiert, durch welche die beiden SNP-

Allele unterschiedlich fluoreszenzmarkiert und unabhängig gemessen oder gescannt werden [10]. Bei der CNV-Analyse mit SNP-Arrays werden 2 Parameter analysiert:

- die Gesamt-Fluoreszenzintensität der Marker und
- deren SNP-Allelverteilung.

Die Analyse der Intensitätswerte entspricht weitestgehend der Array-CGH: Erniedrigte Intensitätswerte (bei Illumina „ $\log_2$  R ratio“ genannt) zeigen Kopienzahlverluste/Deletionen an, erhöhte  $\log_2$ -R-Ratios demgegenüber Kopienzahlzugewinne, also Duplikationen/Amplifikationen (■ **Abb. 2a, b**). Zusätzlich zeigt die Allelverteilung (bei Illumina „B allele frequency“ oder BAF) Abweichungen von der für Diploidie typischen trimodalen Verteilung mit 3 Punktwolken an, bei denen die homozygoten SNPs (AA: BAF  $\cong 0$ , BB: BAF  $\cong 1$ ) die obere und untere Wolke bilden und die heterozygoten SNPs (AB: BAF  $\cong 0,5$ ) die mittlere: Bei der Deletion fehlt die Punktwolke der heterozygoten SNPs, bei der Duplikation ist sie wegen der zusätzlichen Kopie ersetzt durch zwei neue Wolken für AAB und ABB.

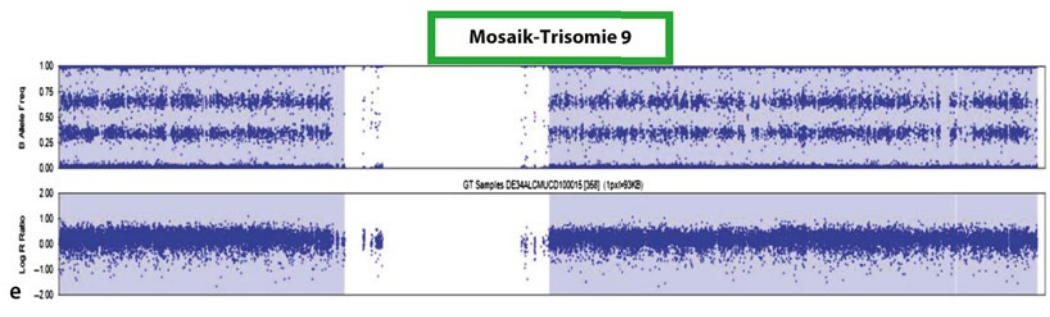
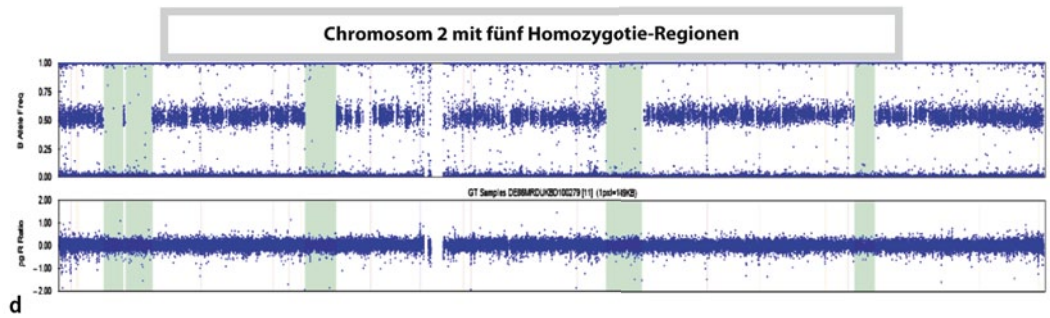
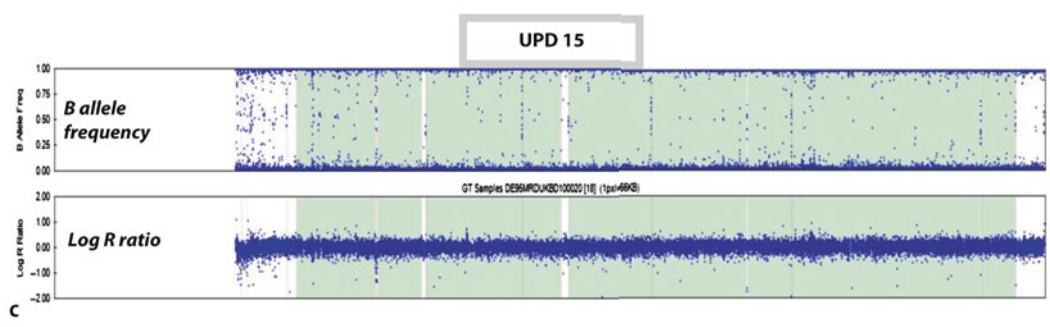
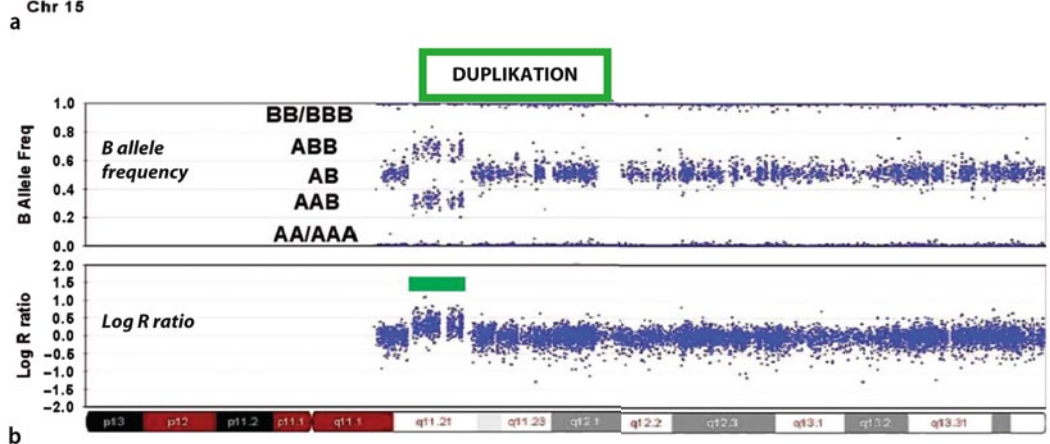
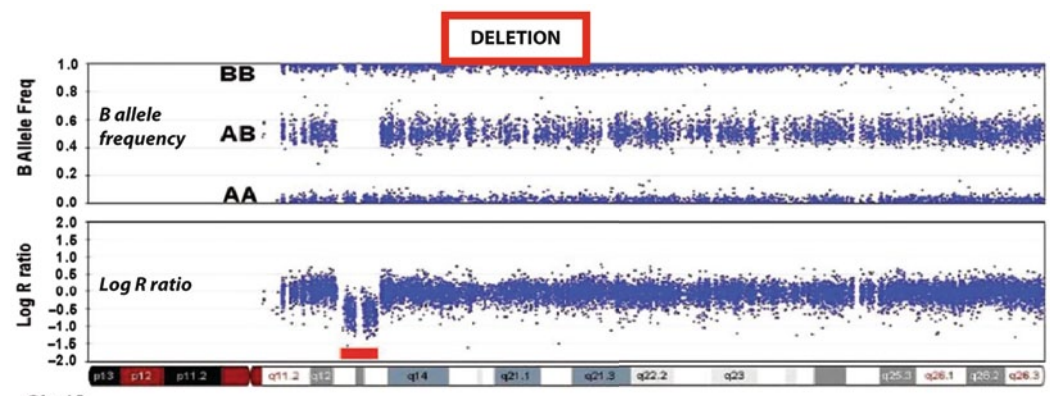
Diese für die SNP-Arrays typische Analyse der Allelverteilung ermöglicht aber auch die Detektion von kopienzahlneutralen Aberrationen wie UPD sowie von Homozygotieregionen. Zeigt der SNP-Array bei normaler Gesamtintensität und damit unveränderter Kopienzahl ein Fehlen der heterozygoten Punktwolke bei einem ganzen Chromosom oder großen Teilen eines Chromosoms, ist dies ein starker Hinweis auf eine isodisome uniparentale Disomie/UPD (■ **Abb. 2c**). Tritt ein solches Muster bei mehreren Regionen unterschiedlicher Chromosomen auf, deutet dies auf Homozygotieregio-

nen hin (■ **Abb. 2d**). Dies kann in diagnostischen Situationen, z. B. bei rezessiven Verdachtsdiagnosen mit ausgeprägter Lokusheterogenität, hilfreich sein, in denen die Homozygotieregionen die Zahl der möglichen ursächlichen Gene verringert („phenotype to genotype“). Selten ermöglicht dies sogar eine Genotype-to-phenotype-Diagnose durch die Identifizierung einer Homozygotieregion, die ein bekanntes Gen für ein rezessiv vererbtes Krankheitsbild enthält, das zuvor nicht die Verdachtsdiagnose war [1]. Allerdings können so auch zuvor unbekannte Verwandtschaftsbeziehungen aufgedeckt werden. Dies macht entsprechende Anpassungen z. B. bei der Aufklärung und Einverständniserklärung der Patienten sowie ggf. der Befundmitteilung notwendig [10]. Ein weiterer Vorteil der SNP-Arrays ist die im Vergleich zur reinen Intensitätsmessung höhere Sensitivität bezüglich Mosaiksituationen [2]. Sie ermöglicht zumindest bei gezielter Auswertung die Detektion von bis zu 7% igen Trisomie-Zelllinien bei günstiger Allelverteilung und von bis zu 12% igen Trisomie-Zelllinien bei ungünstiger Allelverteilung (■ **Abb. 2e**). Erste systematische Untersuchungen mit SNP-Arrays deuten darauf hin, dass besonders Mosaik-Aneuploidien mit 1% und auch UPD mit 0,4% häufiger als gemeinhin angenommen in Patientenkollektiven auftreten [2, 3].

### EBM-Kapitel 11.4

Das EBM-Kapitel 11.4 (GOP 11,500) schreibt eine minimale Auflösung der Arrays von 200 kb vor, damit die genetische Diagnostik bei entsprechender medizinischer Indikation erstattungsfähig ist. Diese Grenze wird nur von den sog. BAC-Arrays („bacterial artificial chromosome“)

**Abb. 2** ▶ SNP-Arrays: Beispielhafte Intensitäts- und Genotypprofile bei **a** Deletion, **b** Duplikation, **c** isodisomer UPD 15, **d** großen Homozygotiebereichen, **e** Mosaik-Aneuploidie. *Jeweils oben* Genotypprofile („B allele frequency“, BAF), *jeweils unten* Intensitätsprofile („ $\log_2$ -R-Ratio“). **a** Deletion in 15q13 (rote Markierung): Verringerte Intensitätswerte/ $\log_2$ -R-Ratio und Abwesenheit heterozygoter Genotypen („B allele frequency“). **b** Duplikation in 22q11.21 (grüne Markierung): Erhöhte Intensität/ $\log_2$ -R-Ratio, zusätzlich ersetzen bei der BAF 2 Wolken für AAB (BAF  $\cong 0,33$ ) und ABB (BAF  $\cong 0,66$ ) die mittlere, heterozygote Punktwolke (AB). **c** Isodisome UPD: Unveränderte Intensitätswerte ( $\log_2$ -R-Ratio  $\cong 0$ ) und (bis auf wenige Artefakte) Abwesenheit heterozygoter Genotypen mit BAF  $\cong 0,5$  für das gesamte Chromosom 15. **d** Chromosom 2 mit 5 Homozygotieregionen (hellgrüne Markierungen): Unveränderte Intensitätswerte ( $\log_2$ -R-Ratio  $\cong 0$ ) und Abwesenheit heterozygoter Genotypen für 5 Regionen. **e** Mosaik-Trisomie 9: Die Intensitätswerte sind weniger deutlich erhöht als bei einer durchgängigen Trisomie und zeigen eine größere Streuung, während das Allelmuster/BAF darauf hinweist, dass die beiden Haplotypen der euploiden Zelllinie auch in der triploiden Zelllinie vorliegen. Die BAF streuen um etwa 0,6 und 0,35 und weichen damit ebenfalls von der bei einer durchgängigen Trisomie zu erwartenden Situation ab. Die Haplotypverteilungen und das Fehlen von sichtbaren Rekombinationsereignissen sind mit einer mitotischen Entstehung des Mosaiks vereinbar



Ab#	Chr	C-band	Upst. Fl. Probe	St. Pos	Probe Name (start)	End Pos	Probe Name (end)	Dwns. Fl. Probe	# of Prb.	Amp/Del	# of Genes	Size (start-end)	Size (flnk. probes)	Links (U/D)
* #1	chr1	q21.3	150.819.451	150.822.873	-6678	150.848.709	-5273	150.852.858	5	1.102622	2	25.836	33.407	1 2 3 4
					ROI	CNV (DGV)								
* #2	chr4	q13.2	68.905.670	69.057.535	-8724	69.166.014	-4384	69.643.272	11	-2.530525	7	108.479	737.602	1 2 3 4
					ROI	CNV (DGV)								
* #3	chr4	q13.2	69.120.805	69.138.837	-4373	69.166.014	-4384	69.643.272	4	-5.165963	3	27.177	522.467	1 2 3 4
					ROI	CNV (DGV)								
* #7	chr9	p24.3	146.264.846	152.931	-4629	256.245	-5624	264.517	9	0.559206	5	103.314	-146.000.329	1 2 3 4
					ROI	CNV (DGV)								
* #8	chr9	q34.2	136.463.017	136.471.996	-4460	136.472.396	-4460	136.480.963	3	0.962387	1	400	17.946	1 2 3 4
					ROI	CNV (DGV)								
* #12	chr14	q32.33	105.033.584	105.040.857	-4404	105.041.257	-4404	105.045.371	3	1.337543	1	400	11.787	1 2 3 4
					ROI	CNV (DGV)								
* #13	chr17	q21.31	41.515.622	41.521.344	-9993	41.707.070	-1346	41.841.187	16	0.986284	4	185.726	325.565	1 2 3 4
					ROI	CNV (DGV)								
* #14 X	chr17	q25.2	72.248.697	72.255.320	-3150	72.268.707	-3417	72.277.456	3	-0.788412	1	13.387	28.759	1 2 3 4
					ROI	unbekannt; Mfsd11								
						Custom Array	72.255.520-72.277.515	priority						
							(familial (mat) - inherited from normal mother)							

a

18q23

Chip	Gender	Sex Mismatch	Amplification/Deletion	Upstream Flanking Probe	Start Position	End Position	Downstream Flanking Probe	Length	Links
13191_10981_1			0.98865	75.746.292	75.759.625	75.789.414	75.795.917	29.789	1 2 3 4
16688_2185015016_2			0.915055	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
17039_2185016301_1			0.896968	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
14969_14152_1			0.881269	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
15919_14149_1			0.843057	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
17167_2185016543_1			0.83177	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
14858_14090_2			-0.801355	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
17542_2185016563_1_2			-0.879623	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
16892_2185015065_1			-0.911162	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
17598_2185016564_1_1			-0.933618	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4
17171_2185015538_1			-1.065778	75.746.292	75.759.625	75.779.459	75.783.563	19.834	1 2 3 4

b

**Abb. 3** ▲ Selbstentwickelte Software zur komfortableren Handhabung von Listen mit Kopienzahlvarianten. **a** Partieller Datensatz einer Array-CGH-Analyse (Beispiel aus 1c). Per Klick („Links“) können alle Varianten z. B. im „UCSC Genome Browser“ dargestellt werden. Es gibt Kommentarfelder, in denen vermerkt werden kann, ob es sich um einen Polymorphismus handelt oder ob eine unbekannte Variante vorliegt (rot umrandet). **b** Interne Datenbankabfrage nach einer nicht in den Datenbanken verzeichneten Variante auf 18q23. In diesem Fall wurde die Veränderung mehrfach beobachtet und es handelt sich um einen Polymorphismus

überschritten, bei denen einzelsträngige BACs als Sonden verwendet werden. Aberrationen unterhalb von 200 kb sind mit den derzeit verwendeten Systemen in den meisten Fällen problemlos detektierbar (Tab. 1). Vereinfacht zusammengefasst gelten folgende Voraussetzungen als Indikation für eine postnatale molekulare Karyotypisierung: eine isolierte Intelligenzminderung (IQ<70), geistige Behinderung in Kombination mit dysmorphologischen Merkmalen, Entwicklungsstörungen des Autismus-Formenkreises, Fehlbildung und schwere Funktionsstörung des Gehirns, multiple angeborene Fehlbildungen und multiple dysmorphologische Merkmale.

### Kopienzahlvarianten

In der molekularen Karyotypisierung werden sog. Kopienzahlvarianten (CNVs) detektiert. Dieser Begriff beschränkt sich nicht auf ein Gen, sondern kann für jedes beliebige Intervall auf einem Chromosom stehen, auch für ein ganzes Chromosom, z. B. die Trisomie 21. Jedoch werden nach bisherigem Kenntnisstand nicht alle detektierten CNVs als pathogen eingestuft, sondern das Gegenteil ist der Fall. Die meisten Varianten, die durch eine Analyse gefunden werden, gehören zur Klasse der Kopienzahlpolymorphismen, also Veränderungen, die mit einer Frequenz >1% in der Bevölkerung vorkommen.

Möglicherweise tragen die Polymorphismen zur phänotypischen Variabilität bei, sie werden jedoch gegenwärtig als benigne Varianten betrachtet, solange ein kausaler Zusammenhang zu einer Erkrankung (noch) nicht nachgewiesen wurde. Die Zahl der derzeit kartierten CNVs ist nicht genau bekannt. In der Literatur wurde 2010 die Zahl mit rund 50.000 CNVs angegeben, die sowohl pathogene als auch benigne und unklassifizierte familiäre Varianten beinhalten [9]. Die Zahl der bekannten CNVs wurde zeitgleich in der Database of Genomic Variants mit etwa 67.000 beziffert.

**Tab. 1** Herstellerangaben zu derzeit verwendeten Arrays/SNP-Arrays (unvollständig).

Hersteller	Produkt	CGH-Marker	SNP-Marker	CNV-Auflösung	LOH-/UPD-Auflösung
Nimblegen	3×1,4 M ISCA Plus	~500.000	64.000	15–20 kb	>5 Mb
Nimblegen	Human CNV 3×720 k v1.0	720.000	n. v.	19–24 kb (5 Sonden)	
Affymetrix	SNP Array 6.0	946.000	906.600	100–200 kb	<3 Mb
Agilent	2×400 k CGH + SNP	300.000	120.000	21–28 kb (4 Sonden)	5–10 Mb
Agilent	2×400 k CGH (Design 21.850)	411.000	n. v.	16–21 kb (4 Sonden)	n. v.
Agilent	4×180 k CGH + SNP (ISCA Design)	120.000	60.000	~100 kb (4 Sonden)	5–10 Mb
Agilent	4×180 k CGH (Design 22.060)	170.000	n. v.	39–52 kb (4 Sonden)	
Illumina	HumanOmni1-Quad	~390.000	~745.000	k. A.	<3 Mb

n. v. nicht vorhanden, k. A. keine Angaben. Eine Vergleich verschiedener Systeme mit Arrays maximaler Auflösung wurde von Haraksingh et al. publiziert [7].

### Pathogen oder nichtpathogen?

Bevor die molekulare Karyotypisierung die Analyse von Genomen möglich machte, waren in erster Linie syndromale Mikrodeletionssyndrome bekannt, deren klinische Diagnosen per FISH, mittels MLPA oder anderer Methoden überprüft wurden, z. B. das Williams-Beuren-Syndrom oder das 22q11.2-Deletions-Syndrom. Findet man eine solche bekannte Mikrodeletion, geht man davon aus, dass das entsprechende Syndrom vorliegt. Die reziproken Ereignisse (Mikroduplikationen) können ebenfalls zu Entwicklungsstörungen führen. Klinisch sind sie aufgrund ihrer hohen Heterogenität bisher nicht oder kaum erkennbar. Manche Aberrationen können allerdings auch von nichtbetroffenen Eltern vererbt werden. Das Vorhandensein von CNVs, wie z. B. an den Loci 15q13.3 und 16p11.2, bei einem nichtbetroffenen Elternteil und bei einem oder mehreren betroffenen Kindern wird mit einer unvollständigen Penetranz erklärt. Hier ist jedoch weitere Ursachenforschung erforderlich. Ein „second hit“ wurde zwar postuliert, konnte aber in vielen Fällen nicht gefunden werden.

Familiäre Varianten können auch zur Demaskierung rezessiver Erkrankungen führen. Liegt eine parental vererbte Deletion vor, die als Ursache aufgrund der darin liegenden Gene in Frage käme, würde man von einem rezessiven Pathomechanismus ausgehen. Die Abklärung kann dann in einem ersten Schritt per Sequenzierung der Kandidatengene erfolgen.

Unsere Erfahrung zeigt, dass im Großteil der Fälle bisher keine weiteren Mutationen gefunden werden. Es ist möglich, dass epigenetische Phänomene auf dem zweiten Allel zur Pathogenität dieser familiären Variante beitragen. Beim Prader-Willi- und dem Angelman-Syndrom können die epigenetischen Veränderungen in Lymphozyten nachgewiesen werden. Da epigenetische Phänomene ebenso gewebespezifisch sein können, nehmen wir an, dass epigenetische Regulation im relevanten (Hirn-)Gewebe Pathogenität bedingen könnte, jedoch ist Hirngewebe für eine Untersuchung aus offensichtlichen Gründen nicht zugänglich.

Für viele kleinere Duplikationen unter 200 kb ist derzeit praktisch keine Vorhersage bezüglich ihrer Pathogenität möglich. Informationen über die Bruchpunkte, den Gehalt an Genen und die Integrationsorte halten wir für wichtige Daten, um künftig die Pathogenität von Duplikationen besser bewerten zu können. Analog zu Sequenzierdaten kann man in diesen Fällen derzeit in der klinischen Anwendung nur von unklassifizierten Varianten sprechen, deren pathogenes Potenzial durch weitere Fälle mit derselben Veränderung abgeschätzt werden kann. Nach heutigem Wissensstand ist davon auszugehen, dass die Mehrzahl dieser Varianten nicht krankheitsverursachend ist [6].

Die S2-Leitlinie „Humangenetische Diagnostik und genetische Beratung“ [12] enthält in 8.4.4. folgende Empfehlung, „dass sowohl die zum Zeitpunkt der

medgen 2012 · 24:86–93  
DOI 10.1007/s11825-012-0326-z  
© Springer-Verlag 2012

K. Hackmann · H. Engels · E. Schröck

## Molekulare Karyotypisierung in der genetischen Diagnostik. Technologien und Anwendungen

### Zusammenfassung

Die molekulare Karyotypisierung durch Array-CGH („comparative genomic hybridization“) und SNP-Arrays (SNP: „single nucleotide polymorphism“) ermöglicht die hochauflösende Untersuchung des gesamten Genoms, um so Gewinne und/oder Verluste (Kopienzahlvarianten, „copy number variants“, CNVs) zu detektieren, die die Ursache einer genetischen Erkrankung sein können. Diese Technik wird in erster Linie zur Ursachenklärung bei syndromalen und nichtsyndromalen (geistigen) Entwicklungsstörungen und zur genetischen Charakterisierung von Tumoren eingesetzt. Auch in der pränatalen Diagnostik könnte die molekulare Karyotypisierung bei auffälligerem sonographischem Befund zur Klärung der Ursachen hilfreich sein. Der Artikel gibt eine kurze Übersicht über die grundlegenden Methoden, deren Grenzen und Stärken sowie einen Ausblick in die Zukunft.

## Molecular karyotyping in genetic diagnosis. Techniques and applications

### Abstract

Molecular karyotyping by array comparative genomic hybridization (CGH) and single nucleotide polymorphism (SNP) arrays allows for a high resolution scan of the entire genome. It detects gains or losses (copy number variants) that might be the underlying cause of a genetic disorder. This technique is mainly applied to cases with syndromal or non-syndromal impaired (intellectual) development and is used to characterize genetic aberrations of tumor samples. Furthermore, molecular karyotyping might be useful for resolving prenatal cases with abnormal ultrasound findings. The purpose of this article is to explain the basic techniques, their limitations and strengths and to provide an outlook on future prospects.

Analyse in den Datenbanken als nichtpathogen gelisteten CNVs als auch familiäre CNVs zumindest dokumentiert und auf Anfrage zur Verfügung gestellt werden, da diese im Einzelfall relevant sein oder aufgrund neuer Erkenntnisse relevant wer-

den können.“ Es sollte also ein Hinweis auf das Vorliegen von derzeit nicht interpretierbaren Ergebnissen in den Befund eingefügt werden. Wir dokumentieren unklassifizierte Varianten im Befund mit dem Hinweis, dass sie derzeit hinsichtlich ihrer Pathogenität nicht beurteilt werden können. Eine Reevaluierung sollte nach einem Zeitintervall von 1–2 Jahren erfolgen, um zu sehen, ob zu diesen Varianten neue Informationen vorliegen.

Zur Bewertung der mittels SNP- oder Oligonukleotid-Arrays gefundenen CNVs werden die experimentellen Ergebnisse mit den Daten aus einer Reihe von CNV-Datenbanken verglichen. Können ähnliche Patienten gefunden werden, ist die ursächliche Beteiligung des CNVs am Phänotyp wahrscheinlich. Gibt es hinsichtlich der Pathogenität in diesen Datenbanken keine Hinweise, müssen alle Gene einzeln geprüft werden, ob sie möglicherweise als Kandidatengen in Frage kommen (Pubmed, RefSeq, Genecards, allgemeine Suchmaschinen usw.). Die Publikation von Einzelfällen und die Speicherung der Daten der molekularen Karyotypisierung in den einschlägigen Datenbanken wird dazu führen, dass immer mehr CNVs eindeutig als pathogen oder nichtpathogen bewertet werden können.

### CNV-Datenbanken

Kopienzahlvarianten werden in verschiedenen Datenbanken gesammelt [4]. Die „Database Of Genomic Variants“ (DGV, <http://projects.tcag.ca/variation/>) sammelt allgemein alle Varianten sowie deren Frequenz, unabhängig davon, ob es sich um pathogene oder nichtpathogene Veränderungen handelt. Eine kleinere und z. T. überlappende Datenbank dieser Art ist die CHOP-Datenbank (<http://cnv.chop.edu/>), die Daten von 2026 gesunden Individuen enthält. Die DECIPHER-Datenbank (<http://decipher.sanger.ac.uk/>) sammelt in erster Linie Daten, die von Personen mit Entwicklungsstörungen erhoben wurden. Je nach Fall können unter Beachtung des Datenschutzes Patientenbilder begutachtet und Informationen über den klinischen Phänotyp, sowie über die familiäre Segregation der Varianten erhalten werden. Ähnliche, meist ausführlichere Daten sind in der ECARUCA-Datenbank vorrätig ([\[tern.umcn.nl:8080/ecaruca/ecaruca.jsp\]\(http://tern.umcn.nl:8080/ecaruca/ecaruca.jsp\)\). Auch das ISCA \(International Standards for Cytogenomic Arrays\) Consortium hat eine eigene Datenbank aufgebaut, in der Arraydaten und klinische Informationen zu Individuen mit geistiger Behinderung, Autismus und Entwicklungsverzögerung enthalten sind \(<https://www.iscaconsortium.org/>\). Der UCSC Genome Browser ist ein wertvolles Onlinetool, das zur Bewertung von CNVs hilfreich sein kann. Hier können DGV, DECIPHER und ISCA parallel als „Tracks“ angezeigt werden. Jedoch kommt es vor, dass die spezifischen Webseiten aktuellere Informationen enthalten. Des Weiteren können diverse Genmodelle, OMIM-Einträge, regulatorische Sequenzen, chromosomale Banden und viele andere Tracks eingeblendet werden.](http://umcecaruca01.ex-</a></p></div><div data-bbox=)

Sehr hilfreich sind auch eigene Datenbanken, in denen arrayspezifische Kopienzahlvarianten aufgelistet sind, die nicht in den öffentlichen Datenbanken zu finden sind. Die eigene Datenbank kann ebenfalls als Track in den UCSC Genome Browser eingespeist werden. Wir verwenden eine spezielle Software zur Auswertung, in der alle bisherigen Array-Ergebnisse gespeichert sind und die per Klick einen Zugriff auf den UCSC-Browser ermöglicht, ohne dass die Intervalle mühsam von Hand kopiert und eingefügt werden müssen (■ **Abb. 3a**). Dadurch kann Analysezeit eingespart werden, es wird die Beurteilbarkeit der CNVs verbessert und die Entdeckung überlappender CNVs aus früheren Analysen ist unmittelbar möglich (■ **Abb. 3b**).

### Maßgeschneiderte Array-CGH

Neben einer Palette von Standardarrays mit verschiedenen Auflösungen können auch maßgeschneiderte Arrays (Custom-Array-CGH) für konkrete Fragestellungen entworfen und angewendet werden. Das ISCA Consortium beispielsweise favorisiert eine höhere Sondendichte in Regionen mit Genen, die als krankheitsverursachend bekannt sind. Der erwartete Vorteil liegt darin, dass diese Gene mit einer höheren Auflösung untersucht werden und daher häufiger diagnostisch relevante Varianten entdeckt werden. Der Nachteil ist, dass unbekannt pathogene Aberrationen seltener entdeckt wer-

den. Wenn sie eine bestimmte Größe nicht erreichen, werden sie nicht wahrgenommen. Aus diesem Grund hat sich das Low-Lands-Consortium (Niederlande) auf einen möglichst gleichmäßigen Sondenabstand verständigt, um sich die Chance zu wahren, neue Syndrome leichter zu entdecken.

Prinzipiell kann für jede Fragestellung, bei der es um Kopienzahlvarianten geht, ein Custom-Array entworfen werden, für Oligonukleotid-Arrays wie für SNP-Arrays. Insbesondere für die Validierung und die Bestimmung der familiären Segregation von kleineren Duplikationen und Deletionen sind Custom-Arrays eine geeignete Methode. Häufig sind diese Varianten so klein, dass sie per FISH nicht bestätigt werden können. Quantitative Polymerasekettenreaktion („polymerase chain reaction“, PCR) wird nicht von jedermann als geeignet eingeschätzt, um ein Kopienzahlverhältnis von 1:2 oder 3:2 zu detektieren. Auch sind häufig keine MLPA-Kits erhältlich, die die zu untersuchende Region abdecken. Weiterhin haben MLPA-Tests eine begrenzte Anzahl von Sonden zur Verfügung und häufig sind verschiedene Gene in einem Kit kombiniert. Durch den maßgeschneiderten Array können die Bruchpunkte in Abhängigkeit der Sondendichte auf weniger als 1 kb genau bestimmt werden (■ **Abb. 1d**). Zwar ist der Preis für die Arrays höher, dafür ist jedoch auch die Auflösung einzelner Gene und der Bruchregionen um ein Vielfaches höher (■ **Abb. 1b**). In der Regel sind Custom-Array-Analysen für die Routinediagnostik jedoch zu aufwendig. Einige Institute legen bei Ihren Analysen Minimalgrößen von Aberrationen fest. Diese Größe ist in der Regel mit FISH (Deletionen) nachweisbar. Neben der klassischen Verwendung von BACs als Sonden sind auch Fosmide (40 kb) und Oligo-FISH-Sonden zur Detektion von Deletionen unterhalb der Auflösungsgrenze von BAC-Sonden geeignet.

### Ausblick

Die molekulare Karyotypisierung hat rasant Einzug gehalten in die genetische Diagnostik. Doch schon drängt sich eine andere Technologie in den Vordergrund, die das Potenzial hat, die konventionelle

und molekulare Karyotypisierung zu verdrängen. Neben ihrem Einsatz in der wissenschaftlichen Forschung wird die Hochdurchsatzsequenzierung bereits in der genetischen Diagnostik genutzt [11]. Insbesondere für Erkrankungen, die durch Mutationen großer oder durch zahlreiche verschiedene Gene ausgelöst bzw. beeinflusst werden können, bietet sich diese Technik an, z. B. für familiären Brust- und Eierstockkrebs, Retinitis pigmentosa oder Fanconi-Anämie. Es gibt aber auch Ansätze, mittels genomweiter NGS-Sequenzdaten („next-generation sequencing“) Rückschlüsse auf Kopienzahlvarianten zu ziehen [13]. Damit lassen sich die SNP- und Oligonukleotid-Arrays wahrscheinlich innerhalb der nächsten 2–3 Jahre vollständig ersetzen. Die Tiefe der Sequenzierung und damit die Anzahl der Reads pro Base bedingt die Auflösungsgrenze für die Identifizierung von Deletionen und Duplikationen wie auch die Kosten der Analyse. Die problematischen kleinen Duplikationen können hinsichtlich ihres Integrationsorts exakt bestimmt werden. Bei einer höheren Anzahl von Reads können auch Translokationen und andere strukturelle Aberrationen erkannt werden.

Bis zum Einsatz in der genetischen Diagnostik sind jedoch noch eine Reihe von Aufgaben zu bearbeiten und zu lösen: Die Menge an Daten pro Genom und Individuum hat einen erheblichen Speicherplatzbedarf. Die Analyse dieser massiven Datensätze verlangt nach Computersystemen, deren Leistung kaum groß genug sein kann. Der Vergleich der erhaltenen Datensätze mit dem Referenzgenom bzw. die De-novo-Sequenzierung verlangt nach Softwarelösungen, die auch von Nichtinformatikern bedient werden können. Wenn all diese Voraussetzungen geschaffen worden sind und ein Genom so wie heute ein Gen von einer einzelnen Person hinsichtlich der Abweichung vom Referenzgenom beurteilt werden kann, ist erst ein vergleichsweise kleiner Schritt gelungen. Die größte Herausforderung wird es sein, den Daten „Sinn“ zu geben.

Je mehr Details des Humangenoms bekannt sind, desto mehr neue Fragen werden sich daraus ergeben. Wie bewertet man die verschiedenen Kombinationen aus Allelen, Einzelnukleotidvarianten (SNV, SNP), CNVs, Indels und Inversio-

nen? Epigenetische Phänomene erhöhen die Komplexität zudem. Die Frage nach den gewebespezifischen Auswirkungen der Varianten des Humangenoms könnte sich letztlich sogar auf die zellspezifischen Funktionen der Varianten erweitern. Das Genom wird also noch lange ein zu untersuchender Gegenstand in der Wissenschaft sein, und auch in der genetischen Diagnostik wird es eine ständige und sehr schnelle Weiterentwicklung geben. Es wird sehr spannend bleiben, wir werden immer mehr diagnostizieren können, und mithilfe der Zellbiologie, Biostatistik und Bioinformatik, Systembiologie und weiterer Fachgebiete werden wir langfristig auch Therapien genetischer Erkrankungen entwickeln können.

---

### Korrespondenzadresse

---

#### Dr. K. Hackmann

Institut für Klinische Genetik  
 Medizinische Fakultät Carl Gustav Carus  
 Technische Universität Dresden  
 Fetscherstr. 74, 01307 Dresden  
 karl.hackmann@tu-dresden.de

---

**Interessenkonflikt.** Der korrespondierende Autor gibt für sich und seine Koautoren an, dass kein Interessenkonflikt besteht.

### Literatur

1. Bruno DL et al (2011) Extending the scope of diagnostic chromosome analysis: detection of single gene defects using high-resolution SNP microarrays. *Hum Mutat* 32(12):1500–1506
2. Bruno DL et al (2011) Pathogenic aberrations revealed exclusively by single nucleotide polymorphism (SNP) genotyping data in 5000 samples tested by molecular karyotyping. *J Med Genet* 48(12):831–839
3. Conlin LK et al (2010) Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet* 19(7):1263–1275
4. Leeuw N de et al (2012) Diagnostic interpretation of array data using public databases and internet sources. *Hum Mutat*
5. Deak KL, Horn SR, Rehder CW (2011) The evolving picture of microdeletion/microduplication syndromes in the age of microarray analysis: variable expressivity and genomic complexity. *Clin Lab Med* 31(4):543–64, viii
6. Hanemaaijer NM et al (2012) Practical guidelines for interpreting copy number gains detected by high-resolution array in routine diagnostics. *Eur J Hum Genet* 20(2):161–165
7. Haraksingh RR et al (2011) Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms. *PLoS One* 6(11):e27859

8. Miller DT et al (2010) Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 86(5):749–764
9. Morrow EM (2010) Genomic copy number variation in disorders of cognitive development. *J Am Acad Child Adolesc Psychiatry* 49(11):1091–1104
10. Schaaf CP, Wiszniewska J, Beaudet AL (2011) Copy number and SNP arrays in clinical diagnostics. *Annu Rev Genomics Hum Genet* 12:25–51
11. Su Z et al (2011) Next-generation sequencing and its applications in molecular diagnostics. *Expert Rev Mol Diagn* 11(3):333–343
12. Deutsche Gesellschaft für Humangenetik e. V., Berufsverband Deutscher Humangenetiker e. V. (2011) S2-Leitlinie Humangenetische Diagnostik. *Med Genet* 23:281–323
13. Scheinin et al. (2012) Posterpräsentation ISCO-Konferenz 2012