

medgen 2014 · 26:264–272
 DOI 10.1007/s11825-014-0449-5
 Online publiziert: 30. Juni 2014
 © Springer-Verlag Berlin Heidelberg 2014

A.W. Kuß

Institut für Humangenetik, Universitätsmedizin Greifswald, und Interfakultäres Institut für Genetik und Funktionelle Genomforschung, Universität Greifswald

Das methodische Potenzial der neuen Sequenziertechnologien jenseits der Mutationsuche

Die neuen Sequenziertechnologien haben nicht nur das Potenzial, neue krankheitsassoziierte Marker zu identifizieren oder krankheitsverursachende Mutationen zu finden, um ein Vielfaches gesteigert, sie eröffnen auch mannigfaltige neue Möglichkeiten für die Untersuchung der funktionellen Auswirkungen krankheitsrelevanter Genveränderungen auf die betroffenen Genprodukte. Darüber hinaus können sie, je nach Fragestellung, auch als wirkungsvolle Werkzeuge für die Analyse übergeordneter Funktionszusammenhänge im Rahmen der Ätiologie genetisch bedingter Erkrankungen eingesetzt werden.

■ **Tab. 1** zeigt einen Überblick über die verschiedenen Techniken und die dazugehörigen Anwendungsmöglichkeiten, die in diesem Beitrag vorgestellt und anhand beispielhafter Fragestellungen näher erläutert werden.

DNA-Sequenzierung

Die Prinzipien der DNA-Sequenzierung mit den verschiedenen NGS-Plattformen wurden im Rahmen dieses Themenhefts von K. Neveling und A. Hoischen bereits eingehend beschrieben.

In der humangenetischen Forschung steht bei diesen Anwendungen naturgemäß v. a. die Suche nach krankheitsassoziierten Markern bzw. neuen krankheitsverursachenden Mutationen im Vordergrund. So ist den NGS-Techniken Z. B. ein deutlicher Anstieg in der Aufklärungsrate bei seltenen genetisch bedingten Erkrankungen zu verdanken [5].

Es soll an dieser Stelle jedoch nicht unerwähnt bleiben, dass diese neue Technologie auch in anderen, z. T. überlappen-

den Bereichen der Lebenswissenschaften eine enorme Dynamik induziert hat. So wurde beispielsweise die Entschlüsselung von bis vor kurzem noch kaum bzw. ganz unerforschten Genomen deutlich erleichtert. Zur Illustration sei hier nur erwähnt, dass seit dem Jahr 2005 als die erste NGS-Plattform (Genome Sequencer 20, Fa. Roche) auf den Markt gebracht wurde, die Zahl vollständig sequenzierter Mikrobengenome von weniger als etwa 250 [22] auf inzwischen mehr als 17.000 angewachsen ist (Genomes Online Database, http://genomesonline.org/cgi-bin/GOLD/index.cgi?page_requested=Complete+Genome+Projects. Zugegriffen: 23. Mai 2014). Diese Entwicklung trägt nicht zuletzt dazu bei, dass die integrierte Analyse von verschiedenen Genomen, etwa im Rahmen metagenomischer Studien, wie Z. B. bei der Untersuchung des menschlichen

Mikrobioms, erhebliche Fortschritte gemacht hat bzw. in dieser Form überhaupt erst möglich wurde [7].

RNA-Sequenzierung (RNA-Seq)

RNA-Sequenzierung mit NGS (RNA-Seq) erlaubt die Erfassung des gesamten Transkriptoms und kann daher auch bei der Suche nach Mutationen helfen [30]. Für weitergehende Untersuchungen sind jedoch v. a. Genexpressionsanalysen mittels RNA-Seq von Bedeutung [24, 29]. So kann über die Analyse des Expressionsmusters eine quasi phänotypische Erweiterung des klinischen Erscheinungsbilds genetischer Erkrankungen erreicht werden [19]. Es ist aber auch möglich, den Einfluss von pathogenen Punktmutationen in Transkriptionsfaktoren auf deren Aktivität zu untersuchen. Damit können

Tab. 1 Next Generation Sequencing (NGS): Methoden und ihre Anwendungsmöglichkeiten in der humangenetischen Forschung

| Methode | Begriffserläuterung | Anwendungsmöglichkeiten |
|----------------------|--|---|
| RNA-Seq | RNA-Sequenzierung mittels NGS | Untersuchung des Transkriptoms (z. B. Erstellung von Expressionsprofilen), Sequenzierung nichtkodierender RNA-Spezies (z. B. microRNAs) |
| ChIP-Seq | Chromatinimmunpräzipitation gefolgt von NGS | Genomweite Proteinbindungsstudien |
| „Ribosome profiling“ | Gezieltes NGS von mRNA-Molekülen, die zu einem bestimmten Zeitpunkt im Rahmen der Translation an Ribosomen gebunden sind | Analyse von Transkriptionsstartpunkten oder der Syntheseraten von Proteinen |
| MeDIP-Seq | Immunpräzipitation methylierter DNA-Segmente („methylated DNA immunoprecipitation“) gefolgt von NGS | Epigenetische Fragestellungen (z. B. Erstellung von Methylierungsprofilen) |
| HiC | Hochauflösende Weiterentwicklung der Chromosome-conformation-capture(3C)-Methode unter Verwendung von NGS | Bestimmung der Chromosomenkonformation im Interphasenuklear |

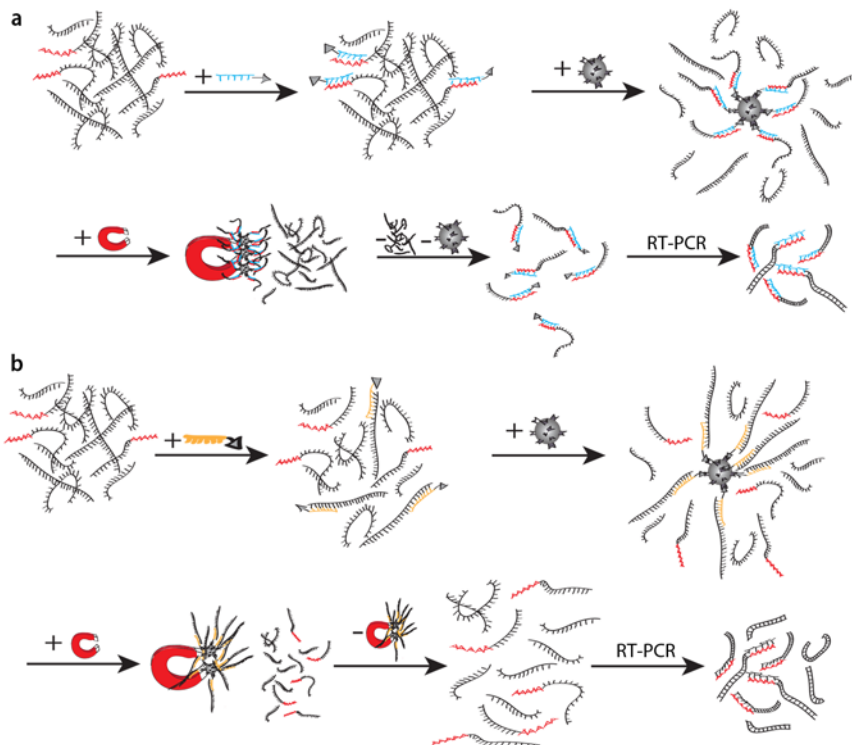


Abb. 1 ▲ mRNA-Isolierung. **a** Spezifische Isolierung polyadenylierter RNA-Spezies (rot) unter Verwendung biotinylierter Oligo-dT-Sequenzen (hellblau), die nach erfolgter Hybridisierung mit den Poly-A-Sequenzen der mRNA über magnetische mit Streptavidin versehene Mikrosphären aus dem Gesamt-RNA-Präparat isoliert werden. **b** Entfernung ribosomaler RNA aus dem Gesamt-RNA-Präparat durch biotinylierte rRNA-spezifische Sonden (gelb) über magnetische mit Streptavidin versehene Mikrosphären. RT reverse Transkriptase

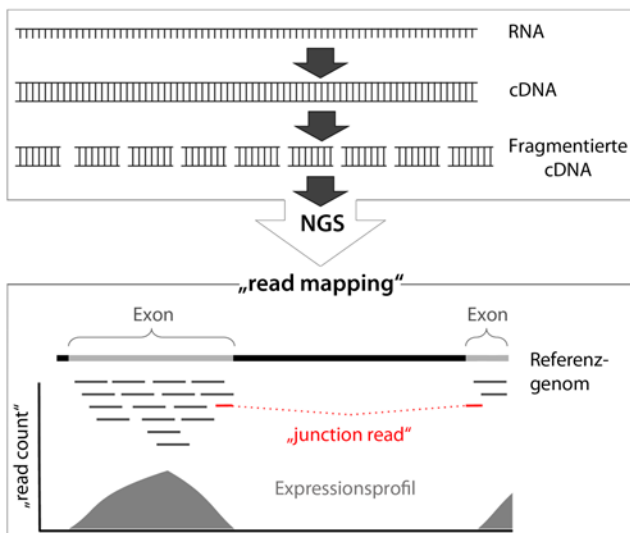


Abb. 2 ▲ Schematische Darstellung des prinzipiellen Vorgehens bei der RNA-Sequenzierung mittels NGS (RNA-Seq). Zunächst werden angereicherte RNA-Spezies über eine reverse Transkriptionsreaktion in cDNA umgeschrieben. Anschließend wird eine Fragmentierung vorgenommen und eine Sequenzierbibliothek für NGS erstellt (oben). Nach erfolgter Sequenzierung werden die gewonnenen Einzelsequenzen mit der Referenzsequenz aligniert (Read-Mapping). Man beobachtet dabei dann Cluster von Einzelsequenzen in den exonischen Regionen. Außerdem fallen „junction reads“ an, die die Grenzen zwischen 2 aneinander gespleißten Exons überspannen (unten)

Informationen über molekulare Zusammenhänge bei der Krankheitsentstehung gewonnen werden.

Anders als die bisher häufig eingesetzten Array-Applikationen zur Transkriptomanalyse gestattet diese Methode darüber hinaus auch die Detektion von bisher unbekanntem Transkripten und Fusionstranskripten [11, 27]. Letztere werden häufig bei der Untersuchung von Tumoren als Resultate von Translokationen beobachtet und man geht davon aus, dass ihnen insbesondere im Rahmen der personalisierten Medizin bei der Diagnose, Vorbeugung und Behandlung von Tumorerkrankungen eine Rolle zukommen könnte [26, 28].

Um überhaupt eine Sequenzierbibliothek erzeugen zu können, muss bei der RNA-Seq-Technik zunächst die ribosomale RNA (rRNA), die bis zu mehr als 80% der gesamten RNA einer Zelle ausmachen kann, entfernt werden. Dies kann zum einen durch die spezifische Isolierung polyadenylierter RNA-Spezies erfolgen, wodurch der größte Teil der proteinkodierenden Sequenzen erfasst wird. Hierfür verwendet man beispielsweise biotinylierte Oligo-dT-Sequenzen, die nach zunächst erfolgter Hybridisierung mit den Poly-A-Sequenzen der mRNA über magnetische mit Streptavidin versehene Mikrosphären aus dem Gesamt-RNA-Präparat isoliert werden können (■ Abb. 1). Diese Vorgehensweise birgt allerdings den Nachteil, dass alle RNA-Spezies ohne Poly-A-Schwanz verloren gehen und nicht für die Sequenzanalyse zur Verfügung stehen. Um dies zu vermeiden, können alternativ auch biotinylierte rRNA-spezifische Sonden eingesetzt werden [8]. Diese erlauben es dann, ebenfalls über eine Biotin-Streptavidin-Bindung und magnetische Mikrosphären, die rRNA aus dem Gesamt-RNA-Präparat zu entfernen (■ Abb. 1). Mit dieser Methode können u. U. selbst formalinfixierte und in Paraffin eingebettete Proben noch für die RNA-Seq aufbereitet werden [25].

In beiden Fällen müssen die verbleibenden RNA-Spezies erst noch über eine reverse Transkriptionsreaktion in cDNA umgeschrieben werden, bevor dann eine Sequenzierbibliothek erstellt werden kann (■ Abb. 2).

Nach erfolgter Sequenzierung werden, analog der DNA-Sequenzierung, die gewonnenen Einzelsequenzen (Reads) mit der Referenzsequenz aligniert (Read-Mapping). Man beobachtet dabei dann Cluster von Einzelsequenzen in den exonischen Regionen. Außerdem fallen „junction reads“ an, die die Grenzen zwischen 2 aneinander gespleißten Exons überspannen (■ **Abb. 2**). Über solche Sequenzen können z. B. auch neue Spleißvarianten und Fusionstranskripte identifiziert werden.

Generell erlaubt die Durchführung von Paired-end-Sequenzierungen, bei denen die cDNA-Fragmente von beiden Seiten aus ansequenziert werden, eine größere Sicherheit beim Read-Mapping. Sie führt außerdem dazu, dass insgesamt weniger Reads als bei Single-end-Sequenzierungen, bei denen die cDNA-Fragmente nur von einem Ende her ansequenziert werden, erforderlich sind, um vergleichbare Ergebnisse zu erzielen. Im Rahmen der RNA-Sequenzierung wird dadurch v. a. auch die Detektion von Fusionstranskripten und Spleißvarianten begünstigt. Die Menge der zu den Exons einzelner Gene gehörenden Einzelsequenzen („read count“) gibt Aufschluss über die Menge des zugehörigen Transkripts in der Probe und liefert über alle transkribierten Bereiche gesehen das Expressionsprofil. Wichtige Einschränkungen, die bei der Datenanalyse berücksichtigt werden müssen, ergeben sich aus möglichen sequenzbedingten systematischen Verzerrungen und stochastischen Messfehlern [3, 6, 12, 16, 20, 31]. Dies gilt letztlich für alle in diesem Beitrag geschilderten Methoden.

Die für die RNA-Sequenzierung benötigte Anzahl an Reads hängt von der Größe des zu untersuchenden Genoms, der Komplexität bzw. Homogenität der zu analysierenden Proben, der verwendeten Read-Länge und -art (Paired-end- bzw. Single-end-Sequenzierung) und nicht zuletzt auch von der verwendeten Technologie ab. Für die SOLID™-5500XL-Plattform empfiehlt der Hersteller eine Menge von 50 Mio. eindeutig in der Referenzsequenz zuordenbarer Reads. Eine klare Aussage darüber, wie viele solcher Reads vonnöten sind, um tatsächlich alle in einer Probe vorhandenen Transkripte zu

verlässig zu erfassen, lässt sich jedoch derzeit nicht treffen.

Man kann mittels RNA-Sequenzierung auch kleine RNA-Moleküle, wie z. B. microRNA-Moleküle (miRNA), untersuchen. Dabei handelt es sich um eine Klasse von etwa 22 Nukleotide umfassenden nichtkodierenden RNA-Molekülen. Sie spielen bei der Feinregulation der Genexpression eine Rolle. Der Wirkmechanismus beruht dabei auf der Sequenzkomplementarität der miRNA-Moleküle mit ihren jeweiligen Zieltranskripten: In Abhängigkeit vom Grad der Übereinstimmung zwischen miRNA-Sequenz und der Basenabfolge in der entsprechenden Bindungsstelle kommt es entweder zum Abbau der betroffenen mRNA oder einer Hemmung der Translation. Für die Sequenzierung kleiner RNA-Moleküle ist eine spezielle Form der Probenaufbereitung erforderlich, wofür es jedoch auch kommerziell erhältliche Lösungen gibt. Weiterführende Informationen zu Funktion und Wirkungsweise von miRNA-Moleküle sowie zu Besonderheiten bei der Probenaufbereitung für die Sequenzierung kleiner RNA-Moleküle finden sich z. B. in der Übersichtsarbeit von Gurtan und Sharp [13] bzw. bei [15] oder [38].

„Ribosome profiling“

Beim „ribosome profiling“ handelt es sich um eine Methode zur gezielten Sequenzierung von mRNA-Molekülen, die zu einem bestimmten Zeitpunkt im Rahmen der Translation an die Ribosomen gebunden sind [18]. Dies kann von Bedeutung sein, wenn man beispielsweise Transkriptionsstartpunkte untersuchen oder die Syntheseraten von Proteinen bestimmen möchte. „Ribosome profiling“ wurde z. B. in der Grundlagenforschung bereits erfolgreich verwendet, um Genorte zu identifizieren, an denen mindestens 2 translatierte offene Leserahmen vorliegen, bzw. um bisher unbekannte proteinkodierende Sequenzen zu erkennen [23]. Ein anderes Beispiel zeigt die Nützlichkeit der Methode bei der Untersuchung von langen nichtkodierenden intergenisch lokalisierten RNA-Molekülen, die ebenfalls in ribosomalen Komplexen gefunden, jedoch nicht in Proteine translatiert werden [14].

medgen 2014 · 26:264–272
DOI 10.1007/s11825-014-0449-5
© Springer-Verlag Berlin Heidelberg 2014

A.W. Kuß

Das methodische Potenzial der neuen Sequenzier-technologien jenseits der Mutationsuche

Zusammenfassung

In diesem Beitrag wird eine Reihe wichtiger Anwendungen der neuen Sequenzier-technologien bzw. des Next Generation Sequencing (NGS) vorgestellt. An ausgewählten Beispielen werden für jede Methode die Anwendungsmöglichkeiten in der humangenetischen Forschung dargestellt, jeweils das prinzipielle Vorgehen beschrieben und mögliche Quellen für ausführliche Arbeitsanweisungen vorgestellt. Die beschriebenen Techniken umfassen im Einzelnen: RNA-Sequenzierung mittels NGS („RNA-Seq“), Chromatinimmunpräzipitation in Kombination mit NGS („ChIP-Seq“), „ribosome profiling“, Immunpräzipitation methylierter DNA-Segmente in Kombination mit NGS („methylated DNA immunoprecipitation“ bzw. „MeDIP-Seq“) und die HiC-Technik, eine Weiterentwicklung der Chromosome-Conformation-Capture(3C)-Methode.

Schlüsselwörter

Chromatin · Molekulare Konformation · Immunpräzipitation · Hochdurchsatznukleotidsequenzierung · Bioinformatik

Methodological potential of new sequencing technologies beyond the search for mutations

Abstract

In this article, a number of important applications for next generation sequencing (NGS)-based techniques are presented. For each method the technical principles are introduced, the application options in human genetic research using selected examples are illustrated and possible sources for detailed protocols are indicated. The following methods are described: RNA sequencing using NGS (RNA-seq), chromatin immunoprecipitation in combination with NGS (ChIP-seq), ribosome profiling, methylated DNA immunoprecipitation in combination with NGS (MeDIP-seq) and the HiC technique, an extension of the chromosome confirmation capture (3C) method.

Keywords

Chromatin · Molecular conformation · Immunoprecipitation · High-throughput nucleotide sequencing · Bioinformatics

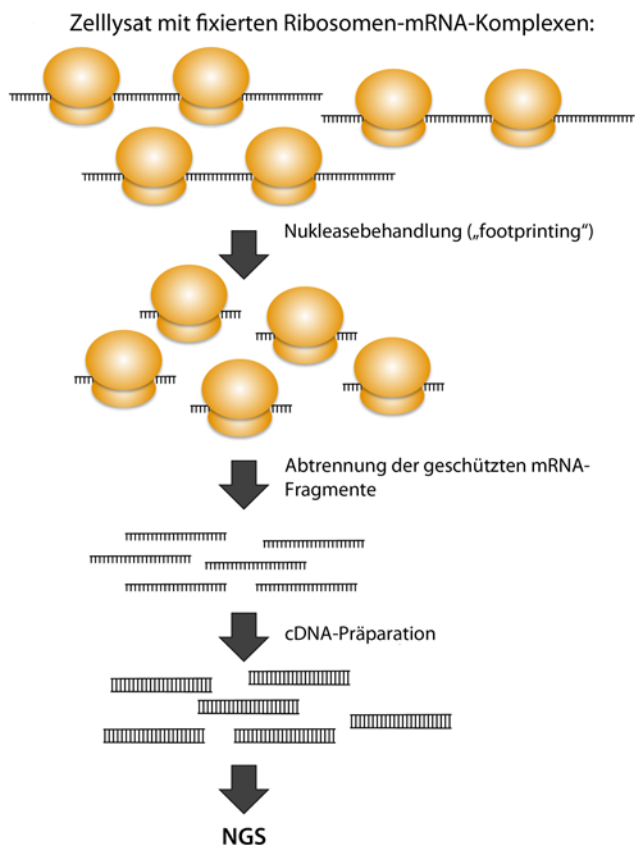


Abb. 3 ▲ Prinzipielles Vorgehen beim „ribosome profiling“. Nach Fixierung der in der Translation begriffenen Ribosomen und Lyse der Zellen wird das Lysat mit Nucleasen behandelt. Dadurch werden alle Bereiche der RNA, die nicht durch die Komplexierung mit Ribosomen geschützt sind, zerschnitten („footprinting“). Es bleiben nur solche Sequenzen intakt, die an Ribosomen gebunden waren. Nach Isolierung der Poly- und Monosomen (z. B. durch Ultrazentrifugation) und Abtrennung der „geschützten“ mRNA-Fragmente wird cDNA hergestellt, die die Grundlage für eine NGS-Sequenzierbibliothek bildet. NGS Next Generation Sequencing

Darüber hinaus stellt die NGS-Technologie mit dieser Methode ein weiteres interessantes Instrument für die Aufklärung von Pathomechanismen genetischer Erkrankungen zur Verfügung. So konnte z. B. mithilfe dieser Methode für eine Patientin mit einer Mutation in der mitochondrialen tRNA für Tryptophan gezeigt werden, dass die aller Wahrscheinlichkeit nach krankheitsverursachende tRNA-Veränderung zu einer eingeschränkten Ribosomenaktivität führt [32]. Dies hatte möglicherweise die, bei der im Alter von 13 Monaten verstorbenen Patientin beobachtete, Oxidative-phosphorylation(OXPHOS)-Symptomatik [35] zur Folge.

Methodisch geht man beim „ribosome profiling“ i. d. R. folgendermaßen vor: Zunächst muss eine Behandlung der Zel-

len erfolgen, die dazu führt, dass die in der Translation begriffenen Ribosomen auf der RNA fixiert werden. Dies wird z. B. durch eine Behandlung mit Cycloheximid erreicht. Anschließend werden die Zellen lysiert und das Lysat mit Nucleasen behandelt. Dadurch werden alle Bereiche der RNA, die nicht durch die Komplexierung mit Ribosomen geschützt sind, zerschnitten und man erreicht ein „footprinting“. Dabei bleiben nur solche Sequenzen intakt, die an Ribosomen gebunden waren. Nach Isolierung der Poly- und Monosomen (z. B. durch Ultrazentrifugation) und Abtrennung der „geschützten“ mRNA-Fragmente wird cDNA hergestellt, die die Grundlage für eine NGS-Sequenzierbibliothek bildet. Eine grob schematische Darstellung des Vorgehens ist in

■ **Abb. 3** dargestellt. Eine sehr detaillierte Arbeitsanweisung für diese Methode wurde z. B. im Jahr 2013 von Nicholas Ingolia und Kollegen veröffentlicht [17]. Bei dieser Methode sind ebenfalls die bereits zuvor erwähnten Einschränkungen bei der RNA-Sequenzierung zu beachten und im Rahmen der bioinformatischen Auswertung zu berücksichtigen. Auch hier lässt sich bezüglich der benötigten Anzahl an Reads keine allgemeingültige Aussage treffen. Es ist jedoch davon auszugehen, dass eine möglichst große Read-Zahl die Auswertung begünstigt [17].

ChIP-Seq

ChIP-Seq (Chromatinimmunpräzipitation gefolgt von NGS) ist eine Strategie, die die Untersuchung der Interaktionen zwischen DNA-bindenden Proteinen und DNA ermöglicht. Dabei geht es insbesondere um die Identifizierung spezifischer Erkennungssequenzen und in deren Nähe liegender Zielgene solcher Proteine. Aber auch für die funktionelle Einordnung von Transkriptionsfaktoren bei genetisch bedingten Erkrankungen kann diese Technik herangezogen werden. So konnte beispielsweise mittels eines ChIP-Seq-basierten Versuchsansatzes gezeigt werden, dass NRF1 („nuclear respiratory factor 1“) diverse Zielgene reguliert, die im Zusammenhang mit den Ätiologien verschiedener neurodegenerativer Erkrankungen stehen, wie etwa M. Alzheimer oder Chorea Huntington [33]. NRF1 ist als Transkriptionsfaktor charakterisiert, der die Expression einer Reihe von Genen steuert, die jeweils essenziell für die mitochondriale Biogenese und Funktion sind. Weitere Beispiele aus dem Bereich der Krebsforschung zeigen die Anwendung von ChIP-Seq bei der Untersuchung von tumorspezifischen Transkriptionsfaktormutationen [1, 39].

Methodisch handelt es sich bei der ChIP-Seq-Technik um die Kombination der klassischen Chromatinimmunpräzipitation (ChIP) mit NGS. Für ChIP müssen zunächst die zu untersuchenden Zellen mit Formaldehyd fixiert werden, wodurch es auch zu einer Quervernetzung zwischen der DNA und daran gebundenen Proteinen kommt. Danach werden die so behandelten Zellen lysiert, eine DNA-Iso-

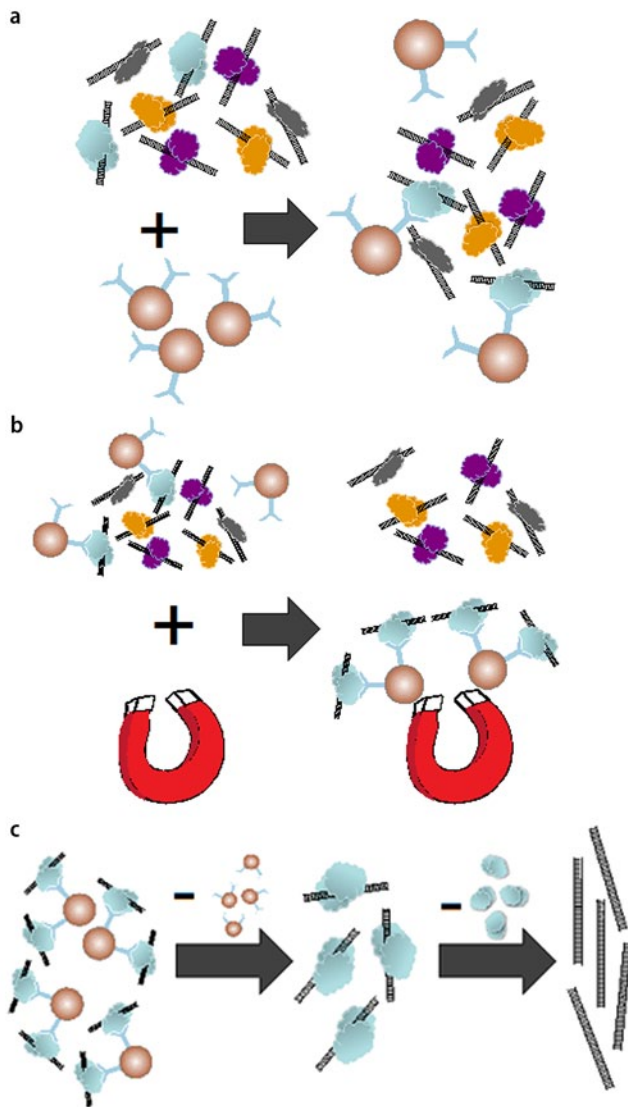


Abb. 4 ▲ Vorbereitung der ChIP-Seq. **a** Fragmentiertes Chromatin aus einem Lysat von mit Formaldehyd fixierten Zellen wird mit magnetischen Mikrosphären, die mit Antikörpern gegen das DNA-bindende Zielprotein bestückt sind, zusammengebracht. **b** Isolierung der Antikörper-Protein-DNA-Komplexe über einen Magneten. **c** Entfernung der Mikrosphären, Auflösung der isolierten Protein-DNA-Komplexe und Isolierung der gebundenen DNA-Sequenzen. Diese stehen dann für die Herstellung einer Next-Generation-Sequencing-Sequenzierbibliothek zur Verfügung

lierung durchgeführt und das Chromatin z. B. durch Ultraschall fragmentiert. Im Anschluss können wieder mithilfe magnetischer Mikrosphären, die in diesem Fall mit Antikörpern gegen das jeweilige Zielprotein bestückt sind (■ **Abb. 4a**), die Protein-DNA-Komplexe isoliert werden (siehe z. B. [34]). Dies geschieht im nächsten Schritt über einen Magneten (■ **Abb. 4b**). Danach werden die Protein-DNA-Komplexe aufgelöst, um schließlich aus den so isolierten und spezifisch

an das Zielprotein bindenden DNA-Sequenzen (■ **Abb. 4c**) eine Sequenzierbibliothek herzustellen, die nachfolgend für NGS verwendet werden kann.

Die dabei erzielten Einzelsequenzen zeigen nach Alignieren mit der Referenzsequenz Anhäufungen an denjenigen Stellen im Genom, an denen das untersuchte Protein in der Zelle zuvor gebunden hatte.

MeDIP-Seq

MeDIP-Seq („methylated DNA immunoprecipitation“ gefolgt von NGS) ist eine Spielart der ChIP-Seq-Methode, die das Studium des Methylierungsgrads bzw. -musters auf der DNA gestattet. Dies ist ein wichtiges Instrument für die Bearbeitung epigenetischer Fragestellungen, bei denen es darum geht, die Verteilung und Häufigkeit des Vorkommens von methylierten Cytosinbausteinen an der 5'-Position des Kohlenstoffatoms in der DNA zu untersuchen.

Das methylierte Cytosin ist meist Bestandteil von im DNA-Strang paarweise aufeinanderfolgenden Cytosin und Guaninbausteinen (CpG). Oft werden Anhäufungen von CpG, sog. CpG-Inseln, in den Promotorbereichen von Genen beobachtet, wo sie eine wichtige Rolle bei der Genregulation oder beim genomischen Imprinting spielen.

Als Beispiel für den erfolgreichen Einsatz von MeDIP-Seq kann u. a. die Arbeit von Bell et al. [4] dienen, die zeigen konnten, dass es speziesspezifische nichtpolymorphe CpG-Nukleotidpaare im Genom, sog. „CpG beacons“, gibt. Beim Menschen wurde dabei beobachtet, dass Anhäufungsbereiche solcher speziesspezifischen „CpG beacons“ wiederholt mit Genorten zusammenfallen, die bei verschiedenen erblichen neurologischen Störungen und hereditären Formen der geistigen Behinderung krankheitsverursachende Mutationen tragen. Darüber hinaus waren auch Loci betroffen, die mit komplexen Erkrankungen assoziiert sind. Insgesamt schlussfolgern die Autoren, dass „CpG beacons“ epigenetische Charakteristika darstellen, die im engen Zusammenhang mit der Evolution von CpG-Inseln im humanen Genom stehen und mit der Ausprägung menschlicher Merkmale bzw. spezifischer Erkrankungen assoziiert sind. So wurden differenziell methylierte genomische Regionen auch bei Schizophrenie und bipolaren Störungen mittels MeDIP-Seq gefunden [40].

Im Einzelnen geht man beim MeDIP-Seq-Verfahren analog der ChIP-Seq-Methode vor: Es wird zunächst DNA extrahiert, fragmentiert und danach die doppelsträngige DNA denaturiert. Anschließend wird über einen 5-Methyl-Cyti-

Hier steht eine Anzeige.



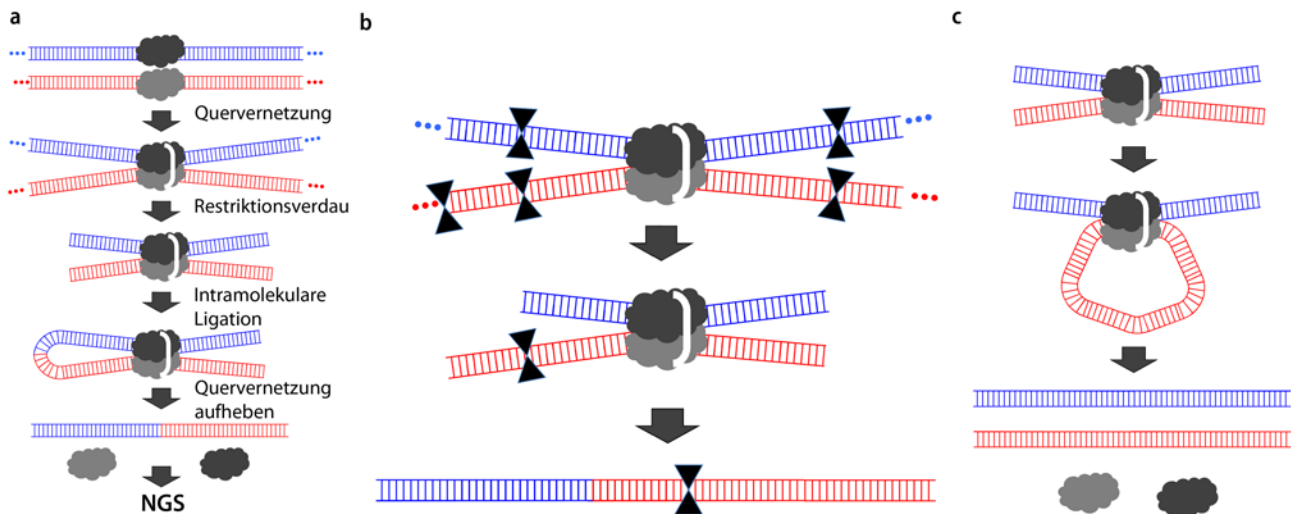


Abb. 5 ▲ Übersicht zur HiC-Methode und möglichen Fehlerquellen. **a** Prinzipielles Vorgehen bei HiC. Zunächst wird eine Fixierung der zu untersuchenden Zellen mit Formaldehyd vorgenommen. Dabei kommt es zur Ausbildung kovalenter Bindungen zwischen nebeneinanderliegenden Proteinen aus benachbart angeordneten Chromatinabschnitten (Quervernetzung). Mit daraus isolierter quervernetzter DNA wird dann ein Restriktionsverdau durchgeführt, bei dem einzelsträngige Überhänge an den DNA-Fragmenten entstehen („sticky ends“). Diese Überhänge werden durch eine DNA-Polymerase mit biotinylierten Nucleotiden aufgefüllt. Die Biotinylierung dient der späteren Aufreinigung der Fragmente mittels streptavidintragender magnetischer Mikrosphären. Dadurch entstehen überhangslose sog. Blunt-end-Fragmente. Zwischen den Enden der Fragmente von den verschiedenen DNA-Strängen wird nachfolgend eine (intramolekulare) Ligation herbeigeführt. Im Anschluss werden die ursprüngliche Quervernetzung der Proteine untereinander sowie deren Bindung an die DNA aufgehoben; die frei werdende DNA kann anschließend für NGS verwendet werden. **b** Fehlerquelle unvollständige Restriktion. Bei der Restriktionsreaktion kann es dazu kommen, dass nicht immer die der Quervernetzungsstelle nächstgelegene Schnittstelle benutzt wird. Dadurch können unterschiedliche Fragmente entstehen, die jedoch zum selben Berührungspunkt der verschiedenen DNA-Stränge gehören. **c** Fehlerquelle Ligation Cis-liegender DNA-Moleküle. Es lässt sich nie ganz verhindern, dass auch die Enden desselben DNA-Abschnitts miteinander ligieren. Hierdurch entstehen Fragmente, die nicht für die Ermittlung benachbart liegender Chromatinabschnitte geeignet sind. NGS Next Generation Sequencing

Antikörper die Immunpräzipitation durchgeführt. Die so gewonnenen DNA-Fragmente bilden dann die Grundlage für eine NGS-Sequenzierbibliothek. Diese Experimente führt man vergleichend durch, indem z. B. die MedIP mit Patientenzellen und mit entsprechenden Kontrollen parallel durchgeführt werden, um dadurch möglicherweise krankheitsrelevante bzw. -spezifische Unterschiede bezüglich Häufigkeit und Methylierungsgrad genomischer Positionen zu identifizieren.

Eine detaillierte Arbeitsanweisung für MedIP-Seq wurde beispielsweise von Oluwatosin Taiwo und Kollegen publiziert [36].

HiC

Bei der HiC-Technik handelt es sich um eine weitere Anwendung der NGS-Technologie, die es in diesem Falle erlaubt, die räumliche Konformation der Chromosomen bzw. des gesamten Genoms, wie

es im Zellkern während der Interphase des Zellzyklus (Interphasenuklear) vorliegt, zu analysieren. HiC geht ursprünglich zurück auf die schon länger etablierte sog. Chromosome-conformation-capture(3c)-Methode [10], die mehrfach modifiziert und weiterentwickelt wurde und nun in Kombination mit NGS eine neue genomweite Dimension erreicht hat.

Das Studium der strukturellen Eigenschaften und räumlichen Organisation der Chromosomen während der Interphase des Zellzyklus spielt bei der Aufklärung der molekularen Grundlagen der DNA-Replikation, Genregulation, Positionseffektvariegation (PEV), DNA-Reparaturmechanismen und von Rekombinationsvorgängen eine wichtige Rolle.

Ein anschauliches Beispiel für die Bedeutung der Interaktion weit voneinander entfernt liegender Chromosomenabschnitte bei der Genregulation stellen die „long range regulators“ dar. Dabei handelt es sich um regulatorisch aktive Sequenzabschnitte, die sich in räumlich

großer Entfernung zu den von ihnen regulierten Genen auf anderen Chromosomenabschnitten oder gar anderen Chromosomen befinden. Man geht davon aus, dass diese Sequenzelemente erst durch die räumliche Konformation der Chromosomen im Interphasenuklear in die Nähe ihrer Zielgene gelangen müssen, um Einfluss auf deren Expressionseigenschaften nehmen zu können.

Solche „long range regulators“ spielen u. a. bei der Ätiologie von erblichen Fehlbildungen der Extremitäten eine Rolle. So sind z. B. sowohl für bestimmte Formen der Brachydaktylie [9] als auch der Polydaktylie [2] krankheitsverursachende Mutationen in Long-range-regulator-Elementen beschrieben worden. Diese Beobachtungen belegen, wie Untersuchungsansätze zur Aufklärung des Interaktionsmusters im Chromatin des Interphasenuklear auch für die Erforschung der genetischen Grundlagen von erblichen Erkrankungen von besonderem Interesse sein können und weisen damit ein

weiteres Mal darauf hin, wie durch NGS-basierte Methoden neue Untersuchungsmöglichkeiten in der humangenetischen Forschung eröffnet werden.

Eine ausführliche Arbeitsanweisung zur HiC-Methode wurde von van Berkum et al. [37] publiziert. Dazu steht unterstützendes Videomaterial zur Verfügung (<http://www.jove.com/video/1869>, Zugriffen: 23. Mai 2014). Kurz zusammengefasst werden folgende Arbeitsschritte durchgeführt, bevor die eigentliche NGS-Analyse stattfindet (■ **Abb. 5a**):

- Zunächst wird eine Fixierung der zu untersuchenden Zellen mit Formaldehyd vorgenommen. Dabei kommt es zur Ausbildung kovalenter Bindungen zwischen nebeneinanderliegenden Proteinen aus benachbart angeordneten Chromatinabschnitten.
- Anschließend wird ein Restriktionsverdau durchgeführt.
- Danach folgt eine Ligationsreaktion zwischen den durch Restriktion erzeugten DNA-Fragmenten der Protein-DNA-Komplexe.
- Als nächstes wird die ursprüngliche Quervernetzung der Proteine wieder aufgehoben. Die frei werdende DNA kann anschließend weiter verarbeitet werden.

Im Folgenden werden die zugrundeliegenden Prinzipien der einzelnen Arbeitsschritte näher erläutert.

Für den Restriktionsschritt wird je nach angestrebter Fragmentlänge ein Restriktionsenzym ausgewählt, für das es möglichst gleichmäßig über das Genom verteilte Schnittstellen gibt, wie beispielsweise Hind III [21]. Grundsätzlich sollte das Restriktionsenzym einen einzelsträngigen Endüberhang an den Enden der geschnittenen DNA-Doppelstränge („sticky end“) erzeugen. Dieser Überhang kann durch eine DNA-Polymerase mit biotinylierten Nukleotiden aufgefüllt werden. Dadurch entstehen sog. Blunt-end-Fragmente. Die Biotinylierung dient dabei der späteren Aufreinigung der entstandenen Fragmente mittels streptavidintragender magnetischer Mikrosphären.

Im nächsten Schritt wird sodann – gesteuert über die Reaktionsbedingungen – eine Ligation zwischen den nun überhangslosen doppelsträngigen Blunt-End-

Fragmenten der verschiedenen DNA-Stränge hergestellt. Danach kann die ursprünglich hergestellte proteinvermittelte Quervernetzung aufgehoben werden.

Es liegen nun lineare DNA-Fragmente vor, die aus Sequenzabschnitten unterschiedlicher genomischer bzw. chromosomaler Herkunft zusammengesetzt sind. Diese können über das zuvor eingebaute Biotin isoliert werden und als Grundlage für eine NGS-Sequenzierbibliothek dienen, die nachfolgend einer Paired-end-Sequenzierung unterworfen wird. Die dadurch erhaltenen Cluster aus Einzelsequenzen alignieren mit der Referenzsequenz an denjenigen Stellen, die höchstwahrscheinlich bei der Mehrzahl der untersuchten Zellen an intra- oder interchromosomalen Interaktionen im Interphasenuklear beteiligt sind.

Aufgrund der geschilderten speziellen Probenaufbereitung gibt es auch bei dieser Methode besondere Fehlerquellen, die bei der bioinformatischen Auswertung berücksichtigt werden müssen. So kann es je nach Reaktionsbedingungen dazu kommen, dass nicht immer die der Quervernetzungsstelle nächstgelegene Schnittstelle benutzt wird. Dadurch können unterschiedliche Fragmente entstehen, die aber zum selben Berührungspunkt der verschiedenen DNA-Stränge gehören (■ **Abb. 5b**).

Außerdem lässt sich nie ganz verhindern, dass auch die Enden vom selben DNA-Abschnitt miteinander ligieren, wodurch Fragmente entstehen, die nicht für die Ermittlung benachbart liegender Chromatinabschnitte geeignet sind (■ **Abb. 5c**).

Für HiC muss eine Paired-end-Sequenzierung durchgeführt werden. Die beiden zusammengehörigen Paired-end-Reads müssen dann jeweils separat entlang der Referenzsequenz zugeordnet („gemappt“) werden. Zur weiteren Auswertung werden nur solche Reads verwendet, die in der Nähe der Schnittstellen des verwendeten Restriktionsenzym zugeordnet wurden. Außerdem fasst man im Rahmen der Auswertung mehrere Restriktionsenzymstellen zusammen („binning“). Die Read-Zahlen in den dadurch erzeugten „Bins“ müssen daraufhin nach GC-Gehalt, Fragmentlänge und „mappability“ (Eindeutigkeit

der Zuordnung von Reads im Referenzgenom) normalisiert werden. Abschließend kann dann die statistische Signifikanz der gefundenen Interaktionen berechnet werden.

Fazit für die Praxis

- Die neuen Sequenziertechnologien bieten mit ihrem methodischen Potenzial vielfältige Anwendungsmöglichkeiten in der humangenetischen Forschung. Das vorhandene Methodenspektrum wird v. a. mit Blick auf die Aufklärung von Pathomechanismen und die Analyse funktioneller Charakteristika der Transkriptions- und Translationsprodukte von Genen, die krankheitsverursachende Mutationen tragen, im Bereich der Sequenzanalytik deutlich erweitert.
- Der wichtigste Fortschritt liegt insbesondere in der Erschließung der gesamtexomischen und -transkriptomischen bzw. genomischen Dimension für die Bearbeitung der vorgeannten Fragestellungen. Wie bei allen NGS-Techniken ist jedoch auch bei den hier vorgestellten Methoden besonders zu beachten, dass für ihre erfolgreiche Anwendung bioinformatische Expertise im Umgang mit den jeweils anfallenden NGS-Daten unerlässlich ist.

Korrespondenzadresse

A.W. Kuß

Institut für Humangenetik, Universitätsmedizin Greifswald, und Interfakultäres Institut für Genetik und Funktionelle Genomforschung, Universität Greifswald
Fleischmannstr. 42–44, 17475 Greifswald
kussa@uni-greifswald.de

Danksagung. Der Autor dankt Prof. U. Felbor für konstruktive Kommentare.

Einhaltung ethischer Richtlinien

Interessenkonflikt. A. W. Kuß gibt an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren.

Literatur

1. Adomas AB, Grimm SA, Malone C et al (2014) Breast tumor specific mutation in GATA3 affects physiological mechanisms regulating transcription factor turnover. *BMC Cancer* 14:278
2. Albuissou J, Isidor B, Giraud M et al (2011) Identification of two novel mutations in Shh long-range regulator associated with familial pre-axial polydactyly. *Clin Genet* 79:371–377
3. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106
4. Bell CG, Wilson GA, Butcher LM et al (2012) Human-specific CpG „beacons“ identify loci associated with human-specific traits and disease. *Epigenetics* 7:1188–1199
5. Boycott KM, Vanstone MR, Bulman DE et al (2013) Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 14:681–691
6. Bullard JH, Purdom E, Hansen KD et al (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94
7. Cho I, Blaser MJ (2012) The human microbiome: at the interface of health and disease. *Nat Rev Genet* 13:260–270
8. Cui P, Lin Q, Ding F et al (2010) A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* 96:259–265
9. Dathe K, Kjaer KW, Brehm A et al (2009) Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am J Hum Genet* 84:483–492
10. Dekker J, Rippe K, Dekker M et al (2002) Capturing chromosome conformation. *Science* 295:1306–1311
11. Edgren H, Murumagi A, Kangaspeka S et al (2011) Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol* 12:R6
12. Fu GK, Xu W, Wilhelmy J et al (2014) Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. *Proc Natl Acad Sci U S A* 111:1891–1896
13. Gurtan AM, Sharp PA (2013) The role of miRNAs in regulating gene expression networks. *J Mol Biol* 425:3582–3600
14. Guttman M, Russell P, Ingolia NT et al (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154:240–251
15. Hafner M, Renwick N, Farazi TA et al (2012) Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods* 58:164–170
16. Hansen KD, Brenner SE, Dudoit S (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res* 38:e131
17. Ingolia NT, Brar GA, Rouskin S et al (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 7:1534–1550
18. Ingolia NT, Brar GA, Rouskin S et al (2013) Genome-wide annotation and quantitation of translation by ribosome profiling. *Curr Protoc Mol Biol* Chapter 4:Unit 4 18
19. Jensen LR, Bartenschlager H, Rujirabanjerd S et al (2010) A distinctive gene expression fingerprint in mentally retarded male patients reflects disease-causing defects in the histone demethylase KDM5C. *Pathogenetics* 3:2
20. Li J, Jiang H, Wong WH (2010) Modeling non-uniformity in short-read rates in RNA-Seq data. *Genome Biol* 11:R50
21. Lieberman-Aiden E, Berkum NL van, Williams L et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293
22. Markowitz VM, Korzeniewski F, Palaniappan K et al (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34:D344–D348
23. Michel AM, Choudhury KR, Firth AE et al (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res* 22:2219–2229
24. Montgomery SB, Sammeth M, Gutierrez-Arcelus M et al (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464:773–777
25. Morlan JD, Qu K, Sinicropi DV (2012) Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One* 7:e42882
26. Nakagawa H (2013) Prostate cancer genomics by high-throughput technologies: genome-wide association study and sequencing analysis. *Endocr Relat Cancer* 20:R171–R181
27. Norton N, Sun Z, Asmann YW et al (2013) Gene expression, single nucleotide variant and fusion transcript discovery in archival material from breast tumors. *PLoS One* 8:e81925
28. Parker BC, Engels M, Annala M et al (2014) Emergence of FGFR family gene fusions as therapeutic targets in a wide spectrum of solid tumours. *J Pathol* 232:4–15
29. Pickrell JK, Marioni JC, Pai AA et al (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464:768–772
30. Quinn EM, Cormican P, Kenny EM et al (2013) Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 genomes data. *PLoS One* 8:e58815
31. Roberts A, Trapnell C, Donaghey J et al (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12:R22
32. Rooijers K, Loayza-Puch F, Nijtmans LG et al (2013) Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat Commun* 4:2886
33. Satoh J, Kawana N, Yamamoto Y (2013) Pathway analysis of ChIP-Seq-Based NRF1 target genes suggests a logical hypothesis of their involvement in the pathogenesis of neurodegenerative diseases. *Gene Regul Syst Bio* 7:139–152
34. Siegel TN, Hekstra DR, Kemp LE et al (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* 23:1063–1076
35. Smits P, Mattijssen S, Morava E et al (2010) Functional consequences of mitochondrial tRNA Trp and tRNA Arg mutations causing combined OXPHOS defects. *Eur J Hum Genet* 18:324–329
36. Taiwo O, Wilson GA, Morris T et al (2012) Methylation analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc* 7:617–636
37. Berkum NL van, Lieberman-Aiden E, Williams L et al (2010) Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp* 39:1869
38. Dijk EL van, Jaszczyszyn Y, Thermes C (2014) Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res* 322:12–20
39. Watanabe K, Biesinger J, Salmans ML et al (2014) Integrative ChIP-seq/microarray analysis identifies a CTNNB1 target signature enriched in intestinal stem cells and colon cancer. *PLoS One* 9:e92317
40. Xiao Y, Camarillo C, Ping Y et al (2014) The DNA methylome and transcriptome of different brain regions in schizophrenia and bipolar disorder. *PLoS One* 9:e95875