



Michael Zeschnigk · Bernhard Horsthemke

Institut für Humangenetik, Universität Duisburg-Essen und Universitätsklinikum Essen, Essen, Deutschland

Next-Generation-Sequencing in der Epigenetik

Einleitung

Das menschliche Genom enthält ca. 20.000 Gene, von denen viele – mit Ausnahme von konstitutionell exprimierten Genen – nur in bestimmten Entwicklungsstufen, in bestimmten Zelltypen und zu bestimmten Zeiten aktiv sind. Das koordinierte An- und Ausschalten von Genen erfolgt durch generegulatorische Netzwerke, oft angestoßen durch extrazelluläre Signale. Dabei bestimmen Transkriptionsfaktoren einerseits die Aktivität und lokalen Chromatinzustände ihrer Zielgene, andererseits beeinflussen die Chromatinzustände die Kinetik der Bindung von Transkriptionsfaktoren an ihre Bindestellen.

Chromatinzustände sind hauptsächlich charakterisiert durch das Methylierungsmuster der DNA, die posttranskriptionale Modifikation von Histonen sowie die relative Position von Nukleosomen (Abb. 1a). Die DNA-Methylierung betrifft fast ausschließlich die Methylierung der Base Cytosin in CpG-Dinukleotiden. Histone können an verschiedenen Aminosäureseitenketten acetyliert, methyliert oder anderweitig modifiziert sein. Nukleosomenfreie DNA-Abschnitte erleichtern die Bindung von Transkriptionsfaktoren und der RNA-Polymerase an die DNA. Schließlich können Enhancer/Transkriptionsfaktorkomplexe über eine große Distanz hinweg mit Promotoren interagieren, und zwar durch Schleifenbildung des Chromatins (Abb. 1b); d.h., auch die 3-D-Struktur des Genoms spielt eine Rolle bei der Regulation der Genexpression. Diese epigenetischen Veränderungen sind in der Regel mitochondrial stabil, können aber von Transkriptionsfaktoren durch die Rekrutierung von DNA-Methyltransferasen und -De-

methylasen, von histonmodifizierenden Enzymen sowie von Nukleosom-Remodelling-Komplexen verändert werden. Die Gesamtheit aller epigenetischen Veränderungen (das Epigenom) spiegelt die Identität von Zelltypen in normalen, altersbedingten und krankheitsbedingten Situationen wider.

Aberrante epigenetische Muster an spezifischen Genorten werden Epimutation genannt [12], wobei zwischen primären und sekundären Epimutationen unterschieden wird [13]. Sekundäre Epimutationen sind Folge einer genetischen Mutation am oder in der Nähe des betroffenen Genorts. Ein Beispiel für eine sekundäre Epimutation ist die Methylierung des *MSH2*-Promoters infolge einer Deletion des 3'-Endes des benachbarten *EPCAM*-Gens in einigen Familien mit Lynch-Syndrom. Die Deletion führt zu einer *EPCAM*-Transkription durch den *MSH2*-Promoter und so zu dessen De-novo-Methylierung [20]. Weiterhin kann eine sekundäre Epimutation die Folge einer Mutation in einem epigenetischen Regulator (z.B. DNA-Methyltransferasen oder histonmodifizierende Enzyme) sein [17], wodurch in der Regel sehr viele Loci betroffen sind. Sekundäre Epimutationen sind sehr viel häufiger als primäre, die ohne eine genetische Veränderung aufgetreten sind. Klassisches Beispiel für eine primäre Epimutation sind Imprintingfehler, bei denen eine aberrante DNA-Methylierung vorliegt, aber eine Mutation im Imprintingzentrum ausgeschlossen wurde [2]. Da fast alle Epimutationen postzygotisch entstehen, liegen sie oft im Mosaik vor, was ihren Nachweis erschwert. Es ist umstritten, inwieweit Blut verwendet werden kann, um Epimutationen an Genen zu identifizieren, die nur in anderen Gewe-

ben, z.B. im Gehirn, exprimiert werden. Krebszellen weisen in der Regel eine globale DNA-Hypomethylierung [8] und eine lokale Hypermethylierung (Methylierung von Tumorsuppressorgenen) auf [9]. Bislang wurde bei Krankheitsgeschehen und Alterungsprozessen nur nach aberranten DNA-Methylierungsmustern gesucht. Im Prinzip sind aber auch aberrante Histonmodifikationen denkbar, insbesondere die Trimethylierung von Histon H3 an Lysin 27 (H3K27me3) betreffend (siehe auch unten). Diese Markierung charakterisiert Polycomb-reprimiertes Chromatin, das ähnlich stabil wie die DNA-Methylierung ist. Diese Möglichkeit wurde bislang beim Menschen nicht erforscht.

Die Analyse des gesamten Epigenoms ist erst mit der Entwicklung von Hochdurchsatz-Sequenziertechniken und spezieller Auswertealgorithmen möglich geworden. Durch die Anstrengungen vieler Arbeitsgruppen im Rahmen des International Human Epigenome Consortiums (IHEC) und anderer Konsortien stehen inzwischen große Datensätze von vielen menschlichen Geweben und Zellen zur Verfügung. Leider wurden und werden allzu oft nicht genau spezifizierte Mischungen verschiedener Zelltypen (Blut oder andere Gewebe) untersucht. Eine unterschiedliche Zellzusammensetzung verschiedener Gewebsproben kann bei vergleichenden Analysen epigenetische Unterschiede vortäuschen, die nicht existieren. Zur Vermeidung dieses Problems sollte eine möglichst homogene Zellpopulation untersucht werden, z.B. Monozyten, B-Zellen oder T-Zellen statt Vollblut. Außerdem wird oft übersehen, dass genetische Variation einen großen Einfluss auf die epigenetische Variation hat [23]. Zum Ausschluss genetisch be-

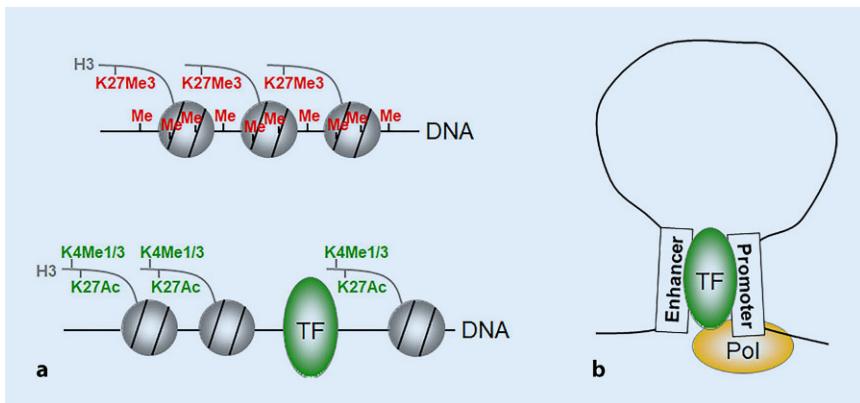


Abb. 1 ▲ Epigenetische Markierungen.

a Jeweils 146 bp DNA (schwarze Linien) sind um ein Histon-Oktamer gewickelt (Nukleosomen; graue Kugeln), die eng benachbart (oben) oder weiter auseinander (unten) liegen können, sodass Transkriptionsfaktoren (TF; grüne Ellipse) an ihre Zielsequenzen binden können. Nur sogenannte Pionierfaktoren können an geschlossenem Chromatin binden und dessen Öffnung mithilfe von Nukleosom-Remodellern bewerkstelligen. CpG-Dinukleotide sind entweder methyliert (Me) oder unmethyliert. Histonseitenketten (graue Linien; hier nur für H3 gezeigt) können an verschiedenen Aminosäuren, wie z.B. Lysin4 (K4) oder Lysin27 (K27), mono-, di- oder trimethyliert (hier me1 und me3 gezeigt) oder acetyliert sein (hier für K27ac gezeigt).

b Durch eine Schleifenbildung des Chromatins (hier der Einfachheit halber ohne Nukleosomen gezeigt) können Enhancer mit weit entfernt liegenden Promotoren interagieren.

Pol RNA-Polymerase

5-Methylcytosin in Uracil umgewandelt. Bei einer nachfolgenden Polymerasekettenreaktion (PCR) wird Thymin anstelle von Uracil in die DNA eingebaut. Zeigt die anschließende DNA-Sequenzierung im Vergleich zur Referenzsequenz ein T statt C, war das ursprüngliche Cytosin unmethyliert. Zeigt sie ein C, war das ursprüngliche Cytosin methyliert oder hydroxymethyliert. 5-Hydroxymethylcytosin, das bis auf embryonale Stammzellen und Neuronen oft nur in sehr geringer Konzentration vorliegt, wird deshalb in der Regel vernachlässigt. Es kann aber durch oxidative Bisulfitsequenzierung bestimmt werden. Bis sich bisulfitfreie Verfahren (enzymatische Methoden zur Unterscheidung von Cytosin, 5-Methylcytosin und 5-Hydroxymethylcytosin oder der direkte Nachweis modifizierter Basen während der Sequenzierung) durchgesetzt haben, bleibt die Bisulfitsequenzierung der Goldstandard für die basengenaue DNA-Methylierungsanalyse.

„Whole-genome bisulfite sequencing“

Als Goldstandard für genomweite Methylierungsstudien wird die Methode des „whole-genome bisulfite sequencing“ (WGBS) angesehen. Mit dieser Technik wird der Methylierungsstatus aller CpG-Dinukleotide (28.000.000 im menschlichen Genom) in Einzelnukleotidauflösung bestimmt. Je nach erzielter Sequenzertiefe (für epigenetische Karten >30X) erhält man ein sehr umfassendes und detailliertes Bild vom Methylierungsmuster der untersuchten DNA. Für viele Fragestellungen wie z.B. Klassifizierungen oder Vergleiche reichen oft auch niedrigere Sequenzertiefen (5-20X) aus.

Das ursprüngliche Verfahren [21] zur Herstellung der Bibliotheken für die genomweite Methylierungsanalyse beginnt mit einer Fragmentierung der DNA. Nach Ligation gerätespezifischer Adaptersequenzen erfolgt die Behandlung mit Natriumbisulfit („pre-bisulfite library preparation“), gefolgt von einer PCR. Um die Zahl der PCR-Schritte und somit den Anteil der Duplikate möglichst gering zu halten, muss zu Beginn

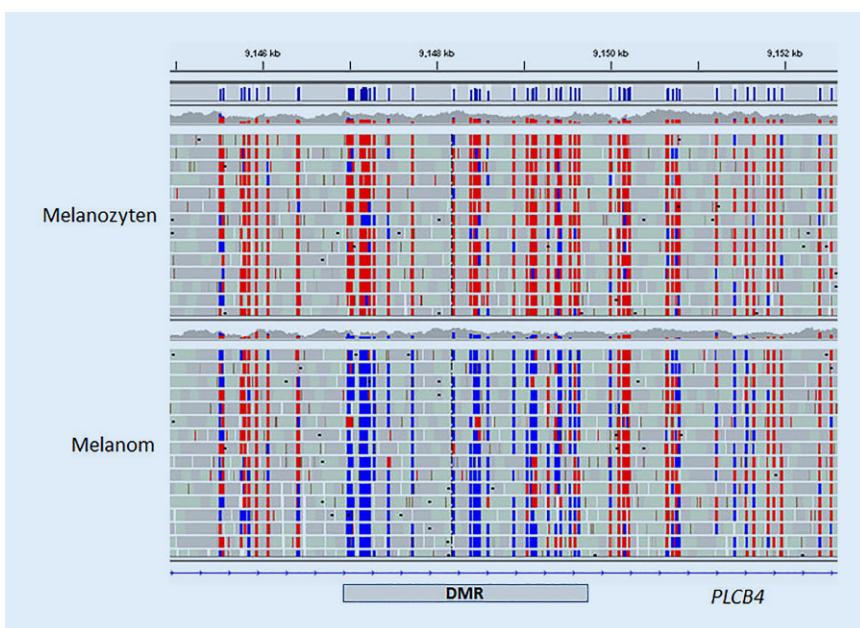


Abb. 2 ▲ Differenziell methylierte Region (DMR). Screenshot (Integrated Genomics Viewer, Broad Institute, Cambridge, MA, USA) einer Region im Intron des *PLCB4*-Gens, die zwischen Melanozyten und Melanomzellen unterschiedlich methyliert ist. Rot methylierte CpGs; blau unmethylierte CpGs

dingter epigenetischer Variation sollte deshalb angestrebt werden, isogene Vergleiche durchzuführen, z. B. ein Vergleich von verschiedenen Zelltypen ein und denselben Individuum oder ein Vergleich desselben Zelltyps eines Individuum vor und nach einer Intervention.

DNA-Methylierung

Methoden zur Untersuchung der DNA-Methylierung basieren meist auf der Behandlung von einzelsträngiger DNA mit Natriumbisulfit [4]. Durch diese Behandlung wird Cytosin, aber nicht

eine relativ hohe Menge genomischer DNA (ca. 2 µg) eingesetzt werden. Bei diesem Verfahren ist die Abdeckung von CpG-Inseln (CGIs) relativ schlecht. Eine Alternative ist die Erstellung einer „post-bisulfite library preparation“. Hier wird die DNA (50–100 ng) initial mit Natriumbisulfit behandelt. Die Adaptersequenzen werden dann mittels „random priming“ eingeführt. Diese Bibliotheken bevorzugen CGIs. Durch Zusammenführung von Daten aus „pre- und post-bisulfite“ Bibliotheken kann eine recht gleichmäßige Abdeckung des Genoms erreicht werden.

Neuere Verfahren zur Herstellung der Bibliotheken wie z. B. „Tagmentation“ (T-WGBS; [27]) kommen mit sehr wenig Ausgangsmaterial aus (ca. 10 ng DNA) und ergeben trotzdem eine relativ gleichmäßige Abdeckung des Genoms. Bei dieser Methode wird die genomische DNA durch eine Tn5-Transposase fragmentiert, und im gleichen Schritt werden die Sequenzierungssadapter angehängt. Die ersten Einzelzellmethylierungsanalysen [5] wurden unter Verwendung eines „scBS-seq library preparation“ Protokolls durchgeführt. Wegen der geringen Abdeckung (<0,3X) ist eine vergleichende Analyse nur bedingt möglich.

Für die Datenanalyse müssen die Sequenzen, ähnlich wie bei der Genomsequenzierung, auf das Referenzgenom kartiert werden. Für die Kartierung von Bisulfitsequenzen, die Markierung von Duplikaten, die Errechnung der Abdeckung sowie das Auszählen der „reads“, die Methylierung bzw. Nichtmethylierung eines jeden CpGs anzeigen, gibt es etablierte bioinformatische Datenanalyseprozesse (z. B. [25]). „MethylSeekR“ kann benutzt werden, um das Genom in niedrig, partiell und hoch methylierte Regionen zu segmentieren [3]. Der Methylierungsgrad von CpGs in ausgewählten genomischen Regionen kann mit Programmen wie dem Integrated Genomics Viewer (IGV) visualisiert werden (Abb. 2).

Bei einer vergleichenden Methylierungsanalyse werden die DNA-Methylierungsniveaus zwischen unterschiedlichen Proben (z. B. Fälle gegen Kontrollen) bestimmt. In den letzten Jahren

medgen 2019 · 31:205–211 <https://doi.org/10.1007/s11825-019-0245-3>
 © Der/die Autor(en) 2019

M. Zeschnigk · B. Horsthemke

Next-Generation-Sequencing in der Epigenetik

Zusammenfassung

Molekulare Epigenetik bezieht sich auf die chemische und physikalische Modifikation von DNA und Chromatin. Transkriptionsfaktoren bestimmen die Chromatinzustände ihrer Zielgene, während Chromatinzustände andererseits die Bindung von Transkriptionsfaktoren an die DNA beeinflussen. Chromatinzustände sind gekennzeichnet durch spezifische Muster der DNA-Methylierung, Histonmodifikation, den Abstand der Nukleosomen und die 3-D-Faltung von Chromatin. Seit dem Aufkommen des Next-Generation-Sequencing (NGS) ist

es möglich, diese Muster im genomweiten Maßstab mit einer beispiellosen Auflösung zu analysieren. Solche epigenetischen Karten sind wertvoll, um regulatorische Elemente zu identifizieren, normale und gestörte Entwicklungsvorgänge zu verstehen und Epimutationen zu entdecken.

Schlüsselwörter

Epigenetik · DNA-Methylierung · Histone-Modifikation · Nukleosomen · 3-D-Chromatinfaltung

Next-generation sequencing in epigenetics

Abstract

Molecular epigenetics refers to the chemical and physical modification of DNA and chromatin. Transcription factors determine the chromatin states of their target genes, and chromatin states affect the binding of transcription factors to DNA. Chromatin states are characterized by specific patterns of DNA methylation, histone modification, the spacing of nucleosomes and the 3-D-folding of chromatin. Since the advent of next-generation sequencing (NGS), it has

become possible to analyze these patterns on a genome-wide scale and at unprecedented resolution. Such epigenetic maps are important for detecting regulatory elements, understanding normal and abnormal development and identifying epimutations.

Keywords

Epigenetics · DNA methylation · Histone modification · Nucleosomes · 3-D Chromatin folding

wurden hierfür mehrere Software-Werkzeuge entwickelt, z. B. BSmooth [11] und metilene [15]. Jede Methode setzt dabei bestimmte Eigenschaften der Daten voraus. Die Herausforderungen liegen in der zuverlässigen Schätzung des Methylierungsgrades an jedem CpG in jeder Probe bzw. pro Klasse bei in der Regel kleiner Probenzahl (oft nur 2 gegen 2) und geringer Sequenziertiefe.

Um plattformabhängige Verzerrungen bei der DNA-Methylierungsanalyse zu vermeiden, werden vergleichende Analysen am besten nur mit Datensätzen derselben Plattform durchgeführt. Wegen der begrenzten Teststärke (Power) solcher Vergleiche werden statt einzelner CpGs in der Regel differentiell methylierte Regionen (DMRs) betrachtet, und zwar nur solche mit mehr als 4 benachbarten, im Methylierungsgrad korrelierenden CpGs sowie einem mittleren Methylier-

rungsunterschied von >0,3. Mit einem solchen Vorgehen und nach unabhängiger Validierung sind bereits wichtige biologische Mechanismen aufgedeckt worden [26]. Durch diese Studien hat sich die früher vorherrschende Ansicht geändert, dass DMRs vorwiegend in CGIs anzutreffen sind. Insbesondere Methylierungsveränderungen im Rahmen einer normalen Zelldifferenzierung (mit Ausnahme von Unterschieden zwischen Körper- und Keimzellen) sind selten in CGIs, sondern meistens in Enhancern zu finden.

Mithilfe von WGBS lassen sich auch Epimutationen bei syndromalen Erkrankungen mit bekannten oder unbekannten Krankheitsgenen detektieren [1]. Da hier nichtisogene Vergleiche angestellt werden müssen, bleibt die genomweite Suche nach Epimutationen eine große Herausforderung.

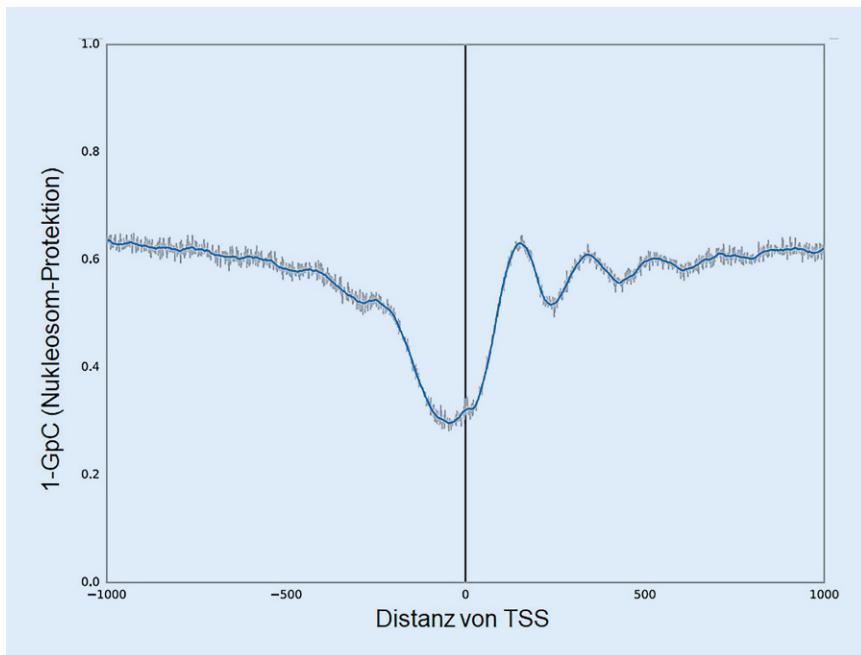


Abb. 3 ▲ Nukleosomenprofil von Monozyten. Aufgetragen ist der komplementäre mittlere Methylierungsgrad der GpC-Dinukleotide ($1-GpC$) aller Gene im Abstand von -1000 bis $+1000$ bp um ihre Transkriptionsstartstelle (TSS) innerhalb von CpG-Inseln. Peaks zeigen nukleosomengeschützte Bereiche an

„Deep bisulfite sequencing“

Eine gezielte Methylierungsanalyse wird häufig für die Validierung von DMRs genutzt, die mittels WGBS oder Arrays identifiziert wurden, oder für einen Kandidatengen-Ansatz, wenn eine Beteiligung von lokusspezifischen epigenetischen Veränderungen vermutet wird. Bei der tiefen Amplikong-Bisulfitsequenzierung („deep bisulfite sequencing“, DBS) wird die Zielregion nach Modifikation mit Natriumbisulfit unter Verwendung von lokusspezifischen Primern amplifiziert, die am 5'-Ende universelle „Tag“-Sequenzen besitzen. In einer 2. PCR werden über diese „Tags“ die gerätespezifischen Adaptersequenzen zusammen mit probenspezifischen Identifier-Sequenzen (ID) an das PCR-Produkt angefügt (detailliertes Protokoll siehe [18]). Durch Verwendung unterschiedlicher IDs an beiden Enden des Amplikons können in einem Lauf Hunderte unterschiedliche Proben parallel analysiert werden. Die hohe Sequenzierkapazität der NGS-Geräte erlaubt es zudem, die Amplikons sehr tief zu analysieren, d. h. mit weit mehr als 1000 „reads“ je Amplikon. Dies ergibt eine sehr detaillierte und quantita-

tive Darstellung des Methylierungsstatus jedes einzelnen CpGs. Für die komfortable Analyse der Sequenzdaten und die Visualisierung der Ergebnisse steht die Software amplikyzer2 zur Verfügung, die speziell für diese Anwendung geschrieben wurde [14]. Amplikyzer2 bestimmt die mittleren Methylierungswerte a) aller CpG-Dinukleotide in einer Region oder b) jedes CpG-Dinukleotids über alle „reads“ oder c) einen ausgewählten Satz von „reads“. Es ist auch möglich, Methylierungswerte zwischen verschiedenen Proben, aber auch zwischen verschiedenen Allelen innerhalb einer Probe zu vergleichen.

Nukleosomenlandschaft

NOMe-seq

Eine Variante von WGBS ist das „nucleosome occupancy and methylome sequencing“ (NOMe-seq) [16], mit dem sowohl die genomweite DNA-Methylierung als auch die genomweite Nukleosomenverteilung bestimmt werden kann. Hierbei wird Chromatin (nicht DNA) mit der viralen GpC-Methyltransferase M.CviPI behandelt, die Cytosinreste in

GpC-Dinukleotiden methyliert, allerdings nur in nukleosomenfreien DNA-Abschnitten. Anschließend wird die DNA isoliert und eine WGBS durchgeführt. Um die endogene DNA-Methylierung zu bestimmen, werden, wie oben beschrieben, CpG-Dinukleotide analysiert. Die Methylierung von GpC-Dinukleotiden hingegen zeigt nukleosomenfreie DNA-Bereiche an (GpC ist in Säugetieren nicht methyliert). Da bei GpCpG-Trinukleotiden nicht zwischen endogener und artifizieller Methylierung unterschieden werden kann, werden diese Nukleotide nicht betrachtet. NOMe-seq ist nicht so weit verbreitet wie andere Methoden zur Identifizierung nukleosomenfreier Bereiche (z. B. ATAC-seq s. unten), es ist aber die einzige Methode, mit der die Chromatinzugänglichkeit und die DNA-Methylierung in einem einzigen Experiment bestimmt werden können [10]. Das ist z. B. wesentlich für Einzelzellanalysen, die in zunehmendem Maße an Bedeutung gewinnen.

Analysen von NOMe-seq-Daten fokussieren auf Unterschiede innerhalb einer Probe und finden beispielsweise nukleosomenangereicherte und nukleosomenabgereicherte Regionen („nucleosome depleted regions“, NDRs). Hierzu gibt es Methoden und Implementierungen, etwa das R-Paket aaRon [24]. □ Abb. 3 zeigt die nukleosomenabgereicherte Region um die Transkriptionsstartstellen (TSS) innerhalb von CpG-Inseln humarer Monozyten. Rechts (3') der TSS sind deutlich die Position der Nukleosomen +1 bis +5 zu sehen, links etwas weniger deutlich die Nukleosome -1, bis -4. Die gezielte Analyse von ausgewählten Regionen, ähnlich zum tiefen Amplikong-Bisulfitsequenzieren, ist auch mit NOMe-seq möglich. Die amplikyzer2 Software (s. oben) bietet auch hierfür Analysemöglichkeiten.

Bislang gibt es kein Werkzeug zur genomweiten Identifizierung differentiell nukleosomenfreier Regionen (DNRs) aus NOMe-seq-Daten zwischen verschiedenen Proben. Die DNR-Entdeckung ist aus mehreren Gründen schwieriger als die DMR-Entdeckung: Es gibt viermal mehr GpCs als CpGs im menschlichen Genom (122 Mio. vs. 28 Mio.), und die Effizienz der DNA-Behandlung mit

der GpC-Methylase M.CviPI unterliegt experimentellen Schwankungen, sodass zuerst eine Normalisierung der Daten durchgeführt werden muss. Außerdem sind die Messwerte stärker verrauscht. Die Anwendung von DMR-Suchalgorithmen, wie z.B. metilene zur DNR-Entdeckung auf NOME-seq-Daten, ist formal möglich, scheint aber ebenfalls zu unterschiedlichen Resultaten, je nach eingesetzter Methode, zu führen. Unklar ist bislang, ob sich die Ansätze überhaupt erfolgreich übertragen lassen oder ob man für DNRs grundsätzlich andere Methoden entwickeln muss.

ATAC-seq

Ähnlich wie WGBS ist NOME-seq sehr teuer. Sehr viel preiswerter ist die Begrenzung der Sequenzierung auf die nukleosomenfreien Regionen. Historisch wurde hier die limitierte Verdauung des Chromatins mit DNaseI, gefolgt von der massiv parallelen Sequenzierung der dadurch generierten kleinen DNA-

Fragmente, eingesetzt. Diese Methode lässt sich aber sehr schwer einstellen und normalisieren. Eine technisch einfachere Lösung ist das ATAC-seq-Verfahren („Assay for Transposase-Accessible Chromatin using sequencing“), das eine Transposase verwendet. Die Transposase fragmentiert nukleosomenfreie DNA und hängt an die generierten kleinen DNA-Fragmente Sequenzieradaptoren an. Nach Amplifikation der DNA-Fragmente werden diese dann massiv parallel sequenziert und die erhaltenen Sequenzen dann gegen das Referenzgenom kartiert. Anschließend wird die Anreicherung von „reads“ entlang des Genoms bestimmt. Angereicherte Bereiche waren in den Ausgangszellen nukleosomenarm. Ein Problem ist die massive Anreicherung von mitochondrialen DNA-Fragmenten in der Sequenzierbibliothek (die zwar identifiziert und verworfen werden können), was das Sequenzieren unnötig verteuert. Nach Möglichkeit sollten also Mitochondrien vorher abgetrennt

werden, wie z.B. mit der Methode des Omni-ATAC [6].

Histonmodifikationen

Die DNA, und zwar jeweils 146 bp, ist um Histon-Oktamere gewickelt, meistens bestehend aus zwei Histonen H2A, zwei Histonen H2B, zwei Histonen H3 und zwei Histonen H4. Die N-terminalen Enden der Histone sind in der Regel an vielen Aminosäureseitenketten posttranslational modifiziert. Lysinseitenketten z.B. können acetyliert oder methyliert sein, Serinseitenketten z.B. phosphoryliert. Insgesamt gibt es dabei eine Vielzahl von Kombinationsmöglichkeiten, obwohl viele Modifikationen hochgradig korreliert oder antikorreliert sind. Häufig werden folgende Modifikationen untersucht: H3K27ac (Acetylierung von H3 an Lysin 27), H3K4me1, H3K4me3, H3K36me3, H3K27me3 und H3K9me3 (Bedeutung analog). H3K27ac markiert aktive Promotoren und Enhancer. H3K4me1 Enhancer, H3K4me3

Hier steht eine Anzeige.

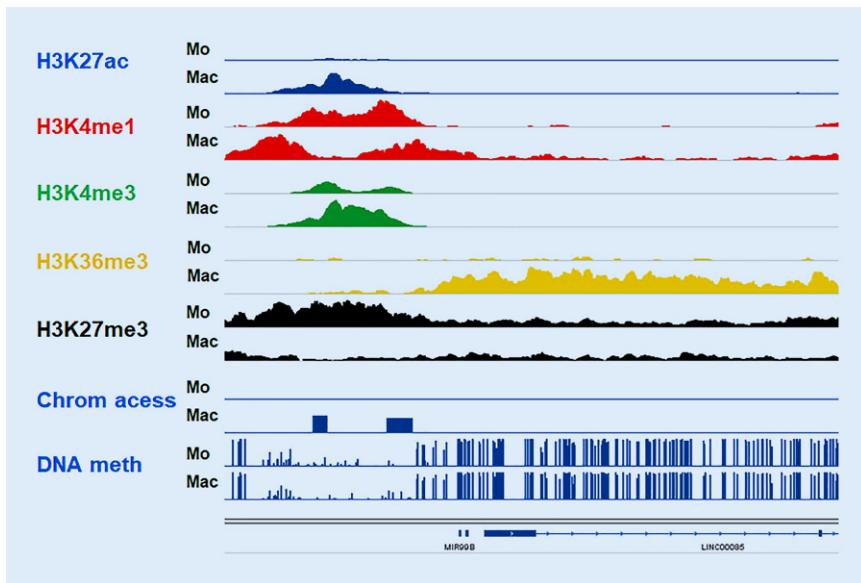


Abb. 4 ▲ Epigenetische Markierungen an einem miRNA/IncRNA-Gencluster in Monozyten (Mo) und Makrophagen (Mac). Screenshot (Integrated Genome Viewer) von fünf Histonmarkierungen, der Chromatinzugänglichkeit („chrom access“; NOME-seq „peaks“ in Makrophagen) und der DNA-Methylierung (DNA meth)

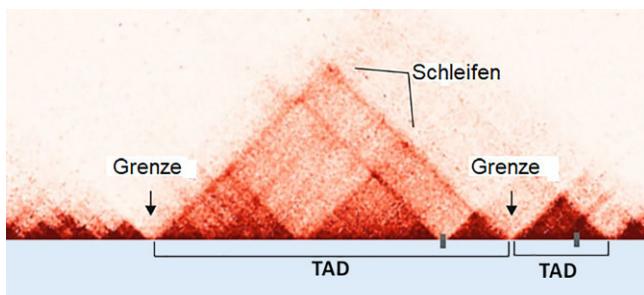


Abb. 5 ▲ Matrix von DNA-Kontakthäufigkeiten. Die Farbtiefe jedes Pixels gibt die Kontakthäufigkeit von DNA-Sequenzen an. Topologisch assoziierende Domänen (TAD) werden voneinander durch Grenzelemente getrennt. Kontaktpunkte von Chromatinschleifen sind deutlich sichtbar. (Überlassen von Prof. Dr. Stefan Mundlos, Berlin)

Promotoren, H3K36me3 transkribierte DNA, H3K27me3 Polycomb-reprimierte DNA und H3K9me3 Heterochromatin. Durch die Bestimmung dieser sechs Modifikationen können verschiedene Chromatinzustände einer Region mittels Kombinatorik und des ChromHMM-Algorithmus [7] noch weiter differenziert werden, z. B. in aktive TSS, aktive TSS-Umgebung, bivalente TSS, intergener Enhancer, transkribierter Enhancer, bivalenter Enhancer, starke Transkription, schwache Transkription, ZNF-Gene und Repeats, starke Polycomb-Repression, schwache Polycomb-Repression, Heterochromatin und stummes Chromatin. Mit dieser Information können regulatorische Elemente im Genom identifiziert

und die funktionelle Auswirkung von Sequenzvarianten abgeschätzt werden, die beim „whole-genome sequencing“ (WGS) außerhalb von Exonen gefunden werden.

Zur genomweiten Bestimmung der Anreicherung bestimmter Histonmodifikationen in bestimmten Regionen wird eine ChIP-seq durchgeführt. Dafür wird das Chromatin zunächst mit Formaldehyd fixiert, wobei die Histone kovalent an die DNA gebunden werden. Nach Fragmentierung des Chromatins auf eine Größe von 150–500 bp wird eine Immunopräzipitation mithilfe von Antikörpern gegen bestimmte Histonmodifikationen wie z. B. H3K4me3 durchgeführt. Es ist von äußerster Wichtigkeit,

dass hier hochspezifische Antikörper eingesetzt werden. Aus dem gereinigten Präzipitat werden dann die zuvor gebundenen DNA-Fragmente freigesetzt und nach Herstellung einer Sequenzierbibliothek massiv parallel sequenziert. Für die spätere Normalisierung ist eine ChIP mit unspezifischem IgG notwendig. Es sollten mindestens 50.000.000 reads mit mindestens 36 bp-Länge erzeugt werden. Die erhaltenen Sequenzen werden dann gegen das Referenzgenom kartiert. Anschließend wird die Anreicherung von „reads“ entlang des Genoms bestimmt. □ Abb. 4 zeigt einen IGV-Screenshot einer Region auf Chromosom 19 in Monozyten (Mo) und den daraus differenzierten Makrophagen (Mac). Es ist leicht erkennbar, dass ein Cluster von miRNA- und IncRNA-Genen auf Chromosom 19 während der Differenzierung aktiviert wird: das Signal der aktiven Histonmarkierungen (obere vier Doppelspuren) nimmt zu, das Signal für Polycomb-reprimiertes Chromatin (fünfte Doppelspur) nimmt ab. Das 5' liegende Chromatin (ein CGI) ist schon in Monozyten frei von DNA-Methylierung, aber das Chromatin öffnet sich erst während der Differenzierung (NOME-seq „peaks“).

3-D-Chromatinfaltung

Die dreidimensionale Faltung des Chromatins kompartimentiert das Genom und kann entfernt liegende funktionelle Elemente, wie z. B. Enhancer und Promotoren, räumlich zusammenbringen (□ Abb. 1b). HiC ist eine „chromosome conformation capture“ Methode, mit der genomweit räumlich benachbarte DNA-Sequenzen identifiziert werden können [19]. Bei dieser Methode werden Zellen mit Formaldehyd fixiert, sodass interagierende Loci durch kovalente DNA-Protein-Vernetzungen verbunden bleiben. Die DNA wird dann mit einem häufig schneidendem Restriktionsenzym geschnitten, die Enden mit biotinylierten Nukleotiden aufgefüllt und diese in großer Verdünnung ligiert. Dadurch wird eine Bibliothek von Ligationsprodukten = räumlich benachbarten DNA-Sequenzen generiert, die mittels Streptavidin aufgereinigt und nach Adaptor-

ligation massiv parallel sequenziert und gegen das Referenzgenom kartiert werden. Je häufiger zwei DNA-Sequenzen interagieren, desto häufiger sind sie in der Bibliothek vertreten. Aus diesen Daten lassen sich Chromosomenterritorien sowie topologisch assoziierende Domänen (TAD) identifizieren, die durch Grenzen („boundaries“) voneinander abgegrenzt sind (Abb. 5). In der Regel können Enhancer nur mit Promotoren in derselben TAD interagieren. Der Verlust einer Grenze durch eine Deletion kann es Enhancern innerhalb einer TAD ermöglichen, in aberranter Weise mit Promotoren einer anderen TAD zu interagieren. Eine solche „enhancer adoption“ [22] kann zu genetischen Erkrankungen führen. Dies ist natürlich keine Epimutation, sondern ein neu entdeckter Mechanismus, über den sich eine DNA-Veränderung (Deletion) auf die Genexpression auswirken kann.

Korrespondenzadresse



Prof. Dr.

Bernhard Horsthemke

Institut für Humangenetik,
Universität Duisburg-Essen
und Universitätsklinikum
Essen
Hufelandstr. 55, 45147 Essen,
Deutschland
bernhard.horsthemke@
uni-due.de

Danksagung. Die Autoren danken Dr. Klein-Hitpass für die Sequenzierung unserer NGS-Bibliotheken, Dr. Christopher Schröder und Prof. Dr. Sven Rahmann für die bioinformatischen Analysen sowie allen Kolleginnen und Kollegen des Deutschen Epigenom Programms (DEEP) und des International Human Epigenome Consortiums (IHEC) für die gute Zusammenarbeit. Die eigenen Arbeiten wurden von der DFG und dem BMBF gefördert.

Einhaltung ethischer Richtlinien

Interessenkonflikt. M. Zeschnigk und B. Horsthemke geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autoren keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfäl-

tigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz befügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- Aref-Eshghi E, Bend FG, Colaiacovo S et al (2019) Diagnostic utility of genome-wide DNA methylation testing in genetically unsolved individuals with suspected hereditary conditions. *Am J Hum Genet* 104:e685–700
- Buiting K, Dittrich B, Gross S et al (1998) Sporadic imprinting defects in Prader-Willi syndrome and Angelman syndrome: implications for imprint-switch models, genetic counseling, and prenatal diagnosis. *Am J Hum Genet* 63:170–180
- Burger L, Gaidatzis D, Schübeler D et al (2013) Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res* 41:e155
- Clark SJ, Harrison J, Paul CL et al (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 22:2990–2997
- Clark SJ, Smallwood SA, Lee HJ et al (2017) Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc* 12:534–547
- Corces MR, Trevino AE, Hamilton EG et al (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 14:959–962
- Ernst J, Kellis M (2017) Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 12:2478–2492
- Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89–92
- Greger V, Passarge E, Höpping W et al (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet* 83:155–158
- Gu C, Liu S, Wu Q et al (2019) Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. *Cell Res* 29:110–123
- Hansen KD, Langmead B, Irizarry RA (2012) BS-mooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13:R83
- Holliday R (1987) The inheritance of epigenetic defects. *Science* 238:163–170
- Horsthemke B (2006) Epimutations in human disease. *Curr Top Microbiol Immunol* 310:45–59
- <https://Bitbucket.Org/Svenrahmann/Amplikyzer/Wiki/Home> (Zugegriffen: 1.Mai 2019)
- Jühl F, Kretzmer H, Bernhart SH et al (2016) metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res* 26:256–262
- Kelly TK, Liu Y, Lay FD et al (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* 22:2497–2506
- Kleefstra T, Brunner HG, Amiel J et al (2006) Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am J Hum Genet* 79:370–377
- Leitao E, Beygo J, Zeschnigk M et al (2018) Locus-specific DNA methylation analysis by targeted deep bisulfite sequencing. *Methods Mol Biol* 1767:351–366
- Lieberman-Aiden E, Van Berkum NL, Williams L et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293
- Ligtenerberg MJ, Kuiper RP, Chan TL et al (2009) Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet* 41:112–117
- Lister R, O'Malley RC, Tonti-Filippini J et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133:523–536
- Lupianez DG, Kraft K, Heinrich V et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161:1012–1025
- Schröder C, Leitao E, Wallner S et al (2017) Regions of common inter-individual DNA methylation differences in human monocytes: genetic basis and potential function. *Epigenetics Chromatin* 10:37
- Statham AL, Taberlay PC, Kelly TK et al (2015) Genome-wide nucleosome occupancy and DNA methylation profiling of four human cell lines. *Genom Data* 3:94–96
- Tsuji J, Weng Z (2016) Evaluation of preprocessing, mapping and postprocessing algorithms for analyzing whole genome bisulfite sequencing data. *Brief Bioinformatics* 17:938–952
- Wallner S, Schröder C, Leitao E et al (2016) Epigenetic dynamics of monocyte-to-macrophage differentiation. *Epigenetics Chromatin* 9:33
- Wang Q, Gu L, Adey A et al (2013) Fragmentation-based whole-genome bisulfite sequencing. *Nat Protoc* 8:2022–2032