



(10) **DE 20 2016 009 134 U1** 2022.10.13

(12) **Gebrauchsmusterschrift**

(21) Aktenzeichen: **20 2016 009 134.5**

(22) Anmeldetag: **02.12.2016**

(67) aus Patentanmeldung: **EP 21 15 7421.5**

(47) Eintragungstag: **02.09.2022**

(45) Bekanntmachungstag im Patentblatt: **13.10.2022**

(51) Int Cl.: **C12Q 1/6869** (2018.01)  
**C12N 15/10** (2006.01)

(30) Unionspriorität:  
**62/263,532 04.12.2015 US**

(74) Name und Wohnsitz des Vertreters:  
**Mewburn Ellis LLP, 80333 München, DE**

(73) Name und Wohnsitz des Inhabers:  
**10X Genomics, Inc., Pleasanton, US**

**Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.**

(54) Bezeichnung: **Systeme und Zusammensetzungen zur Nukleinsäureanalyse**

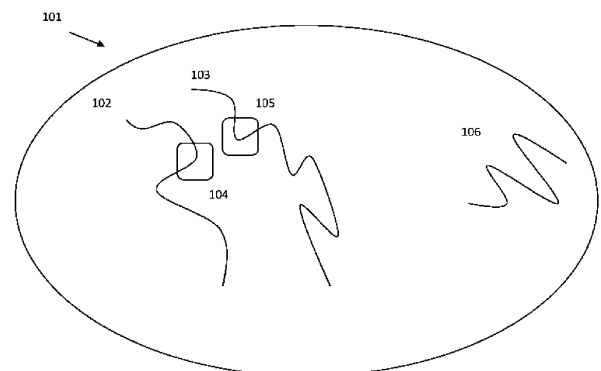
(57) Hauptanspruch: System, das dazu konfiguriert ist, Nukleinsäuren, die aus einer Formalin-fixierten in Paraffin eingebetteten (FFPE) Gewebeprobe gewonnen wurden, unter Beibehaltung von räumlichem Kontext zu analysieren, wobei die Analyse Folgendes umfasst:

a) Partitionieren der aus der FFPE-Gewebeprobe erhaltenen Nukleinsäuren in eine Vielzahl von Vertiefungen; wobei Nukleinsäuren in räumlicher Nähe zueinander in der FFPE-Gewebeprobe in dieselbe Vertiefung eingebracht werden;

b) Barcodieren partitionierter Nukleinsäuren, um eine Vielzahl barcodierter Nukleinsäuren zu bilden, wobei barcodierte Nukleinsäuren innerhalb einer gegebenen diskreten Vertiefung jeweils eine gemeinsame partitionsspezifische Barcodesequenz umfassen, sodass die Barcodesequenzen Nukleinsäuren aus einer gegebenen Vertiefung identifizieren;

c) Gewinnen von Sequenzinformationen von der Vielzahl barcodierter Nukleinsäuren, wobei die Sequenzinformationen von der Vielzahl barcodierter Nukleinsäuren Sequenzinformationen der partitionsspezifischen Barcodesequenz umfassen; und

d) Zuschreiben der Vielzahl barcodierter Nukleinsäuren zu einer Region räumlicher Nähe, wobei barcodierte Nukleinsäuren, die von einer Region räumlicher Nähe in der FFPE-Gewebeprobe abgeleitet sind, dieselbe partitionsspezifische Barcodesequenz umfassen.



**Beschreibung****QUERVERWEIS AUF VERWANDTE  
ANMELDUNGEN**

**[0001]** Diese Anmeldung beansprucht den Vorteil der vorläufigen US-Anmeldung Nr. 62/263, 532, eingereicht am 4. Dezember 2015, die hiermit in ihrer Gesamtheit für alle Zwecke durch Bezugnahme aufgenommen wird.

**HINTERGRUND DER ERFINDUNG**

**[0002]** Die Polynukleotidsequenzierung findet nach wie vor zunehmende Verwendung bei medizinischen Anwendungen wie etwa beim genetischen Screening und bei der Genotypisierung von Tumoren. Viele Polynukleotid-Sequenzierungsverfahren beruhen auf Probenverarbeitungstechniken der ursprünglichen Probe, einschließlich zufälliger Fragmentierung von Polynukleotiden. Diese Verarbeitungstechniken können Vorteile in Bezug auf Durchsatz und Effizienz bieten, den daraus resultierenden Sequenzinformationen, die aus diesen verarbeiteten Proben gewonnen werden, können jedoch wichtige Kontextinformationen in Bezug auf den Ort bestimmter Sequenzen innerhalb der weitergefassten linearen (zweidimensionalen) Sequenz des ursprünglichen Nukleinsäuremoleküls fehlen, der diese Sequenzen enthielt. Bei vielen Probenverarbeitungs- und Sequenzierungstechniken geht auch struktureller Kontext innerhalb des dreidimensionalen Raums der ursprünglichen Probe verloren. Es besteht daher ein Bedarf an Sequenzierungstechnologien, die strukturellen und molekularen Kontext der identifizierten Nukleinsäuresequenzen beibehalten.

**KURZDARSTELLUNG DER ERFINDUNG**

**[0003]** Demzufolge stellt die vorliegende Erfindung Verfahren, Systeme und Zusammensetzungen zum Bereitstellen von Sequenzinformationen bereit, die sowohl molekularen als auch strukturellen Kontext des ursprünglichen Nukleinsäuremoleküls beibehalten.

**[0004]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit. Solche Verfahren beinhalten die folgenden Schritte: (a) Bereitstellen einer Nukleinsäuren enthaltenden Probe, wobei die Nukleinsäuren dreidimensionale Strukturen umfassen; (b) Auftrennen von Teilen der Probe in diskrete Partitionen, sodass Teile der dreidimensionalen Nukleinsäurestrukturen ebenfalls in die diskreten Partitionen aufgetrennt werden; (c) Gewinnen von Sequenzinformationen aus den Nukleinsäuren, wodurch Nukleinsäuren unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0005]** In einigen Ausführungsformen beinhalten die Sequenzinformationen aus dem Gewinnungsschritt (c) eine Identifizierung von Nukleinsäuren, die sich in räumlicher Nähe zueinander befinden.

**[0006]** In weiteren Ausführungsformen beinhalten die Sequenzinformationen aus dem Gewinnungsschritt (c) eine Identifizierung von Nukleinsäuren, die sich in räumlicher Nähe zueinander befinden.

**[0007]** In weiteren Ausführungsformen stellt der Gewinnungsschritt (c) Informationen über intrachromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereit.

**[0008]** In weiteren Ausführungsformen stellt der Gewinnungsschritt (c) Informationen über Chromosomenkonformationen bereit.

**[0009]** In weiteren Ausführungsformen werden vor dem Auftrennungsschritt (b) mindestens einige der dreidimensionalen Strukturen verarbeitet, um verschiedene Teile der Nukleinsäuren zu verknüpfen, die innerhalb der dreidimensionalen Strukturen nahe beieinander liegen.

**[0010]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Auftrennungsschritt (b) nicht aus der Probe isoliert.

**[0011]** In allen Ausführungsformen werden die Nukleinsäuren innerhalb der diskreten Partitionen vor dem Gewinnungsschritt (c) barcodiert, um eine Vielzahl barcodierter Fragmente zu bilden, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen, sodass die Barcodes Nukleinsäuren aus einer gegebenen Partition identifizieren.

**[0012]** In weiteren Ausführungsformen umfasst der Gewinnungsschritt (c) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

**[0013]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit, die die folgenden Schritte beinhalten:

(a) Bilden verknüpfter Nukleinsäuren innerhalb der Probe, sodass räumlich benachbarte Nukleinsäuresegmente verknüpft werden;

(b) Verarbeiten der verknüpften Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten; (c) Ablegen der Vielzahl von Ligationsprodukten in diskrete Partitionen; (d) Barcodieren der Ligationsprodukte innerhalb

der diskreten Partitionen zum Bilden einer Vielzahl barcodierter Fragmente, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen, wodurch jedes Fragment der verknüpften Nukleinsäure zugeordnet wird, von der es abgeleitet ist; (e) Gewinnen von Sequenzinformationen aus der Vielzahl barcodierter Fragmente, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0014]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit, die die folgenden Schritte beinhalten:

(a) Bilden verknüpfter Nukleinsäuren innerhalb der Probe, sodass räumlich benachbarte Nukleinsäuresegmente verknüpft werden; (b) Ablegen der verknüpften Nukleinsäuren in diskrete Partitionen; (c) Verarbeiten der verknüpften Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten; (d) Barcodieren der Ligationsprodukte innerhalb der diskreten Partitionen zum Bilden einer Vielzahl barcodierter Fragmente, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen, wodurch jedes Fragment der verknüpften Nukleinsäure zugeordnet wird, von der es abgeleitet ist; (e) Gewinnen von Sequenzinformationen aus der Vielzahl barcodierter Fragmente, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0015]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit, die die folgenden Schritte beinhalten: (a) Vernetzen von Nukleinsäuren innerhalb der Probe zum Bilden vernetzter Nukleinsäuren, wobei das Vernetzen kovalente Verknüpfungen zwischen räumlich benachbarten Nukleinsäuresegmenten bildet; (b) Ablegen der vernetzten Nukleinsäuren in diskrete Partitionen; (c) Verarbeiten der vernetzten Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten; (d) Gewinnen von Sequenzinformationen aus der Vielzahl von Ligationsprodukten, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0016]** In allen Ausführungsformen ist die Probe eine Formalin-fixierte Probe in Paraffin.

**[0017]** In allen Ausführungsformen umfassen die diskreten Partitionen Beads. In weiteren Ausführungsformen sind die Beads Gel-Beads.

**[0018]** In allen Ausführungsformen umfasst die Probe eine Tumorseite.

**[0019]** In allen Ausführungsformen umfasst die Probe eine Mischung aus Tumor- und normalen Zellen.

**[0020]** In allen Ausführungsformen umfasst die Probe eine Kernmatrix.

**[0021]** In allen Ausführungsformen umfassen die Nukleinsäuren RNA.

**[0022]** In allen Ausführungsformen beträgt die Menge an Nukleinsäuren in der Probe weniger als 5, 10, 15, 20, 25, 30, 35, 40, 45 oder 50 ng/ml.

**[0023]** In einigen Aspekten stellt die vorliegende Erfindung ein Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit, wobei das Verfahren die folgenden Schritte beinhaltet: (a) Bereitstellen einer Nukleinsäuren enthaltenden Probe; (b) Aufbringen einer Bibliothek von Tags auf die Probe, sodass verschiedene geografische Regionen der Probe verschiedene Tags oder verschiedene Konzentrationen von Tags erhalten; (c) Auftrennen von Teilen der Probe in diskrete Partitionen, sodass Teile der Bibliothek von Tags und Teile der Nukleinsäuren ebenfalls in die diskreten Partitionen aufgetrennt werden; (d) Gewinnen von Sequenzinformationen von den Nukleinsäuren und (e) Identifizieren von Tags oder Konzentrationen von Tags in den diskreten Partitionen, wodurch Nukleinsäuren unter Beibehaltung von strukturellem Kontext analysiert werden.

#### Figurenliste

**Fig. 1** stellt eine schematische Veranschaulichung von molekularem Kontext und strukturellem Kontext gemäß den hier beschriebenen Verfahren bereit.

**Fig. 2** stellt eine schematische Veranschaulichung eines hier beschriebenen Prozesses dar.

**Fig. 3** veranschaulicht einen typischen Arbeitsablauf zum Durchführen eines Assays zum Erfassen von Sequenzinformationen unter Verwendung der hier offenbarten Verfahren und Zusammensetzungen.

**Fig. 4** stellt eine schematische Veranschaulichung eines Prozesses zum Kombinieren einer Nukleinsäureprobe mit Beads und Partitionieren der Nukleinsäuren und Beads in diskrete Tröpfchen bereit.

**Fig. 5** stellt eine schematische Veranschaulichung eines Prozesses zum Barcodieren und zur Amplifikation chromosomaler Nukleinsäurefragmente bereit.

**Fig. 6** stellt eine schematische Veranschaulichung der Verwendung eines Barcodierens von Nukleinsäurefragmenten beim Zuordnen von Sequenzdaten zu dem Stammnukleinsäuremolekül, von dem sie stammen, bereit.

**Fig. 7** stellt eine schematische Veranschaulichung eines beispielhaften Probenvorbereitungsverfahrens bereit.

#### AUSFÜHRLICHE BESCHREIBUNG DER ERFINDUNG

**[0024]** Bei der Ausführung der vorliegenden Erfindung können, sofern nicht anders angegeben, herkömmliche Techniken und Beschreibungen der organischen Chemie, Polymertechnologie, Molekularbiologie (einschließlich rekombinanter Techniken), Zellbiologie, Biochemie und Immunologie eingesetzt werden, die auf dem Fachgebiet hinreichend bekannt sind. Zu solchen herkömmlichen Techniken gehören Polymer-Array-Synthese, Hybridisierung, Ligation, Phagen-Display und Nachweis von Hybridisierung unter Verwendung eines Labels. Konkrete Veranschaulichungen geeigneter Techniken können unter Bezugnahme auf das nachstehende Beispiel erhalten werden. Es können aber natürlich auch andere äquivalente herkömmliche Vorgehensweisen verwendet werden. Solche herkömmlichen Techniken und Beschreibungen finden sich in Standard-Laborhandbüchern, wie etwa Genome Analysis: A Laboratory Manual Series (Bde. I-IV), Using Antibodies: A Laboratory Manual, Cells: A Laboratory Manual, PCR Primer: A Laboratory Manual, und Molecular Cloning: A Laboratory Manual (alle von Cold Spring Harbor Laboratory Press), Stryer, L. (1995) Biochemistry (4. Aufl.) Freeman, New York, Gait, „Oligonucleotide Synthesis: A Practical Approach“ 1984, IRL Press, London, Nelson und Cox (2000), Lehninger, Principles of Biochemistry 3. Aufl., W. H. Freeman Pub., New York, N. Y. und Berg et al. (2002), Biochemistry, 5. Aufl., W. H. Freeman Pub., New York, N. Y., die alle hiermit in ihrer Gesamtheit für alle Zwecke durch Bezugnahme aufgenommen werden.

**[0025]** Es sei darauf hingewiesen, dass im hier verwendeten Sinne und in den beigefügten Ansprüchen die Singularformen „ein“, „eine“ und „der“, „die“ und „das“ Pluralbezüge beinhalten, es sei denn, der Kontext schreibt eindeutig etwas anderes vor. So bezieht sich zum Beispiel eine Bezugnahme auf „eine Polymerase“ auf ein Mittel oder Gemische solcher Mittel, und eine Bezugnahme auf „das Verfahren“ beinhaltet eine Bezugnahme auf äquivalente Schritte und Ver-

fahren, die dem Fachmann bekannt sind, und so weiter.

**[0026]** Sofern nicht anders definiert, haben alle hier verwendeten technischen und wissenschaftlichen Begriffe die gleiche Bedeutung, wie sie allgemein von einem Durchschnittsfachmann auf dem Gebiet, zu dem diese Erfindung gehört, verstanden wird. Alle hier erwähnten Veröffentlichungen werden hiermit zum Zwecke der Beschreibung und Offenbarung von Vorrichtungen, Zusammensetzungen, Formulierungen und Methodologien, die in der Veröffentlichung beschrieben sind und die in Verbindung mit der vorliegend beschriebenen Erfindung verwendet werden könnten, durch Bezugnahme aufgenommen.

**[0027]** Wenn ein Bereich von Werten bereitgestellt wird, versteht es sich, dass jeder dazwischen liegende Wert auf das Zehntel der Einheit der Untergrenze, es sei denn, der Kontext schreibt eindeutig etwas anderes vor, zwischen der Ober- und Untergrenze dieses Bereichs und jeder andere angegebene oder dazwischen liegende Wert in diesem angegebenen Bereich von der Erfindung eingeschlossen sind. Die Ober- und Untergrenzen dieser kleineren Bereiche können unabhängig in den kleineren Bereichen beinhaltet sein, die ebenfalls von der Erfindung eingeschlossen sind, vorbehaltlich jeglicher konkret ausgeschlossener Grenze in dem angegebenen Bereich. Wenn der angegebene Bereich eine oder beide der Grenzen beinhaltet, sind Bereiche, die eine oder beide dieser beinhalteten Grenzen ausschließen, ebenfalls in der Erfindung beinhaltet.

**[0028]** In der folgenden Beschreibung werden zahlreiche konkrete Details dargelegt, um ein gründlicheres Verständnis der vorliegenden Erfindung bereitzustellen. Für den Fachmann ist es jedoch offensichtlich, dass die vorliegende Erfindung ohne eines oder mehrere dieser konkreten Details praktiziert werden kann. In anderen Fällen wurden hinreichend bekannte Merkmale und Vorgehensweisen, die dem Fachmann hinreichend bekannt sind, nicht beschrieben, um zu vermeiden, dass die Erfindung verschleiert wird.

**[0029]** Im hier verwendeten Sinne soll der Begriff „umfassend“ bedeuten, dass die Zusammensetzungen und Verfahren die aufgeführten Elemente beinhalten, jedoch andere nicht ausschließen. „Im Wesentlichen bestehend aus“ soll, wenn es zur Definition von Zusammensetzungen und Verfahren verwendet wird, den Ausschluss anderer Elemente von wesentlicher Bedeutung für die Zusammensetzung oder das Verfahren bedeuten. „Bestehend aus“ bedeutet den Ausschluss von mehr als Spurenelementen anderer Inhaltsstoffe für beanspruchte Zusammensetzungen und wesentliche Verfahrensschritte. Ausführungsformen, die durch jeden dieser

Übergangsbegriffe definiert sind, liegen innerhalb des Umfangs dieser Erfindung. Dementsprechend ist beabsichtigt, dass die Verfahren und Zusammensetzungen zusätzliche Schritte und Komponenten beinhalten können (umfassend) oder alternativ Schritte und Zusammensetzungen ohne Bedeutung beinhalten (im Wesentlichen bestehend aus) oder alternativ nur die angegebenen Verfahrensschritte oder Zusammensetzungen vorgesehen sind (bestehend aus).

**[0030]** Alle numerischen Bezeichnungen, z. B. pH, Temperatur, Zeit, Konzentration und Molekulargewicht, einschließlich Bereiche, sind Näherungswerte, die in Schritten von 0,1 (+) oder (-) variieren können. Es versteht sich, wenngleich nicht immer ausdrücklich darauf hingewiesen, dass allen Zahlenbezeichnungen der Begriff „etwa“ vorangestellt ist. Der Begriff „etwa“ beinhaltet auch den genauen Wert „X“ zusätzlich zu kleineren Inkrementen von „X“, wie etwa „X + 0,1“ oder „X - 0,1“. Es versteht sich auch, wenngleich nicht immer ausdrücklich darauf hingewiesen, dass die hier beschriebenen Reagenzien lediglich beispielhaft sind und dass Äquivalente davon auf dem Fachgebiet bekannt sind.

## I. Übersicht

**[0031]** Diese Offenbarung stellt Verfahren, Zusammensetzungen und Systeme zur Charakterisierung von genetischem Material bereit. Im Allgemeinen stellen die hier beschriebenen Verfahren, Zusammensetzungen und Systeme Verfahren zum Analysieren von Komponenten einer Probe unter Beibehaltung von Informationen über den strukturellen sowie molekularen Kontext dieser Komponenten, wie ursprünglich in der Probe vorliegend, bereit. Obwohl sich ein Großteil der Erörterung hier auf die Analyse von Nukleinsäuren bezieht, versteht es sich, dass die hier erörterten Verfahren und Systeme so angepasst werden können, dass sie auf andere Komponenten einer Probe angewendet werden können, einschließlich Proteine und andere Moleküle.

**[0032]** Desoxyribonukleinsäure (DNA) ist ein lineares Molekül, und von daher wird das Genom häufig in Bezug auf lineare Dimensionen beschrieben und beurteilt. Chromosomen sind jedoch nicht starr, und die räumliche Entfernung zwischen zwei genomischen Loci muss nicht immer ihrer Entfernung entlang der linearen Sequenz des Genoms entsprechen. Durch viele Megabasen getrennte Regionen können im dreidimensionalen Raum unmittelbar benachbart sein. Vom Standpunkt der Regulation aus kann es nützlich sein, weitreichende Wechselwirkungen zwischen genomischen Loci zu verstehen. Beispielsweise können Gen-Enhancer-, -Silencer- und -Isolatorelemente möglicherweise über große genomische Entfernungen hinweg funktionieren. Die Möglichkeit, sowohl strukturellen als auch

molekularen Kontext von Sequenz-Reads beizubehalten, bietet die Möglichkeit, solche weitreichende Wechselwirkungen zu verstehen.

**[0033]** „Beibehalten von strukturellem Kontext“ bedeutet im hier verwendeten Sinne, dass mehrere Sequenz-Reads oder mehrere Teile von Sequenz-Reads dem ursprünglichen dreidimensionalen relativen Ort dieser Sequenz-Reads innerhalb der Probe zugeordnet werden können. Anders ausgedrückt, die Sequenz-Reads können einem relativen Ort innerhalb der Probe in Bezug auf benachbarte Nukleinsäuren (und in einigen Situationen assoziierte Proteine) in dieser Probe zugeordnet werden. Diese räumlichen Informationen sind durch die hier erörterten Verfahren verfügbar, selbst wenn sich diese benachbarten Nukleinsäuren physisch nicht innerhalb der linearen Sequenz eines einzelnen Stammnukleinsäuremoleküls befinden. Bezug nehmend auf die schematische Veranschaulichung in **Fig. 1**: In einer Probe (101) befinden sich die Sequenzen (104) und (105) innerhalb der linearen Sequenz von zwei verschiedenen Stammnukleinsäuremolekülen ((102) bzw. (103)), befinden sich jedoch in räumlicher Nähe zueinander innerhalb der Probe. Die hier beschriebenen Verfahren und Zusammensetzungen bieten die Möglichkeit, diese Informationen über den strukturellen Kontext von Sequenz-Reads beizubehalten, und ermöglichen somit, dass Reads von den Sequenzen (104) und (105) ihrer relativen räumlichen Nähe innerhalb der ursprünglichen Probe auf den ursprünglichen Nukleinsäuremolekülen (102) und (103) zugewiesen werden können, von denen diese Sequenz-Reads abgeleitet sind.

**[0034]** Die hier erörterten Verfahren und Zusammensetzungen stellen auch Sequenzinformationen bereit, die molekularen Kontext beibehalten. „Beibehaltung molekularen Kontexts“ bedeutet im hier verwendeten Sinne, dass mehrere Sequenz-Reads oder mehrere Teile von Sequenz-Reads einem einzelnen Stammmolekül einer Nukleinsäure zugewiesen werden können. Wenngleich dieses einzelne Molekül einer Nukleinsäure eine beliebige einer Vielfalt von Längen haben kann, ist es in bevorzugten Aspekten ein relativ langes Molekül, das eine Erhaltung eines molekularen Kontexts über große Entfernungen ermöglicht. Insbesondere ist das einzelne Stammmolekül vorzugsweise wesentlich länger als die typische Sequenzlänge eines Short-Reads, z. B. länger als 200 Basen, und ist häufig mindestens 1000 Basen lang oder länger, 5000 Basen lang oder länger, 10.000 Basen lang oder länger, 20.000 Basen lang oder länger, 30.000 Basen lang oder länger, 40.000 Basen lang oder länger, 50.000 Basen lang oder länger, 60.000 Basen lang oder länger, 70.000 Basen lang oder länger, 80.000 Basen lang oder länger, 90.000 Basen lang oder länger oder 100.000 Basen lang oder länger und in einigen Fällen bis zu 1 Megabase lang oder länger.

**[0035]** Im Allgemeinen beinhalten die hier beschriebenen Verfahren ein Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem und molekularem Kontext. Solche Analysen beinhalten Verfahren, bei denen eine Nukleinsäuren enthaltende Probe bereitgestellt wird, wobei die Nukleinsäuren dreidimensionale Strukturen enthalten. Teile der Probe werden in diskrete Partitionen aufgetrennt, sodass Teile der dreidimensionalen Nukleinsäurestrukturen ebenfalls in die diskreten Partitionen aufgetrennt werden - Nukleinsäuresequenzen, die in räumlicher Nähe zueinander sind, neigen daher dazu, in die gleiche Partition aufgetrennt zu werden, und behalten somit die dreidimensionalen Informationen dieser räumlichen Nähe bei, selbst wenn später gewonnene Sequenz-Reads von Sequenzen stammen, die ursprünglich nicht auf demselben individuellen Stammnukleinsäuremolekül waren. Erneut Bezug nehmend auf **Fig. 1**: Wenn die Probe 101, die die Nukleinsäuremoleküle 102 und 103 und 106 enthält, in diskrete Partitionen aufgetrennt wird, sodass Teilmengen der Probe verschiedenen diskreten Partitionen zugeteilt werden, ist es wahrscheinlicher, dass die Nukleinsäuremoleküle 102 und 103 aufgrund des physischen Abstands zwischen dem Nukleinsäuremolekül 106 und 102 und 103 zusammen in der gleichen Partition platziert werden wie das Nukleinsäuremolekül 106. Von daher sind Nukleinsäuremoleküle innerhalb der gleichen diskreten Partitionen diejenigen, die in der ursprünglichen Probe in räumlicher Nähe zueinander waren. Von Nukleinsäuren innerhalb der diskreten Partitionen gewonnene Sequenzinformationen stellen somit eine Möglichkeit bereit, die Nukleinsäuren zu analysieren, beispielsweise durch Nukleinsäuresequenzierung, und diese Sequenz-Reads dem strukturellen Kontext der Stammnukleinsäuremoleküle zuzuweisen.

**[0036]** In weiteren Beispielen kann der strukturelle Kontext (hier auch als „geografischer Kontext“ bezeichnet) durch eine Verwendung von Tags (wie etwa Barcode-Oligonukleotide) zur Codierung der Geografie der Probe beibehalten werden. In einigen Situationen kann dies ein Injizieren einer viralen Bibliothek, die für eine Sammlung von barcodierten Sequenzen (wie etwa mRNA-Sequenzen) codiert, in eine Probe beinhalten. Die Barcodes wandern durch aktive Prozesse oder durch Diffusion durch die Probe. Wenn die Probe dann gemäß den hier beschriebenen und auf dem Fachgebiet bekannten Verfahren weiterverarbeitet wird, können Barcodes mit strukturellen Positionen korreliert werden, um Nukleinsäuresequenzen von demselben geografischen Ort innerhalb der Probe zu identifizieren. In Beispielen, in denen die Barcodes durch aktive Prozesse durch die Probe verteilt werden, können Sequenzen mit demselben Barcode geografisch verbunden und/oder durch denselben Prozess verbunden sein. Wie zu erkennen ist, kann dieses System der Verwendung von Tags zum Codieren von struk-

turellem Kontext allein oder in Kombination mit hier beschriebenen Verfahren verwendet werden, wobei diskrete Partitionen verwendet werden, um strukturellen und molekularen Kontext noch besser beizubehalten. In Beispielen, in denen Tags zum Codieren räumlicher Orte und Barcodes zum Identifizieren von Molekülen verwendet werden, die in die gleichen diskreten Partitionen aufgetrennt sind, werden die Proben im Wesentlichen mit Tag versehen oder „doppelt barcodiert“, wenn ein Satz Barcodes zum Identifizieren räumlicher Orte verwendet wird und ein Satz Barcodes partitionsspezifisch ist. In solchen Beispielen können beide Barcodesätze verwendet werden, um Informationen bereitzustellen, um strukturellen und molekularen Kontext der aus der Probe generierten Sequenz-Reads beizubehalten.

**[0037]** In einigen Beispielen stellen die von den Nukleinsäuren gewonnenen Sequenzinformationen Informationen über intrachromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereit. In weiteren Beispielen beinhalten die Sequenzinformationen Informationen über Chromosomenkonformationen.

**[0038]** In weiteren Beispielen können die Nukleinsäuren in der Probe vor der Auftrennung in die diskreten Partitionen verarbeitet werden, um verschiedene Regionen ihrer dreidimensionalen Strukturen zu verknüpfen, sodass Regionen der Sequenz, die innerhalb dieser dreidimensionalen Struktur nahe beieinander liegen, miteinander verbunden werden. Von daher trennt die Auftrennung der Probe in diskrete Partitionen diese verknüpften Regionen in die gleiche Partition auf, wodurch weiter sichergestellt wird, dass der strukturelle Kontext jeglicher Sequenz-Reads von diesen Nukleinsäuren beibehalten wird.

**[0039]** In einigen Situationen kann die Verknüpfung von Nukleinsäuren unter Verwendung beliebiger, auf dem Fachgebiet bekannter Verfahren erreicht werden, die verwendet werden, um Moleküle in räumlicher Nähe zu vernetzen. Zu solchen Vernetzungsmitteln können Alkylierungsmittel, Cisplatin, Distickstoffoxid, Psoralene, Aldehyde, Acrolein, Glyoxal, Osmiumtetroxid, Carbodiimid, Quecksilberchlorid, Zinksalze, Pikrinsäure, Kaliumdichromat, Ethanol, Methanol, Aceton, Essigsäure und dergleichen gehören, ohne darauf beschränkt zu sein. In konkreten Beispielen werden die Nukleinsäuren unter Verwendung von Protokollen verknüpft, die für die Analyse der dreidimensionalen Architektur von Genomen ausgelegt sind, wie etwa das „Hi-C“-Protokoll, das beispielsweise in Dekker et al., „Capturing chromosome conformation“, *Science* 295:1306-1311 (2002), und Berkum et al., *J. Vis. Exp.* (39), e1869, doi:10.3791/1869 (2010), beschrieben wird, die hiermit jeweils in ihrer Gesamtheit durch Bezugnahme für alle Zwecke und insbesondere für alle Lehren,

die sich auf die Verknüpfung von Nukleinsäuremolekülen beziehen, eingeschlossen werden. Solche Protokolle beinhalten im Allgemeinen ein Erzeugen einer Bibliothek von Molekülen durch Vernetzung der Probe, sodass genomische Loci, die sich in enger räumlicher Nähe befinden, verknüpft werden. In weiteren Ausführungsformen wird die dazwischensliegende DNA-Schleife zwischen der Vernetzung wegverdaut, und dann werden die Intrasequenz-Regionen zur Hinzufügung zu der Bibliothek einem Reverse Crosslinking unterzogen. Die Verdau- und Reverse-Crosslinking-Schritte können vor einem Schritt des Partitionierens der Probe in diskrete Partitionen erfolgen, oder sie können innerhalb der Partitionen nach dem Auftrennungsschritt erfolgen.

**[0040]** In weiteren Beispielen können die Nukleinsäuren einem Tagging- oder Barcodierungsschritt unterzogen werden, der einen gemeinsamen Barcode für alle Nukleinsäuren innerhalb einer Partition bereitstellt. Wie zu erkennen ist, kann diese Barcodierung mit oder ohne den oben erörterten Nukleinsäureverknüpfungs-/vernetzungsschritten erfolgen. Die Verwendung der hier offenbarten Barcodierungstechnik verleiht die einzigartige Möglichkeit, einen individuellen strukturellen und molekularen Kontext für genomische Regionen bereitzustellen - d. h. durch Zuschreiben bestimmter Sequenz-Reads zu einzelnen Nukleinsäuremolekülen einer Probe und durch variantenkoordinierte Assemblierung einen breiteren oder sogar weiter reichenden erschlossenen Kontext unter mehreren Nukleinsäuremolekülen einer Probe und/oder zu einem bestimmten Chromosom bereitzustellen. Der Begriff „genomische Region“ oder „Region“ bezieht sich im hier verwendeten Sinne auf eine beliebige definierte Länge eines Genoms und/oder Chromosoms. Beispielsweise kann sich eine genomische Region auf die Assoziierung (d. h. beispielsweise eine Wechselwirkung) zwischen mehr als einem Chromosom beziehen. Eine genomische Region kann auch ein vollständiges Chromosom oder ein Teilchromosom umfassen. Zusätzlich kann eine genomische Region eine bestimmte Nukleinsäuresequenz auf einem Chromosom (d. h. zum Beispiel ein offenes Leseraster und/oder ein regulatorisches Gen) oder eine intergenische nichtcodierende Region beinhalten.

**[0041]** Die Verwendung von Barcodierung verleiht die zusätzlichen Vorteile der Erleichterung der Möglichkeit, zwischen Minderheitsbestandteilen und Hauptbestandteilen der gesamten aus der Probe extrahierten Nukleinsäurepopulation zu unterscheiden, z. B. zum Nachweis und zur Charakterisierung von zirkulierender Tumor-DNA im Blutstrom, und reduziert auch oder eliminiert Amplifikations-Bias während optionaler Amplifikationsschritte. Darüber hinaus verleiht die Implementierung in einem Mikrofluidikformat die Möglichkeit, mit extrem kleinen Probenvolumina und geringen Eingabemengen an DNA

zu arbeiten, sowie die Möglichkeit, schnell eine große Anzahl von Probenpartitionen (Tröpfchen) zu verarbeiten, um ein genomweites Tagging zu ermöglichen.

**[0042]** Zusätzlich zum Bereitstellen der Möglichkeit, Sequenzinformationen von ganzen oder ausgewählten Regionen des Genoms zu gewinnen, können die hier beschriebenen Verfahren und Systeme auch andere Charakterisierungen von genomischem Material bereitstellen, einschließlich, aber nicht beschränkt auf Haplotyp-Phasierung, Identifizierung von strukturellen Variationen und Kopienzahlvariationen, wie in den USSNs 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; und 14/316,463 beschrieben, die hiermit in ihrer Gesamtheit für alle Zwecke und insbesondere für alle schriftlichen Beschreibungen, Figuren und Arbeitsbeispiele, die auf die Charakterisierung von genomischem Material gerichtet sind, durch Bezugnahme aufgenommen werden.

**[0043]** Im Allgemeinen beinhalten erfindungsgemäße Verfahren Schritte wie in **Fig. 2** veranschaulicht, die einen schematischen Überblick über erfindungsgemäße Verfahren bereitstellt, die hier ausführlicher erörtert werden. Wie zu erkennen ist, ist das in **Fig. 2** skizzierte Verfahren ein Ausführungsbeispiel, das nach Bedarf und wie hier beschrieben geändert oder modifiziert werden kann. Wie in **Fig. 2** gezeigt können die hier beschriebenen Verfahren einen optionalen Schritt 201 beinhalten, bei dem Nukleinsäuren einer Probe verarbeitet werden, um Nukleinsäuren in räumlicher Nähe zueinander zu verknüpfen. Mit oder ohne diesen vorläufigen Verarbeitungsschritt (201) werden die hier beschriebenen Verfahren in den meisten Beispielen einen Schritt beinhalten, in dem enthaltene Nukleinsäuren einer Probe partitioniert werden (202). Im Allgemeinen wird jede Partition, die Nukleinsäuren aus genomischen Regionen von Interesse enthält, einem Prozess unterzogen, der zu Fragmenten führt, die Barcodes enthalten (203). Diese Fragmente können dann vor einem Sequenzieren (205) gepoolt (204) werden. Die Sequenz-Reads aus (205) können im Allgemeinen aufgrund der partitionsspezifischen Barcodes (203) dem strukturellen und molekularen Stammkontext (206) zugewiesen werden. Jede Partition kann in einigen Beispielen mehr als eine Nukleinsäure beinhalten und wird in einigen Fällen mehrere hundert Nukleinsäuremoleküle enthalten. Die barcodierten Fragmente von Schritt 203 können unter Verwendung beliebiger auf dem Fachgebiet bekannter Verfahren erzeugt werden - in einigen Beispielen sind innerhalb der verschiedenen Partitionen Oligonukleotide in den Proben enthalten. Solche Oligonukleotide können Zufallssequenzen umfassen, die dazu bestimmt sind, zahlreiche verschiedene Regionen der Proben zufällig zu primen, oder sie können eine bestimmte Primersequenz umfassen,

die darauf abzielt, stromaufwärts einer Zielregion der Probe zu primen. In weiteren Beispielen enthalten diese Oligonukleotide auch eine Barcode-Sequenz, sodass der Replikationsprozess auch das resultierende replizierte Fragment der ursprünglichen Nukleinsäure einer Probe barcodiert. Ein besonders eleganter Prozess zur Verwendung dieser Barcode-Oligonukleotide beim Amplifizieren und bei der Barcodierung von Proben ist ausführlich in den USSNs 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; und 14/316,463 beschrieben, die hiermit jeweils in ihrer Gesamtheit für alle Zwecke und insbesondere für alle Lehren in Bezug auf Barcodierung und Amplifikation von Oligonukleotiden durch Bezugnahme aufgenommen werden. Extensionsreaktionsreagenzien, z. B. DNA-Polymerase, Nukleosidtriphosphate, Cofaktoren (z. B.  $Mg^{2+}$  oder  $Mn^{2+}$  usw.), die ebenfalls in den Partitionen enthalten sind, verlängern dann die Primersequenz unter Verwendung der Probe als Templat, um ein komplementäres Fragment zu dem Strang des Templates zu erzeugen, an den sich der Primer angelagert (Annealing) hat, und das komplementäre Fragment beinhaltet das Oligonukleotid und seine zugeordnete Barcode-Sequenz. Das Annealing und die Extension mehrerer Primer an verschiedenen Teilen der Probe kann zu einem großen Pool überlappender komplementärer Fragmente der Probe führen, von denen jedes seine eigene Barcode-Sequenz besitzt, die auf die Partition hinweist, in der es erzeugt wurde. In einigen Fällen können diese komplementären Fragmente selbst als Templat verwendet werden, das durch die in der Partition vorhandenen Oligonukleotide geprimt wird, um ein Komplement des Komplements zu erzeugen, das wiederum die Barcode-Sequenz enthält. In weiteren Beispielen ist dieser Replikationsprozess derart konfiguriert, dass er, wenn das erste Komplement dupliziert wird, zwei komplementäre Sequenzen an oder in der Nähe seiner Termini erzeugt, um die Bildung einer Haarnadelstruktur oder partiellen Haarnadelstruktur zu ermöglichen, was die Möglichkeit des Moleküls verringert, die Basis für die Herstellung weiterer iterativer Kopien zu sein. Ein Vorteil der hier beschriebenen Verfahren und Systeme besteht darin, dass das Anbringen eines partitions- oder probenspezifischen Barcodes an den kopierten Fragmenten den ursprünglichen molekularen Kontext der sequenzierten Fragmente bewahrt, wodurch sie ihrer ursprünglichen Partition und somit ihrem Stammnukleinsäuremolekül in der Probe zugewiesen werden können.

**[0044]** Häufig wird die Probe mit einem Satz von Oligonukleotid-Tags kombiniert, die vor dem Partitionierungsschritt lösbar an Beads angebracht sind. Verfahren zur Barcodierung von Nukleinsäuren sind auf dem Fachgebiet bekannt und hier beschrieben. In einigen Beispielen werden Verfahren eingesetzt, wie in Amini et al., 2014, Nature Genetics, Advance Online Publication) beschrieben, das hiermit in sei-

ner Gesamtheit für alle Zwecke und insbesondere für alle Lehren in Bezug auf das Anbringen von Barcodes oder anderen Oligonukleotid-Tags an Nukleinsäuren durch Bezugnahme aufgenommen wird. Verfahren zum Verarbeiten und Sequenzieren von Nukleinsäuren gemäß den in der vorliegenden Anmeldung beschriebenen Verfahren und Systemen sind auch ausführlicher in den USSNs 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; und 14/316,463 beschrieben, die hiermit in ihrer Gesamtheit für alle Zwecke und insbesondere für alle schriftlichen Beschreibungen, Figuren und Arbeitsbeispiele, die auf das Verarbeiten von Nukleinsäuren und das Sequenzieren und andere Charakterisierungen von genomischem Material gerichtet sind, durch Bezugnahme aufgenommen werden.

**[0045]** Zusätzlich zu dem obigen Arbeitsablauf können genomische Zielregionen für weitere Analysen, insbesondere Sequenzierung, unter Verwendung von Verfahren, zu denen sowohl Chip-basierte als auch lösungsbasierte Einfangverfahren gehören, angereichert, isoliert oder getrennt, d. h. „pulled down“ werden. Bei solchen Verfahren werden Sonden eingesetzt, die zu den genomischen Regionen von Interesse oder zu Regionen nahe oder benachbart zu den genomischen Regionen von Interesse komplementär sind. Beispielsweise werden beim Hybrid- (oder Chipbasierten) Einfangen Mikroarrays, die Fangsonden (normalerweise einzelsträngige Oligonukleotide) mit Sequenzen enthalten, die zusammengekommen die Region von Interesse abdecken, an einer Oberfläche fixiert. Genomische DNA wird fragmentiert und kann weiter verarbeitet werden, wie etwa Endreparatur, um Blunt-Ends zu erzeugen, und/oder Hinzufügen zusätzlicher Merkmale, wie etwa universeller Priming-Sequenzen. Diese Fragmente werden an die Sonden auf dem Mikroarray hybridisiert. Nichthybridisierte Fragmente werden gewaschen, und die gewünschten Fragmente werden eluiert oder anderweitig auf der Oberfläche zur Sequenzierung oder für andere Analysen verarbeitet, und somit wird die Population von Fragmenten, die auf der Oberfläche verbleibt, mit Fragmenten angereichert, die die Zielregionen von Interesse enthalten (z. B. die Regionen, die die Sequenzen, die zu den in den Fangsonden enthaltenen komplementär sind). Die angereicherte Population von Fragmenten kann ferner unter Verwendung beliebiger auf dem Fachgebiet bekannter Amplifikationstechnologien amplifiziert werden. Beispielhafte Verfahren für solche gezielten Pulldown-Anreicherungsverfahren sind in USSN 14/927,297, eingereicht am 29. Oktober 2015, beschrieben, das hiermit in seiner Gesamtheit für alle Zwecke und insbesondere für alle Lehren in Bezug auf gezielte Pulldown-Anreicherungsverfahren und Sequenzierungsverfahren, einschließlich aller schriftlichen Beschreibungen, Figuren und Beispiele durch Bezugnahme aufgenommen wird. Die Population der genomischen Zielregionen kann vor



den oben beschriebenen Pulldown-Verfahren weiter angereichert werden, indem Verfahren verwendet werden, um die Abdeckung dieser Zielregionen zu erhöhen. Eine solche erhöhte Abdeckung kann beispielsweise unter Verwendung gezielter Amplifikationsverfahren erreicht werden, einschließlich jener, die beispielsweise in USSN 62/119,996, eingereicht am 24. Februar 2015, beschrieben sind, das hiermit für alle Zwecke und insbesondere für alle Lehren in Bezug auf gezielte Abdeckung von Nukleinsäuremolekülen durch Bezugnahme aufgenommen wird.

**[0046]** In bestimmten Fällen beinhalten die hier beschriebenen Verfahren einen Schritt, in dem ausgewählte Regionen des Genoms vor der Sequenzierung selektiv amplifiziert werden. Diese Amplifikation, die im Allgemeinen unter Verwendung von auf dem Fachgebiet bekannten Verfahren durchgeführt wird (einschließlich, aber nicht beschränkt auf PCR-Amplifikation), stellt eine mindestens 1-fache, 10-fache, 20-fache, 50-fache, 100-fache, 200-fache, 500-fache, 1000-fache, 1500-fache, 2000-fache, 5000-fache oder 10000-fache Abdeckung der ausgewählten Regionen des Genoms bereit, wodurch eine Menge an Nukleinsäuren bereitgestellt wird, um eine De-novo-Sequenzierung dieser ausgewählten Regionen zu ermöglichen. In weiteren Ausführungsformen stellt die Amplifikation mindestens 1- bis 20-fache, 50- bis 100-fache, 200- bis 1000-fache, 1500- bis 5000-fache, 5000- bis 10.000-fache, 1000- bis 10000-fache, 1500- bis 9000-fache, 2000- bis 8000-fache, 2500- bis 7000-fache, 3000- bis 6500-fache, 3500- bis 6000-fache, 4000- bis 5500-fache Abdeckung der ausgewählten Regionen des Genoms bereit.

**[0047]** Die Amplifikation wird im Allgemeinen durch Extension von Primern durchgeführt, die zu Sequenzen innerhalb oder in der Nähe der ausgewählten Regionen des Genoms komplementär sind. In einigen Fällen wird eine Bibliothek von Primern verwendet, die so ausgelegt ist, dass sie die Regionen von Interesse verkachelt - anders ausgedrückt ist die Bibliothek von Primern dafür ausgelegt, Regionen in bestimmten Entfernungen entlang der ausgewählten Regionen des Genoms zu amplifizieren. In einigen Fällen werden bei der selektiven Amplifikation Primer eingesetzt, die zu allen 10, 15, 20, 25, 50, 100, 200, 250, 500, 750, 1000 oder 10000 Basen entlang der ausgewählten Regionen des Genoms komplementär sind. In weiteren Beispielen ist die gekachelte Bibliothek von Primern dafür ausgelegt, eine Mischung von Entfernungen einzufangen - diese Mischung kann eine zufällige Mischung von Entfernungen sein oder intelligent so ausgelegt sein, dass bestimmte Teile oder Prozentsätze der ausgewählten Regionen durch verschiedene Primerpaare amplifiziert werden. Weitere Informationen zur gezielten Abdeckung des Genoms zur Verwendung gemäß hier beschriebener Verfahren wird zum Beispiel in USSN 62/146,834,

eingereicht am 13. April 2015, bereitgestellt, die hiermit in seiner Gesamtheit für alle Zwecke und insbesondere für alle Lehren, die sich auf die gezielte Abdeckung eines Genoms beziehen, durch Bezugnahme aufgenommen wird.

**[0048]** Im Allgemeinen stellen die hier beschriebenen Verfahren und Systeme Nukleinsäuren für Analysen, wie etwa Sequenzierung, bereit. Sequenzierungsinformationen werden unter Verwendung von Verfahren gewonnen, die die Vorteile der extrem niedrigen Sequenzierungsfehlerraten und des hohen Durchsatzes von Short-Read-Sequenzierungstechnologien haben. Wie oben beschrieben wird die Sequenzierung von Nukleinsäuren typischerweise in einer Weise durchgeführt, die den strukturellen und molekularen Kontext von Sequenz-Reads oder Teilen von Sequenz-Reads bewahrt. Damit ist gemeint, dass mehrere Sequenz-Reads oder mehrere Teile von Sequenz-Reads dem räumlichen Ort in Bezug auf andere Nukleinsäuren in der ursprünglichen Probe (struktureller Kontext) und dem Ort dieses Sequenz-Reads entlang der linearen Sequenz eines einzelnen Stammoleküls einer Nukleinsäure (molekularer Kontext) zugewiesen werden können. Wenngleich dieses einzelne Molekül einer Nukleinsäure eine beliebige Länge haben kann, wird es in bevorzugten Aspekten ein relativ langes Molekül sein, was die Bewahrung eines molekularen Kontextes über große Entfernungen ermöglicht. Insbesondere ist das einzelne Stammolekül vorzugsweise wesentlich länger als die typische Sequenzlänge eines Short-Reads, z. B. länger als 200 Basen, und ist häufig mindestens 1000 Basen lang oder länger, 5000 Basen lang oder länger, 10.000 Basen lang oder länger, 20.000 Basen lang oder länger, 30.000 Basen lang oder länger, 40.000 Basen lang oder länger, 50.000 Basen lang oder länger, 60.000 Basen lang oder länger, 70.000 Basen lang oder länger, 80.000 Basen lang oder länger, 90.000 Basen lang oder länger oder 100.000 Basen lang oder länger und in einigen Fällen bis zu 1 Megabase lang oder länger.

**[0049]** Wie oben angemerkt, stellen die hier beschriebenen Verfahren und Systeme einen individuellen molekularen Kontext für Reads von kurzen Sequenzen längerer Nukleinsäuren bereit. Im hier verwendeten Sinne bezieht sich individueller molekularer Kontext auf Sequenzkontext über den konkreten Sequenz-Read hinaus, z. B. in Bezug auf benachbarte oder proximale Sequenzen, die in dem Sequenz-Read selbst nicht enthalten sind, und von daher typischerweise so sind, dass sie weder ganz noch teilweise in einem kurzen Sequenz-Read, z. B. einem Read von etwa 150 Basen oder etwa 300 Basen für gepaarte Reads, enthalten wären. In besonders bevorzugten Aspekten stellen die Verfahren und Systeme einen weitreichenden Sequenzkontext für kurze Sequenz-Reads bereit. Ein solcher

weitreichender Kontext beinhaltet eine Beziehung oder Verknüpfung eines gegebenen Sequenz-Reads mit Sequenz-Reads, die innerhalb einer Entfernung von länger als 1 kb, länger als 5 kb, länger als 10 kb, länger als 15 kb, länger als 20 kb, länger als 30 kb, länger als 40 kb, länger als 50 kb, länger als 60 kb, länger als 70 kb, länger als 80 kb, länger als 90 kb oder sogar länger als 100 kb oder länger zueinander liegen. Wie zu erkennen ist, kann man durch Bereitstellen eines weitreichenden individuellen molekularen Kontexts auch die Phasierungsinformationen von Varianten innerhalb dieses individuellen molekularen Kontexts ableiten, z. B. sind Varianten auf einem besonders langen Molekül definitionsgemäß gemeinsam phasiert.

**[0050]** Durch Bereitstellen eines individuellen weitreichenden molekularen Kontextes stellen die erfindungsgemäßen Verfahren und Systeme auch einen viel längeren erschlossenen molekularen Kontext (hier auch als „langer virtueller Einzelmolekül-Read“ bezeichnet) bereit. Sequenzkontext, wie hier beschrieben, kann ein Abbilden (Mapping) oder Bereitstellen einer Verknüpfung von Fragmenten über verschiedene (im Allgemeinen im Kilobasenmaßstab) Bereiche der vollständigen genomischen Sequenz beinhalten. Diese Verfahren beinhalten ein Abbilden der kurzen Sequenz-Reads auf die einzelnen längeren Moleküle oder Contigs verknüpfter Moleküle sowie die weitreichende Sequenzierung großer Teile der längeren einzelnen Moleküle, z. B. mit zusammenhängenden bestimmten Sequenzen einzelner Moleküle, wobei solche bestimmten Sequenzen länger als 1 kb, länger als 5 kb, länger als 10 kb, länger als 15 kb, länger als 20 kb, länger als 30 kb, länger als 40 kb, länger als 50 kb, länger als 60 kb, länger als 70 kb, länger als 80 kb, länger als 90 kb oder sogar länger als 100 kb sind. Wie beim Sequenzkontext kann die Zuweisung kurzer Sequenzen zu längeren Nukleinsäuren, z. B. sowohl einzelnen langen Nukleinsäuremolekülen als auch Sammlungen verknüpfter Nukleinsäuremoleküle oder Contigs, sowohl eine Abbildung von kurzen Sequenzen auf längere Nukleinsäureabschnitte, um einen Sequenzkontext auf hohem Niveau bereitzustellen, als auch ein Bereitstellen assemblierter Sequenzen von den kurzen Sequenzen bis zu diesen längeren Nukleinsäuren beinhalten.

**[0051]** Wenngleich der weitreichende Sequenzkontext, der mit langen individuellen Molekülen assoziiert ist, genutzt wird, ermöglicht es ein derartiger weitreichender Sequenzkontext auch, einen noch weiter reichenden Sequenzkontext zu erschließen. Beispielsweise ist es möglich, durch Bereitstellen des oben beschriebenen weitreichenden molekularen Kontexts überlappende Variantenteile, z. B. phasierte Varianten, translozierte Sequenzen usw., unter langen Sequenzen von verschiedenen Stammolekülen zu identifizieren, was die erschlossene Ver-

knüpfung zwischen diesen Molekülen ermöglicht. Solche erschlossenen Verknüpfungen oder molekularen Kontexte werden hier als „erschlossene Contigs“ bezeichnet. In einigen Fällen können die erschlossenen Contigs, wenn sie im Zusammenhang mit phasierten Sequenzen erörtert werden, üblicherweise phasierte Sequenzen darstellen, z. B. wenn aufgrund überlappender phasierter Varianten ein phasiertes Contig von wesentlich größerer Länge als die einzelnen Stammoleküle erschlossen wird. Diese phasierten Contigs werden hier als „Phasenblöcke“ bezeichnet.

**[0052]** Durch Beginnen mit längeren Einzelmolekül-Reads (z. B. den oben erörterten „langen virtuellen Einzelmolekül-Reads“) können längere erschlossene Contigs oder Phasenblöcke abgeleitet werden, als dies sonst unter Verwendung von Short-Read-Sequenzierungstechnologien oder anderen Ansätzen zur phasierten Sequenzierung erreichbar wäre. Siehe z. B. die veröffentlichte US-Patentanmeldung Nr. 2013-0157870. Insbesondere ist es unter Verwendung der hier beschriebenen Verfahren und Systeme möglich, erschlossene Contig- oder Phasenblocklängen mit einem N50 (wobei die Summe der Blocklängen, die größer als die angegebene N50-Zahl sind, 50 % der Summe aller Blocklängen beträgt) von mindestens etwa 10 kb, mindestens etwa 20 kb, mindestens etwa 50 kb zu gewinnen. In bevorzugteren Aspekten werden erschlossene Contig- oder Phasenblocklängen mit einer N50 von mindestens etwa 100 kb, mindestens etwa 150 kb, mindestens etwa 200 kb und in vielen Fällen mindestens etwa 250 kb, mindestens etwa 300 kb, mindestens etwa 350 kb, mindestens etwa 400 kb und in einigen Fällen mindestens etwa 500 kb oder mehr erreicht. In wieder anderen Fällen können maximale Phasenblocklängen von über 200 kb, über 300 kb, über 400 kb, über 500 kb, über 1 Mb oder sogar über 2 Mb gewonnen werden.

**[0053]** In einem Aspekt und in Verbindung mit beliebigen der oben und unten hier beschriebenen Verfahren sorgen die hier beschriebenen Verfahren und Systeme für die Kompartimentierung, Ablegung oder Partitionierung von Nukleinsäuren einer Probe oder Fragmenten davon in diskrete Kompartimente oder Partitionen (hier austauschbar als Partitionen bezeichnet), wobei jede Partition ihre eigenen Inhalte von den Inhalten anderer Partitionen getrennt hält. Eindeutige Kennungen, z. B. Barcodes, können zuvor, anschließend oder gleichzeitig an die Partitionen abgegeben werden, die die kompartimentierten oder partitionierten Nukleinsäuren einer Probe enthalten, um die spätere Zuweisung der Eigenschaften, z. B. Nukleinsäuresequenzinformationen, zu den Nukleinsäuren einer Probe zu ermöglichen, die in einem bestimmten Kompartiment enthalten sind, und insbesondere zu relativ langen Abschnitten von zusammenhängenden Nukleinsäuren einer Probe,

die ursprünglich in die Partitionen abgelegt worden sind. Diese spätere Zuweisung ermöglicht ferner eine Zuweisung zum ursprünglichen strukturellen Kontext dieser Nukleinsäuren einer Probe in der ursprünglichen Probe, da es wahrscheinlicher ist, dass Nukleinsäuren, die innerhalb der drei Dimensionen der ursprünglichen Probe nahe beieinander lagen, in derselben Partition abgelegt werden. Somit stellt eine Zuweisung von Sequenz-Reads zu den Partitionen (und den in diesen Partitionen enthaltenen Nukleinsäuren) nicht nur einen molekularen Kontext hinsichtlich des linearen Ortes entlang des ursprünglichen Nukleinsäuremoleküls, von dem dieser Sequenz-Read abgeleitet wurde, bereit, sondern stellt auch einen strukturellen Kontext des Identifizierens von Sequenz-Reads von Nukleinsäuren bereit, die im dreidimensionalen Kontext der ursprünglichen Probe räumlich nahe beieinander lagen.

**[0054]** Die Nukleinsäuren einer Probe, die in den hier beschriebenen Verfahren eingesetzt werden, stellen typischerweise eine Anzahl überlappender Teile der zu analysierenden Gesamtprobe dar, z. B. eines ganzen Chromosoms, Exons oder eines anderen großen Teils eines Genoms. Diese Nukleinsäuren einer Probe können ganze Genome, einzelne Chromosomen, Exome, Amplikons oder beliebige einer Vielzahl verschiedener Nukleinsäuren von Interesse beinhalten. Die Nukleinsäuren einer Probe werden typischerweise derart partitioniert, dass die Nukleinsäuren in den Partitionen in relativ langen Fragmenten oder Abschnitten zusammenhängender Nukleinsäuremoleküle vorliegen. Typischerweise können diese Fragmente der Nukleinsäuren einer Probe länger als 1 kb, länger als 5 kb, länger als 10 kb, länger als 15 kb, länger als 20 kb, länger als 30 kb, länger als 40 kb, länger als 50 kb, länger als 60 kb, länger als 70 kb, länger als 80 kb, länger als 90 kb oder sogar länger als 100 kb sein, was den oben beschriebenen strukturellen und molekularen Kontext mit einem größeren Bereich ergeben kann.

**[0055]** Die Nukleinsäuren einer Probe werden typischerweise auch auf einem Niveau partitioniert, bei dem eine gegebene Partition eine sehr geringe Wahrscheinlichkeit hat, zwei überlappende Fragmente eines genomischen Locus zu enthalten. Dies wird typischerweise dadurch erreicht, dass die Nukleinsäure einer Probe während des Partitionierungsprozesses in einer geringen Eingabemenge und/oder -konzentration bereitgestellt wird. Infolgedessen kann in bevorzugten Fällen eine gegebene Partition eine Anzahl von langen, aber nicht überlappenden Fragmenten der Nukleinsäuren der Ausgangsprobe enthalten. Die Nukleinsäuren einer Probe in den verschiedenen Partitionen werden dann eindeutigen Kennungen zugeordnet, wobei für jede gegebene Partition die darin enthaltenen Nukleinsäuren die gleiche eindeutige Kennung besitzen, wobei jedoch verschiedene Partitionen verschiedene eindeutige

Kennungen enthalten können. Da der Partitionierungsschritt die Probenkomponenten kleinvolumigen Partitionen oder Tröpfchen zuteilt, ist zu erkennen, dass es zum Erzielen der gewünschten Zuteilung wie oben dargelegt nicht erforderlich ist, eine wesentliche Verdünnung der Probe durchzuführen, wie dies bei höhervolumigen Prozessen, z. B. in Röhrchen oder Vertiefungen einer Multiwell-Platte, erforderlich wäre. Da bei den hier beschriebenen Systemen ferner solch ein hohes Maß an Barcode-Diversität eingesetzt wird, können diverse Barcodes einer größeren Anzahl von genomischen Äquivalenten zugeteilt werden, wie oben angegeben. Insbesondere laufen die zuvor beschriebenen Multiwell-Platten-Ansätze (siehe z. B. die veröffentlichte US-Anmeldungen Nrn. 2013-0079231 und 2013-0157870) typischerweise mit nur hundert bis einigen hundert verschiedenen Barcode-Sequenzen sowie einem Grenzverdünnungsprozess der Proben ab, um Barcodes verschiedenen Zellen/Nukleinsäuren zuweisen zu können. Von daher laufen sie im Allgemeinen mit weit weniger als 100 Zellen ab, was typischerweise ein Verhältnis von Genomen: (Barcode-Typ) in der Größenordnung von 1:10 und sicherlich weit über 1:100 liefert. Die hier beschriebenen Systeme können andererseits wegen des hohen Maßes an Barcode-Diversität, z. B. über 10.000, 100.000, 500.000 usw. verschiedene Barcode-Typen, mit Genom:(Barcode-Typ)-Verhältnissen, die in der Größenordnung von 1:50 oder weniger, 1:100 oder weniger, 1:1000 oder weniger liegen, oder sogar noch kleineren Verhältnissen ablaufen, während auch eine höhere Anzahl von Genomen (z. B. in der Größenordnung von mehr als 100 Genome pro Assay, mehr als 500 Genome pro Assay, 1000 Genome pro Assay oder sogar mehr) bei gleichzeitiger weitaus verbesserter Barcode-Diversität pro Genom geladen werden kann.

**[0056]** In weiteren Beispielen können die Oligonukleotide, die in den Teilen der Probe enthalten sind, die in die diskreten Partitionen aufgeteilt werden, mindestens eine erste und eine zweite Region umfassen. Die erste Region kann eine Barcode-Region sein, die zwischen Oligonukleotiden innerhalb einer gegebenen Partition im Wesentlichen die gleiche Barcode-Sequenz sein kann, jedoch zwischen verschiedenen Partitionen eine unterschiedliche Barcode-Sequenz sein kann und in den meisten Fällen ist. Die zweite Region kann ein N-mer sein (entweder ein zufälliges N-mer oder ein N-mer, das dazu ausgelegt ist, auf eine bestimmte Sequenz abzielen), das verwendet werden kann, um die Nukleinsäuren innerhalb der Probe innerhalb der Partitionen zu primen. In einigen Fällen, in denen das N-mer dazu ausgelegt ist, auf eine bestimmte Sequenz abzielen, kann es dazu ausgelegt sein, auf ein bestimmtes Chromosom (z. B. Chromosom 1, 13, 18 oder 21) oder eine Region eines Chromosoms, z. B. ein Exom, oder eine andere Zielregion abzielen. In einigen Fällen kann das N-mer auf

ein bestimmtes Gen oder eine bestimmte genetische Region ausgerichtet sein, wie etwa ein Gen oder eine Region, die mit einer Krankheit oder Störung (z. B. Krebs) assoziiert ist. Innerhalb der Partitionen kann eine Amplifikationsreaktion unter Verwendung des zweiten N-mers durchgeführt werden, um die Nukleinsäureprobe an verschiedenen Stellen entlang der Länge der Nukleinsäure zu primen. Infolge der Amplifikation kann jede Partition amplifizierte Produkte der Nukleinsäure enthalten, die an einem identischen oder nahezu identischen Barcode angebracht sind und die überlappende, kleinere Fragmente der Nukleinsäuren in jeder Partition darstellen können. Der Barcode kann als Marker dienen, der anzeigt, dass ein Satz von Nukleinsäuren aus derselben Partition stammt und somit möglicherweise auch aus demselben Nukleinsäurestrang stammt. Nach der Amplifikation können die Nukleinsäuren gepoolt, sequenziert und unter Verwendung eines Sequenzierungsalgorithmus ausgerichtet werden. Da kürzere Sequenz-Reads aufgrund ihrer zugeordneten Barcode-Sequenzen ausgerichtet und einem einzelnen langen Fragment der Nukleinsäure einer Probe zugewiesen werden können, können alle identifizierten Varianten dieser Sequenz einem einzelnen Stammfragment und einem einzigen Stammchromosom zugewiesen werden. Ferner kann durch Ausrichten mehrerer colokalisierter Varianten über mehrere lange Fragmente dieser chromosomale Beitrag weiter charakterisiert werden. Dementsprechend können dann Rückschlüsse auf die Phasierung bestimmter genetischer Varianten gezogen werden, ebenso wie Analysen über große Bereiche genomischer Sequenzen - zum Beispiel die Identifizierung von Sequenzinformationen über Abschnitte schlecht charakterisierter Regionen des Genoms hinweg. Solche Informationen können auch zum Identifizieren von Haplotypen nützlich sein, die im Allgemeinen ein spezifizierter Satz genetischer Varianten sind, die sich auf demselben Nukleinsäurestrang oder auf unterschiedlichen Nukleinsäuresträngen befinden. Variationen der Kopienzahl können auf diese Weise auch identifiziert werden.

**[0057]** Die beschriebenen Verfahren und Systeme bieten deutliche Vorteile gegenüber aktuellen Nukleinsäure-Sequenzierungstechnologien und den damit verbundenen Probenvorbereitungsverfahren. Ensemble-Probenvorbereitungs- und Sequenzierungsverfahren sind prädestiniert dafür, in erster Linie die Hauptbestandteile in der Probe zu identifizieren und zu charakterisieren, und sind nicht dazu ausgelegt, Minderheitsbestandteile zu identifizieren und zu charakterisieren, z. B. genetisches Material, das von einem Chromosom aus einer schlecht charakterisierten oder hochpolymorphen Region des Genoms beigetragen wird, oder Material aus einer oder wenigen Zellen oder fragmentierte DNA-Moleküle von Tumorzellen, die im Blutstrom zirkulieren und einen kleinen Prozentsatz der gesamten DNA

in der extrahierten Probe ausmachen. Zu den hier beschriebenen Verfahren gehören selektive Amplifikationsverfahren, die das genetische Material aus diesen Minderheitsbestandteilen vermehren, und die Möglichkeit, den molekularen Kontext dieses genetischen Materials beizubehalten, stellt ferner eine genetische Charakterisierung dieser Bestandteile bereit. Die beschriebenen Verfahren und Systeme bieten auch einen erheblichen Vorteil beim Nachweis von Populationen, die in einer größeren Probe vorhanden sind. Von daher sind sie besonders nützlich für eine Beurteilung von Haplotyp- und Kopienzahlvariationen - die hier offenbarten Verfahren sind auch nützlich zum Bereitstellen von Sequenzinformationen über Regionen des Genoms, die schlecht charakterisiert sind oder aufgrund während der Probenvorbereitung eingeführter Bias in einer Population von Nukleinsäurezielen schlecht repräsentiert sind.

**[0058]** Die Verwendung der hier offenbarten Barcodierungstechnik verleiht die einzigartige Möglichkeit, einen individuellen molekularen Kontext für einen gegebenen Satz genetischer Marker bereitzustellen, d. h. einen gegebenen Satz genetischer Marker (im Gegensatz zu einem einzelnen Marker) einzelnen Nukleinsäuremolekülen einer Probe zuzuweisen, und durch variantenkoordinierte Assemblierung einen breiteren oder sogar weiter reichenden erschlossenen individuellen strukturellen und molekularen Kontext unter mehreren Nukleinsäuremolekülen einer Probe und/oder zu einem bestimmten Chromosom bereitzustellen. Diese genetischen Marker können bestimmte genetische Loci beinhalten, z. B. Varianten, wie etwa SNPs, oder sie können kurze Sequenzen beinhalten. Darüber hinaus bietet die Verwendung von Barcodierung die zusätzlichen Vorteile einer erleichterten Unterscheidbarkeit zwischen Minderheitsbestandteilen und Hauptbestandteilen der gesamten aus der Probe extrahierten Nukleinsäurepopulation, z. B. zum Nachweis und zur Charakterisierung von zirkulierender Tumor-DNA im Blutstrom, und reduziert oder eliminiert auch Amplifikations-Bias während optionaler Amplifikationsschritte. Darüber hinaus verleiht eine Implementierung in einem Mikrofluidikformat die Möglichkeit, mit extrem kleinen Probenvolumina und geringen Eingabemengen an DNA zu arbeiten, sowie die Möglichkeit, schnell eine große Anzahl von Probenpartitionen (Tröpfchen) zu verarbeiten, um ein genomweites Tagging zu ermöglichen.

**[0059]** Wie zuvor beschrieben besteht ein Vorteil der hier beschriebenen Verfahren und Systeme darin, dass sie die gewünschten Ergebnisse durch die Verwendung von allgegenwärtig verfügbaren Short-Read-Sequenzierungstechnologien erzielen können. Solche Technologien haben den Vorteil, dass sie innerhalb der Forschungsgemeinschaft leicht verfügbar und weit verbreitet sind, mit Protokollen und Rea-

genziensystemen, die gut charakterisiert und hochwirksam sind. Zu diesen Short-Read-Sequenzierungstechnologien gehören diejenigen, die z. B. von Illumina, Inc. (GALLx, NextSeq, MiSeq, HiSeq, X10), Ion Torrent Division von Thermo-Fisher (Ion Proton und Ion PGM) erhältlich sind, Pyrosequenzierungsverfahren und andere.

**[0060]** Von besonderem Vorteil ist, dass bei den hier beschriebenen Verfahren und Systemen diese Short-Read-Sequenzierungstechnologien eingesetzt werden und dies mit den damit einhergehenden niedrigen Fehlerraten und hohen Durchsätzen. Insbesondere erzielen die hier beschriebenen Verfahren und Systeme die gewünschten Längen individueller molekularer Reads oder Kontexte, wie oben beschrieben, jedoch mit einzelnen Sequenzierungs-Reads, ausgenommen Mate-Pair-Extensionen, die kürzer als 1000 bp, kürzer als 500 bp, kürzer als 300 bp, kürzer als 200 bp, kürzer als 150 bp oder noch kürzer sind; und mit Sequenzierungsfehllerraten für solche Längen individueller molekularer Reads, die weniger als 5 %, weniger als 1 %, weniger als 0,5 %, weniger als 0,1 %, weniger als 0,05 %, weniger als 0,01 %, weniger als 0,005 % oder sogar weniger als 0,001 % betragen.

## II. Übersicht über den Arbeitsablauf

**[0061]** In einem beispielhaften Aspekt sehen die in der Offenbarung beschriebenen Verfahren und Systeme ein Ablegen oder Partitionieren von Proben in diskrete Partitionen vor, wobei jede Partition eine Trennung ihrer eigenen Inhalte von den Inhalten in anderen Partitionen aufrechterhält. Wie hier ausführlicher erörtert, können die Proben Proben umfassen, die von Patienten stammen, wie etwa Zell- oder Gewebeproben, die Nukleinsäuren und in bestimmten Situationen auch assoziierte Proteine enthalten können. In bestimmten Aspekten beinhalten die in den hier beschriebenen Verfahren verwendeten Proben Formalin-fixierte Paraffin-eingebettete (FFPE) Zell- und Gewebeproben und dergleichen sowie alle anderen Probentypen, bei denen das Risiko einer Probenzersetzung hoch ist.

**[0062]** Im hier verwendeten Sinne beziehen sich die Partitionen auf Behälter oder Gefäße, die eine Vielfalt unterschiedlicher Formen annehmen können, z. B. Vertiefungen, Reagenzgläser, Mikro- oder Nanovertiefungen, Durchgangslöcher oder dergleichen. In bevorzugten Aspekten sind die Partitionen jedoch innerhalb von Flüssigkeitsströmen fließfähig. Diese Gefäße können z. B. aus Mikrokapseln oder Mikrovessikeln bestehen, die eine äußere Sperrschicht aufweisen, die ein inneres Flüssigkeitszentrum oder einen Kern umgibt, oder sie können eine poröse Matrix sein, die in der Lage ist, Materialien in ihrer Matrix einzuschließen und/oder zurückzuhalten. In einem bevorzugten Aspekt können diese Partitionen jedoch

Tröpfchen einer wässrigen Flüssigkeit innerhalb einer nichtwässrigen kontinuierlichen Phase, z. B. einer Ölphase, umfassen. Eine Vielfalt unterschiedlicher Gefäße ist beispielsweise in der US-Patentanmeldung Nr. 13/966,150, eingereicht am 13. August 2013, beschrieben. Ebenso sind Emulsionssysteme zum Erzeugen stabiler Tröpfchen in kontinuierlichen nichtwässrigen oder Ölphasen z. B. in der veröffentlichten US-Patentanmeldung Nr. 2010-0105112 ausführlich beschrieben. In bestimmten Fällen sind mikrofluidische Kanalnetzwerke zum Erzeugen von Partitionen besonders geeignet, wie hier beschrieben. Zu Beispielen für solche mikrofluidische Vorrichtungen gehören diejenigen, die ausführlich in der vorläufigen US-Patentanmeldung Nr. 61/977,804, eingereicht am 4. April 2014, beschrieben sind, deren vollständige Offenbarung hiermit in ihrer Gesamtheit für alle Zwecke durch Bezugnahme aufgenommen wird. Alternative Mechanismen können ebenfalls bei der Partitionierung einzelner Zellen verwendet werden, einschließlich poröser Membranen, durch die wässrige Mischungen von Zellen in nichtwässrige Flüssigkeiten extrudiert werden. Solche Systeme sind im Allgemeinen beispielsweise von Nanomi, Inc. erhältlich.

**[0063]** Im Fall von Tröpfchen in einer Emulsion kann ein Partitionieren von Probenmaterialien in diskrete Partitionen im Allgemeinen erreicht werden, indem ein wässriger, eine Probe enthaltender Strom in eine Verbindungsstelle geleitet wird, in die auch ein nichtwässriger Strom von Partitionierungsflüssigkeit fließt, z. B. ein fluoriertes Öl, sodass wässrige Tröpfchen innerhalb des fließenden Stroms von Partitionierungsflüssigkeit erzeugt werden, wobei solche Tröpfchen die Probenmaterialien enthalten. Wie unten beschrieben beinhalten die Partitionen, z. B. Tröpfchen, typischerweise auch copartitionierte Barcode-Oligonukleotide. Die relative Menge an Probenmaterialien innerhalb einer bestimmten Partition kann eingestellt werden, indem eine Vielfalt unterschiedlicher Parameter des Systems gesteuert wird, einschließlich beispielsweise der Konzentration der Probe in dem wässrigen Strom, der Flussrate des wässrigen Stroms und/oder des nichtwässrigen Stroms und dergleichen. Die hier beschriebenen Partitionen werden häufig als extrem kleine Volumina aufweisend gekennzeichnet. Beispielsweise können die Tröpfchen im Fall tröpfchenbasierter Partitionen Gesamtvolumina haben, die kleiner als 1000 pl, kleiner als 900 pl, kleiner als 800 pl, kleiner als 700 pl, kleiner als 600 pl, kleiner als 500 pl, kleiner als 400 pl, kleiner als 300 pl, kleiner als 200 pl, kleiner als 100 pl, kleiner als 50 pl, kleiner als 20 pl, kleiner als 10 pl oder sogar kleiner als 1 pl sind. Bei Copartitionierung mit Beads ist zu erkennen, dass das Probenflüssigkeitsvolumen innerhalb der Partitionen weniger als 90 % der oben beschriebenen Volumina, weniger als 80 %, weniger als 70 %, weniger als 60 %, weniger als 50 %, weniger als 40 %, weniger als

30 %, weniger als 20 % oder sogar weniger als 10 % der oben beschriebenen Volumina betragen kann. In einigen Fällen ist die Verwendung von Partitionen mit niedrigem Reaktionsvolumen besonders vorteilhaft beim Durchführen von Reaktionen mit sehr kleinen Mengen an Ausgangsreagenzien, z. B. Eingabenukleinsäuren. Verfahren und Systeme zum Analysieren von Proben mit Nukleinsäuren mit geringer Eingabe werden in der vorläufigen US-Patentanmeldung Nr. 62/017,580 (Aktenzeichen des Anwalts 43487-727.101), eingereicht am 26. Juni 2014, vorgestellt, deren vollständige Offenbarung in ihrer Gesamtheit hiermit durch Bezugnahme aufgenommen wird.

**[0064]** In Situationen unter Beteiligung von Proben, die einem Abbau unterliegen und/oder niedrige Konzentrationen von Komponenten von Interesse enthalten, können die Proben entweder vor einer Partitionierung oder innerhalb der Partitionen weiterverarbeitet werden, um die Nukleinsäuren und/oder irgendwelche assoziierten Proteine für weitere Analyse freizusetzen. Beispielsweise werden Nukleinsäuren, die in FFPE-Proben enthalten sind, im Allgemeinen unter Verwendung von auf dem Fachgebiet bekannten Verfahren extrahiert. Zur Isolierung längerer Nukleinsäuremoleküle können solche Proben auch durch Zugabe von Organokatalysatoren zur Entfernung von Formaldehyd-Addukten aufgearbeitet werden (siehe beispielsweise Karmakar et al., (2015), *Nature Chemistry*, DOI: 10.1038/NCHEM.2307, die in ihrer Gesamtheit und insbesondere für alle Lehren in Bezug auf eine Behandlung und Verarbeitung von FFPE-Proben durch Bezugnahme aufgenommen wird.)

**[0065]** Sobald die Proben in ihre jeweiligen Partitionen eingeführt wurden, können die Nukleinsäuren einer Probe innerhalb der Partitionen einer Amplifikation unterzogen werden, um die Menge an Nukleinsäuren für nachfolgende Anwendungen (wie hier beschriebene und auf dem Fachgebiet bekannte Sequenzierungsverfahren) zu erhöhen. In bestimmten Ausführungsformen wird diese Amplifikation mit einer Bibliothek von Primern durchgeführt, die auf verschiedene Teile der genomischen Sequenz gerichtet sind, sodass die erhaltenen Amplifikationsprodukte Sequenzen von Unterabschnitten der ursprünglichen Nukleinsäuremoleküle darstellen. In Ausführungsformen, in denen ausgewählte genomische Regionen von Interesse sind, kann diese Amplifikation eine oder mehrere Runden selektiver Amplifikation beinhalten, sodass Regionen des Genoms, die für eine gezielte Abdeckung von Interesse sind, im Vergleich zu anderen Regionen des Genoms in einem höheren Anteil vorhanden sind (obwohl, wie zu erkennen ist, diese anderen Regionen des Genoms ebenfalls amplifiziert werden können, jedoch in geringerem Ausmaß, da sie für eine De-novo-Abdeckung nicht von Interesse sind). In

bestimmten Ausführungsformen stellt die Amplifikation eine mindestens 1-fache, 2-fache, 5-fache, 10-fache, 20-fache, 30-fache, 40-fache oder 50-fache Abdeckung des gesamten Genoms oder ausgewählter Regionen davon bereit. In weiteren Ausführungsformen werden alle Nukleinsäuren innerhalb einer Partition amplifiziert, wobei jedoch ausgewählte genomische Regionen gezielt amplifiziert werden, sodass mindestens 1-5-, 2-10-, 3-15-, 4-20-, 5-25-, 6-30-, 7-35-, 8-40-, 9-45- oder 10-50-mal mehr Amplikons aus diesen ausgewählten genomischen Regionen erzeugt werden als aus anderen Teilen des Genoms.

**[0066]** Gleichzeitig mit oder nach der oben beschriebenen Amplifikation werden die Nukleinsäuren (oder Fragmente davon) innerhalb der Partitionen mit eindeutigen Kennungen versehen, sodass sie nach Charakterisierung dieser Nukleinsäuren als von ihren jeweiligen Ursprüngen stammend zugewiesen werden können. Dementsprechend werden die Nukleinsäuren einer Probe typischerweise mit den eindeutigen Kennungen (z. B. Barcode-Sequenzen) copartitioniert. In besonders bevorzugten Aspekten werden die eindeutigen Kennungen in Form von Oligonukleotiden bereitgestellt, die Nukleinsäure-Barcode-Sequenzen umfassen, die an diesen Proben angebracht werden können. Die Oligonukleotide werden so partitioniert, dass zwischen Oligonukleotiden in einer gegebenen Partition die darin enthaltenen Nukleinsäure-Barcode-Sequenzen gleich sind, aber zwischen unterschiedlichen Partitionen die Oligonukleotide unterschiedliche Barcode-Sequenzen haben können und vorzugsweise haben. In beispielhaften Aspekten wird einer gegebenen Partition nur eine Nukleinsäure-Barcode-Sequenz zugeordnet, obwohl in einigen Fällen zwei oder mehr verschiedene Barcode-Sequenzen vorhanden sein können.

**[0067]** Die Nukleinsäure-Barcode-Sequenzen beinhalten typischerweise 6 bis etwa 20 oder mehr Nukleotide innerhalb der Sequenz der Oligonukleotide. Diese Nukleotide können vollständig zusammenhängend sein, d. h. in einem einzigen Abschnitt benachbarter Nukleotide, oder sie können in zwei oder mehr getrennte Teilsequenzen getrennt sein, die durch ein oder mehrere Nukleotide getrennt sind. Typischerweise können getrennte Teilsequenzen typischerweise eine Länge von etwa 4 bis etwa 16 Nukleotiden aufweisen.

**[0068]** Die copartitionierten Oligonukleotide umfassen typischerweise auch andere funktionelle Sequenzen, die bei der Prozessierung der partitionierten Nukleinsäuren nützlich sind. Diese Sequenzen beinhalten z. B. Ziel- oder zufällige/universelle Amplifikationsprimersequenzen zum Amplifizieren der genomischen DNA von den einzelnen Nukleinsäuren innerhalb der Partitionen, während die zugeordneten Barcode-Sequenzen, Sequenzierungsprimer, Hybri-

disierungs- oder Sondierungssequenzen, z. B. zur Identifizierung des Vorhandenseins von Sequenzen oder zum Pulldown von barcodierten Nukleinsäuren oder einer beliebigen von einer Anzahl anderer potentieller funktioneller Sequenzen angehängt werden. Wiederum ist die Copartitionierung von Oligonukleotiden und zugeordneten Barcodes und anderen funktionellen Sequenzen zusammen mit Probenmaterialien beispielsweise in den USSNs 14/175,935; 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; und 14/316,463 beschrieben, die hiermit in ihrer Gesamtheit für alle Zwecke und insbesondere für alle schriftlichen Beschreibungen, Figuren und Arbeitsbeispiele, die auf eine Verarbeitung von Nukleinsäuren sowie Sequenzierung und andere Charakterisierungen von genomischem Material gerichtet sind, durch Bezugnahme aufgenommen werden.

**[0069]** Kurz gesagt werden in einem beispielhaften Prozess Beads bereitgestellt, die jeweils eine große Anzahl der oben beschriebenen Oligonukleotide enthalten können, die lösbar an den Beads angebracht sind, wobei alle Oligonukleotide, die an einem bestimmten Bead angebracht sind, dieselbe Nukleinsäure-Barcode-Sequenz enthalten können, wobei jedoch eine große Anzahl unterschiedlicher Barcode-Sequenzen über die Population der verwendeten Beads hinweg repräsentiert werden kann. Typischerweise kann die Population von Beads eine Bibliothek unterschiedlicher Barcode-Sequenzen bereitstellen, die mindestens 1000 verschiedene Barcode-Sequenzen, mindestens 10.000 verschiedene Barcode-Sequenzen, mindestens 100.000 verschiedene Barcode-Sequenzen oder in einigen Fällen mindestens 1.000.000 verschiedene Barcode-Sequenzen beinhalten kann. Zusätzlich kann jedes Bead typischerweise mit einer großen Anzahl angebrachter Oligonukleotidmoleküle versehen sein. Insbesondere kann die Anzahl der Oligonukleotidmoleküle, welche die Barcode-Sequenz beinhalten, auf einem einzelnen Bead mindestens etwa 10.000 Oligonukleotide, mindestens 100.000 Oligonukleotidmoleküle, mindestens 1.000.000 Oligonukleotidmoleküle, mindestens 100.000.000 Oligonukleotidmoleküle und in einigen Fällen mindestens 1 Milliarde Oligonukleotidmoleküle betragen.

**[0070]** Die Oligonukleotide können bei Anwendung eines bestimmten Stimulus auf die Beads von den Beads freigesetzt werden. In einigen Fällen kann der Stimulus ein Photostimulus sein, z. B. durch Spaltung einer photolabilen Verknüpfung, die die Oligonukleotide freisetzen kann. In einigen Fällen kann ein thermischer Stimulus verwendet werden, wobei eine Erhöhung der Temperatur der Umgebung der Beads zur Spaltung einer Verknüpfung oder einer anderen Freisetzung der Oligonukleotide aus den Beads führen kann. In einigen Fällen kann ein chemischer Stimulus verwendet werden, der eine Verknüpfung

der Oligonukleotide an die Beads spaltet oder anderweitig zur Freisetzung der Oligonukleotide von den Beads führen kann.

**[0071]** Gemäß den hier beschriebenen Verfahren und Systemen können die Beads, welche die angegebenen Oligonukleotide beinhalten, mit den einzelnen Proben copartitioniert werden, sodass ein einzelnes Bead und eine einzelne Probe in einer einzelnen Partition enthalten sind. In einigen Fällen, in denen Einzelbead-Partitionen gewünscht sind, kann es wünschenswert sein, die relativen Durchflussraten der Flüssigkeiten so zu steuern, dass die Partitionen im Durchschnitt weniger als ein Bead pro Partition enthalten, um sicherzustellen, dass jene Partitionen, die belegt sind, überwiegend einzeln belegt sind. Ebenso kann es wünschenswert sein, die Durchflussrate zu steuern, um dafür zu sorgen, dass ein höherer Prozentsatz von Partitionen belegt ist, z. B. um nur einen kleinen Prozentsatz nicht belegter Partitionen zuzulassen. In bevorzugten Aspekten werden die Flüsse und die Kanalarchitekturen so gesteuert, dass eine gewünschte Anzahl von einfach belegten Partitionen, weniger als ein bestimmtes Maß an nicht belegten Partitionen und weniger als ein bestimmtes Maß an mehrfach belegten Partitionen gewährleistet sind.

**[0072]** Fig. 3 veranschaulicht ein bestimmtes beispielhaftes Verfahren zum Barcodieren und anschließendes Sequenzieren einer Nukleinsäure einer Probe. Zuerst kann eine Nukleinsäure umfassende Probe von einer Quelle gewonnen werden, 300, und es kann ebenfalls ein Satz von barcodierten Beads gewonnen werden, 310. Die Beads sind vorzugsweise mit Oligonukleotiden verknüpft, die eine oder mehrere Barcode-Sequenzen sowie einen Primer, wie etwa ein zufälliges N-mer oder einen anderen Primer, enthalten. Vorzugsweise sind die Barcode-Sequenzen von den barcodierten Beads freisetzbar, z. B. durch Spaltung einer Verknüpfung zwischen dem Barcode und dem Bead oder durch Abbau des darunterliegenden Beads, um den Barcode freizusetzen, oder eine Kombination der beiden. Beispielsweise können die barcodierten Beads in bestimmten bevorzugten Aspekten durch ein Mittel, wie beispielsweise ein Reduktionsmittel, abgebaut oder aufgelöst werden, um die Barcode-Sequenzen freizusetzen. In diesem Beispiel werden eine geringe Menge der Nukleinsäure umfassenden Probe 305, barcodierte Beads 315 und optional andere Reagenzien, z. B. ein Reduktionsmittel 320, kombiniert und einem Partitionieren unterzogen. Beispielsweise kann ein solches Partitionieren ein Einbringen der Komponenten in ein Tröpfchenerzeugungssystem, wie etwa eine mikrofluidische Vorrichtung 325 beinhalten. Mit Hilfe der mikrofluidischen Vorrichtung 325 kann eine Wasserin-Öl-Emulsion 330 gebildet werden, wobei die Emulsion wässrige Tröpfchen enthält, die Nukleinsäure einer Probe 305, Reduktionsmittel

320 und barcodierte Beads 315 enthalten. Das Reduktionsmittel kann die barcodierten Beads auflösen oder abbauen, wodurch die Oligonukleotide mit den Barcodes und zufällige N-mere von den Beads innerhalb der Tröpfchen 335 freigesetzt werden. Die zufälligen N-mere können dann verschiedene Regionen der Nukleinsäure einer Probe primen, was zu amplifizierten Kopien der Probe nach der Amplifikation führt, wobei jede Kopie mit einer Barcode-Sequenz 340 markiert ist. Vorzugsweise enthält jedes Tröpfchen einen Satz von Oligonukleotiden, die identische Barcode-Sequenzen und unterschiedliche zufällige N-mer-Sequenzen enthalten. Anschließend wird die Emulsion gebrochen 345 und zusätzliche Sequenzen (z. B. Sequenzen, die bestimmte Sequenzierungsverfahren unterstützen, zusätzliche Barcodes usw.) können beispielsweise über Amplifikationsverfahren 350 (z. B. PCR) hinzugefügt werden. Dann kann eine Sequenzierung durchgeführt, 355, und ein Algorithmus angewendet werden, um die Sequenzierungsdaten zu interpretieren, 360. Sequenzierungsalgorithmen sind im Allgemeinen in der Lage, zum Beispiel eine Analyse von Barcodes durchzuführen, um Sequenzierungs-Reads auszurichten und/oder die Probe zu identifizieren, zu der ein bestimmter Sequenz-Read gehört. Außerdem können diese Algorithmen, wie hier beschrieben, auch ferner verwendet werden, um die Sequenzen der Kopien ihrem molekularen Stammkontext zuzuweisen.

**[0073]** Wie zu erkennen ist, können die Proben vor oder gleichzeitig mit dem Markieren mit der Barcode-Sequenz 340 gemäß einem der hier beschriebenen Verfahren amplifiziert werden, um eine Abdeckung des gesamten Genoms oder ausgewählter Regionen des Genoms bereitzustellen. Für Ausführungsformen, in denen eine gezielte Abdeckung gewünscht ist, führt die gezielte Amplifikation im Allgemeinen zu einer größeren Population von Amplikons, die Sequenzen der Nukleinsäuren (oder Teile davon) in einer Partition darstellen, die diese ausgewählten Regionen des Genoms enthält, gegenüber Amplikons aus anderen Regionen des Genoms. Infolgedessen gibt es eine größere Anzahl der amplifizierten Kopien, die die Barcode-Sequenz 340 enthalten, innerhalb einer Partition aus den ausgewählten Regionen des Genoms als aus anderen Regionen des Genoms. In Ausführungsformen, in denen eine Amplifikation des gesamten Genoms gewünscht ist, kann die Amplifikation unter Verwendung von Primer-Bibliotheken durchgeführt werden, die dafür ausgelegt sind, Amplifikations-Bias zu minimieren und ein robustes Maß an Abdeckung über das gesamte Genom bereitzustellen.

**[0074]** Wie oben angemerkt ist zwar eine Einzelbelegung der am meisten gewünschte Zustand, es ist jedoch zu erkennen, dass mehrfach belegte Partitionen oder nicht belegte Partitionen oft vorhanden sein

können. Ein Beispiel einer mikrofluidischen Kanalstruktur zum Copartitionieren von Proben und Beads, die Barcode-Oligonukleotide umfassen, ist schematisch in **Fig. 4** veranschaulicht. Wie gezeigt sind die Kanalsegmente 402, 404, 406, 408 und 410 in Fluidverbindung an der Kanalverbindung 412 bereitgestellt. Ein wässriger Strom, der die einzelnen Proben 414 umfasst, wird durch das Kanalsegment 402 in Richtung der Kanalverbindung 412 fließen gelassen. Wie an anderer Stelle hier beschrieben können diese Proben vor dem Partitionierungsprozess in einer wässrigen Flüssigkeit suspendiert werden.

**[0075]** Gleichzeitig wird ein wässriger Strom, der die Barcode-tragenden Beads 416 umfasst, durch das Kanalsegment 404 zur Kanalverbindung 412 fließen gelassen. Eine nichtwässrige Partitionierungsflüssigkeit wird in die Kanalverbindung 412 von jedem der Seitenkanäle 406 und 408 eingeführt und die kombinierten Ströme werden in den Auslasskanal 410 fließen gelassen. Innerhalb der Kanalverbindung 412 werden die zwei kombinierten wässrigen Ströme von den Kanalsegmenten 402 und 404 kombiniert und in Tröpfchen 418 partitioniert, die copartitionierte Proben 414 und Beads 416 enthalten. Wie zuvor angemerkt ist es möglich, durch Steuern der Fließeigenschaften von jeder der Flüssigkeiten, die sich an der Kanalverbindung 412 vereinigen, sowie durch Steuern der Geometrie der Kanalverbindung die Kombination und Partitionierung zu optimieren, um ein gewünschtes Maß an Belegung von Beads, Proben oder beidem innerhalb der Partitionen 418, die erzeugt werden, zu erreichen.

**[0076]** Wie zu erkennen ist, kann eine Reihe anderer Reagenzien zusammen mit den Proben und Beads copartitioniert werden, einschließlich beispielsweise chemischer Stimuli, Nukleinsäureextensions-, Transkriptions- und/oder Amplifikationsreagenzien wie etwa Polymerasen, reverse Transkriptasen, Nukleosidtriphosphate oder NTP-Analoga, Primersequenzen und zusätzliche Cofaktoren wie etwa zweiwertige Metallionen, die bei solchen Reaktionen verwendet werden, Ligationsreaktionsreagenzien, wie etwa Ligaseenzyme und Ligationssequenzen, Farbstoffe, Labels oder andere Markierungsreagenzien. Die Primersequenzen können zufällige Primersequenzen oder zielgerichtete PCR-Primer beinhalten, die auf ein Amplifizieren ausgewählter Regionen des Genoms gerichtet sind, oder eine Kombination davon.

**[0077]** Nach Copartitionierung können die auf dem Bead angeordneten Oligonukleotide verwendet werden, um die partitionierten Proben zu barcodieren und zu amplifizieren. Ein besonders eleganter Prozess zur Verwendung dieser Barcode-Oligonukleotide beim Amplifizieren und Barcodieren von Proben ist ausführlich in den USSNs 14/175,935;



14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; und 14/316,463 beschrieben, deren vollständige Offenbarungen hiermit in ihrer Gesamtheit durch Bezugnahme aufgenommen werden. Kurz gesagt werden in einem Aspekt die auf den Beads vorhandenen Oligonukleotide mit den Proben copartitioniert und von ihren Beads in die Partition mit den Proben freigesetzt. Die Oligonukleotide enthalten typischerweise zusammen mit der Barcode-Sequenz eine Primersequenz an ihrem 5'-Ende. Die Primersequenz kann zufällig oder strukturiert sein. Zufällige Primersequenzen sollen im Allgemeinen zahlreiche unterschiedliche Regionen der Proben zufällig primen. Strukturierte Primersequenzen können eine Reihe unterschiedlicher Strukturen beinhalten, einschließlich definierter Sequenzen, die darauf abzielen, stromaufwärts einer bestimmten Zielregion der Probe zu primen, sowie Primer, die eine Art von teilweise definierter Struktur aufweisen, einschließlich, ohne darauf beschränkt zu sein, Primer, die einen prozentualen Anteil bestimmter Basen enthalten (wie etwa einen prozentualen Anteil an GC-Nukleotiden), Primer, die teilweise oder vollständig degenerierte Sequenzen enthalten, und/oder Primer, die Sequenzen enthalten, die teilweise zufällig und teilweise gemäß einem Teil der Beschreibung hier strukturiert sind. Wie zu erkennen ist, kann einer oder können mehrere der obigen Typen von zufälligen und strukturierten Primern in Oligonukleotiden in beliebiger Kombination enthalten sein.

**[0078]** Nach der Freisetzung kann der Primer-Teil des Oligonukleotids mit einer Komplementärregion der Probe annealen. Extensionsreaktionsreagenzien, z. B. DNA-Polymerase, Nukleosidtriphosphate, Cofaktoren (z. B.  $Mg^{2+}$  oder  $Mn^{2+}$  usw.), die ebenfalls mit den Proben und Beads copartitioniert werden, verlängern dann die Primersequenz unter Verwendung der Probe als Templat, um ein komplementäres Fragment zu dem Strang des Templates zu erzeugen, mit dem der Primer annealte, wobei das komplementäre Fragment das Oligonukleotid und seine zugeordnete Barcode-Sequenz beinhaltet. Das Annealing und die Extension mehrerer Primer mit und von verschiedenen Teilen der Probe kann zu einem großen Pool überlappender komplementärer Fragmente der Probe führen, von denen jedes seine eigene Barcode-Sequenz besitzt, die auf die Partition hinweist, in der es erzeugt wurde. In einigen Fällen können diese komplementären Fragmente selbst als Templat verwendet werden, das durch die in der Partition vorhandenen Oligonukleotide geprimt wird, um ein Komplement des Komplements zu erzeugen, das wiederum die Barcode-Sequenz enthält. In einigen Fällen ist dieser Replikationsprozess so konfiguriert, dass, wenn das erste Komplement dupliziert wird, es zwei komplementäre Sequenzen an oder in der Nähe seiner Termini produziert, um die Bildung einer Haarnadelstruktur oder partiellen Haarnadelstruktur zu ermöglichen, was die Möglichkeit des Moleküls redu-

ziert, als Grundlage für die Erzeugung weiterer iterativer Kopien zu dienen. Eine schematische Veranschaulichung eines Beispiels dafür ist in **Fig. 5** dargestellt.

**[0079]** Wie die Figur zeigt, werden Oligonukleotide, die eine Barcode-Sequenz enthalten, mit einer Nukleinsäure 504 einer Probe z. B. in einem Tröpfchen 502 in einer Emulsion copartitioniert. Wie hier an anderer Stelle angegeben können die Oligonukleotide 508 auf einem Bead 506 bereitgestellt werden, das mit der Nukleinsäure 504 einer Probe partitioniert wird, wobei die Oligonukleotide vorzugsweise von dem Bead 506 freisetzbar sind, wie in Feld A gezeigt. Die Oligonukleotide 508 beinhalten eine Barcode-Sequenz 512 zusätzlich zu einer oder mehreren funktionellen Sequenzen, z. B. den Sequenzen 510, 514 und 516. Zum Beispiel ist das Oligonukleotid 508 so gezeigt, dass es die Barcode-Sequenz 512 sowie die Sequenz 510 umfasst, die als Anbringungs- oder Immobilisierungssequenz für ein gegebenes Sequenzierungssystem fungieren kann, z. B. eine P5-Sequenz, die zum Anbringen in Durchflusszellen eines Hiseq- oder Miseq-Systems von Illumina verwendet wird. Wie gezeigt beinhalten die Oligonukleotide auch eine Primersequenz 516, die ein zufälliges oder zielgerichtetes N-mer zum Priming einer Replikation von Teilen der Nukleinsäure 504 einer Probe beinhalten kann. Ebenfalls in dem Oligonukleotid 508 enthalten ist eine Sequenz 514, die eine Sequenzierungs-Priming-Region bereitstellen kann, wie etwa eine „read1“- oder R1-Priming-Region, die verwendet wird, um Polymerase-vermittelte, Templatgesteuerte Sequenzierung durch Synthesereaktionen in Sequenzierungssystemen zu primen. In vielen Fällen können die Barcode-Sequenz 512, die Immobilisierungssequenz 510 und die R1-Sequenz 514 allen Oligonukleotiden gemeinsam sein, die an einem gegebenen Bead angebracht sind. Die Primersequenz 516 kann für zufällige N-mer-Primer variieren oder kann für bestimmte gezielte Anwendungen den Oligonukleotiden auf einem gegebenen Bead gemeinsam sein.

**[0080]** Basierend auf dem Vorhandensein der Primersequenz 516 sind die Oligonukleotide in der Lage, die Nukleinsäure einer Probe zu primen, wie in Feld B gezeigt, was eine Extension der Oligonukleotide 508 und 508a unter Verwendung von Polymeraseenzymen und anderen Extensionsreagenzien ermöglicht, die ebenfalls mit den Beads 506 und der Nukleinsäureprobe 504 copartitioniert sind. Wie in Feld C gezeigt werden nach Extension der Oligonukleotide, die im Falle zufälliger N-mer-Primer mit mehreren verschiedenen Regionen der Nukleinsäure 504 der Probe annealen würden, mehrere überlappende Komplemente oder Fragmente der Nukleinsäure erzeugt, z. B. die Fragmente 518 und 520. Obwohl diese Konstrukte Sequenzteile beinhalten, die komplementär zu Teilen der Nukleinsäure

der Probe sind, z. B. die Sequenzen 522 und 524, werden diese Konstrukte hier allgemein als Fragmente der Nukleinsäure 504 einer Probe mit den angehängten Barcode-Sequenzen bezeichnet. Wie zu erkennen ist, werden die replizierten Teile der Templatsequenzen, wie oben beschrieben, hier häufig als „Fragmente“ dieser Templatsequenz bezeichnet. Ungeachtet des Vorhergehenden schließt der Begriff „Fragment“ jedoch jede Darstellung eines Teils der Stammnukleinsäuresequenz, z. B. ein Templat oder eine Nukleinsäure einer Probe, ein, einschließlich jener, die durch andere Mechanismen zum Bereitstellen von Teilen der Templatsequenz erzeugt wurden, wie etwa tatsächliche Fragmentierung eines gegebenen Sequenzmoleküls, z. B. durch enzymatische, chemische oder mechanische Fragmentierung. In bevorzugten Aspekten bezeichnen jedoch Fragmente eines Templates oder einer Sequenz einer Nukleinsäure einer Probe replizierte Teile der zugrunde liegenden Sequenz oder Komplemente davon.

**[0081]** Die barcodierten Nukleinsäurefragmente können dann einer Charakterisierung ausgesetzt werden, z. B. durch Sequenzanalyse, oder sie können in dem Prozess weiter amplifiziert werden, wie in Feld D gezeigt. Beispielsweise können zusätzliche Oligonukleotide, z. B. das Oligonukleotid 508b, die ebenfalls von dem Bead 506 freigesetzt werden, die Fragmente 518 und 520 primen. Insbesondere wiederum basierend auf dem Vorhandensein des zufälligen N-mer-Primers 516b in dem Oligonukleotid 508b (der sich in vielen Fällen von anderen zufälligen N-meren in einer gegebenen Partition unterscheiden wird, z. B. Primersequenz 516), annealt das Oligonukleotid mit dem Fragment 518 und wird verlängert, um ein Komplement 526 zu mindestens einem Teil des Fragments 518 zu erzeugen, das die Sequenz 528 beinhaltet, die ein Duplikat eines Teils der Sequenz der Nukleinsäure der Probe umfasst. Die Extension des Oligonukleotids 508b wird solange fortgesetzt, bis es durch den Oligonukleotidabschnitt 508 des Fragments 518 hindurch repliziert wurde. Wie an anderer Stelle hier angemerkt und wie in Feld D veranschaulicht können die Oligonukleotide dazu konfiguriert sein, beispielsweise nach dem Replizieren durch die Sequenzen 516 und 514 des Oligonukleotids 508 hindurch, das in Fragment 518 enthalten ist, an einem gewünschten Punkt einen Stopp in der Replikation durch die Polymerase zu veranlassen. Wie hier beschrieben kann dies durch verschiedene Verfahren erreicht werden, einschließlich zum Beispiel des Einbaus verschiedener Nukleotide und/oder Nukleotidanaloga, die durch das verwendete Polymeraseenzym nicht verarbeitet werden können. Beispielsweise kann dies den Einschluss von Uracil enthaltenden Nukleotiden innerhalb der Sequenzregion 512 beinhalten, um zu verhindern, dass eine Uracilintolerante Polymerase die Replikation dieser Region beendet. Infolgedessen

wird ein Fragment 526 erzeugt, das das Oligonukleotid voller Länge 508b an einem Ende enthält, einschließlich der Barcode-Sequenz 512, der Anbringungssequenz 510, der R1-Primerregion 514 und der zufälligen N-mer-Sequenz 516b. Am anderen Ende der Sequenz wird das Komplement 516' zu dem zufälligen N-mer des ersten Oligonukleotids 508 sowie ein Komplement zu der gesamten oder einem Teil der R1-Sequenz aufgenommen, gezeigt als Sequenz 514'. Die R1-Sequenz 514 und ihr Komplement 514' sind dann in der Lage, miteinander zu hybridisieren, um eine partielle Haarnadelstruktur 528 zu bilden. Wie zu erkennen ist, würde, da sich die zufälligen N-meren zwischen verschiedenen Oligonukleotiden unterscheiden, von diesen Sequenzen und ihren Komplementen nicht erwartet werden, dass sie an einer Haarnadelbildung, z. B. Sequenz 516', die das Komplement zu dem zufälligen N-mer 516 ist, teilnehmen, und würde nicht erwartet werden, komplementär zu zufälliger N-mer-Sequenz 516b zu sein. Dies wäre bei anderen Anwendungen, z. B. zielgerichteten Primern, nicht der Fall, wo die Oligonukleotide innerhalb einer gegebenen Partition die gleichen N-meren aufweisen würden. Durch Bilden dieser partiellen Haarnadelstrukturen wird ermöglicht, dass Duplikate der ersten Ebene der Probensequenz von der weiteren Replikation ausgeschlossen werden, wodurch beispielsweise das iterative Kopieren von Kopien verhindert wird. Die partielle Haarnadelstruktur stellt auch eine nützliche Struktur für die nachfolgende Verarbeitung der erzeugten Fragmente bereit, z. B. Fragment 526.

**[0082]** Alle Fragmente aus mehreren verschiedenen Partitionen können dann zum Sequenzieren auf Hochdurchsatz-Sequenzierern, wie hier beschrieben, gepoolt werden. Da jedes Fragment hinsichtlich seiner Ursprungspartition codiert ist, kann die Sequenz dieses Fragments basierend auf dem Vorhandensein des Barcodes auf seinen Ursprung zurückgeführt werden. Dies ist in **Fig. 6** schematisch dargestellt. Wie in einem Beispiel gezeigt, sind eine Nukleinsäure 604, die von einer ersten Quelle 600 (z. B. einem einzelnen Chromosom, Nukleinsäurestrang usw.) stammt, und eine Nukleinsäure 606, die von einem anderen Chromosom 602 oder Nukleinsäurestrang abgeleitet ist, jeweils zusammen mit ihren eigenen Sätzen von Barcode-Oligonukleotiden partitioniert, wie oben beschrieben.

**[0083]** Innerhalb jeder Partition wird jede Nukleinsäure 604 und 606 dann verarbeitet, um separat einen überlappenden Satz von zweiten Fragmenten des/der ersten Fragments/e bereitzustellen, z. B. die zweiten Fragmentsätze 608 und 610. Diese Verarbeitung versieht auch die zweiten Fragmente mit einer Barcode-Sequenz, die für jedes der zweiten Fragmente gleich ist, die von einem bestimmten ersten Fragment abgeleitet sind. Wie gezeigt wird die Barcode-Sequenz für den zweiten Fragmentsatz

608 mit „1“ bezeichnet, während die Barcode-Sequenz für den Fragmentsatz 610 mit „2“ bezeichnet wird. Eine Bibliothek unterschiedlicher Barcodes kann verwendet werden, um eine große Anzahl unterschiedlicher Fragmentsätze mit unterschiedlichen Barcodes zu versehen. Es ist jedoch nicht erforderlich, dass jeder zweite Fragmentsatz aus einem anderen ersten Fragment mit unterschiedlichen Barcode-Sequenzen barcodiert wird. Tatsächlich können in vielen Fällen mehrere unterschiedliche erste Fragmente gleichzeitig verarbeitet werden, um dieselbe Barcode-Sequenz zu beinhalten. Bibliotheken unterschiedlicher Barcodes werden an anderer Stelle hier ausführlich beschrieben.

**[0084]** Die barcodierten Fragmente, z. B. aus den Fragmentsätzen 608 und 610, können dann zum Sequenzieren unter Verwendung von beispielsweise Sequenz-durch-Synthese-Technologien, erhältlich von Illumina oder Ion Torrent Division von Thermo Fisher, Inc., und dergleichen, gepoolt werden. Nach der Sequenzierung können die Sequenz-Reads aus den gepoolten Fragmenten 612 ihrem jeweiligen Fragmentsatz zugewiesen werden, z. B. wie in den aggregierten Reads 614 und 616 gezeigt, zumindest teilweise basierend auf den enthaltenen Barcodes und optional und vorzugsweise teilweise basierend auf der Sequenz des Fragments selbst. Außerdem können die Sequenz-Reads dem strukturellen Kontext der relativen Position der Nukleinsäure, von der diese Reads abgeleitet sind, in Bezug auf andere Nukleinsäuremoleküle, die sich in enger räumlicher Nähe innerhalb der ursprünglichen Probe befanden, zugewiesen werden. Die zugewiesenen Sequenz-Reads für jeden Fragmentsatz werden dann assembliert, um die assemblierte Sequenz für jedes Probenfragment, z. B. die Sequenzen 618 und 620, bereitzustellen, die wiederum ihren jeweiligen Ursprungschromosomen oder Stammnukleinsäuremolekülen (600 und 602) zugewiesen werden können. Verfahren und Systeme zum Assemblieren genomischer Sequenzen sind beispielsweise in der US-Patentanmeldung Nr. 14/752,773, eingereicht am 26. Juni 2015, beschrieben, deren vollständige Offenbarung hiermit in ihrer Gesamtheit und insbesondere für alle Lehren in Bezug auf eine Assemblierung genomischer Sequenzen durch Bezugnahme aufgenommen wird.

### III. Verfahren und Zusammensetzungen zum Beibehalten von strukturellem Kontext

**[0085]** Diese Offenbarung stellt Verfahren, Zusammensetzungen und Systeme zur Charakterisierung von genetischem Material bereit. Im Allgemeinen stellen die hier beschriebenen Verfahren, Zusammensetzungen und Systeme Verfahren zum Analysieren von Komponenten einer Probe bereit, während Informationen über den strukturellen sowie molekularen Kontext dieser Komponenten, wie sie

ursprünglich in der Probe waren, beibehalten werden. Anders ausgedrückt bezieht sich die Beschreibung hier allgemein auf die räumliche Erfassung von Nukleinsäuren in einer Probe, einschließlich Gewebeprobe, die unter Verwendung von auf dem Fachgebiet bekannten Verfahren fixiert wurden oder werden, wie etwa Formalin-fixierte, Paraffin-eingebettete Proben. Wie zu erkennen ist, kann jedes der in diesem Abschnitt beschriebenen Verfahren mit jedem der oben in den Abschnitten mit den Titeln „Übersicht“ und „Übersicht über den Arbeitsablauf“ beschriebenen Verfahren sowie mit den in nachfolgenden Abschnitten dieser Patentschrift beschriebenen Nukleinsäuresequenzierungsverfahren kombiniert werden.

**[0086]** Im Allgemeinen beziehen sich die hier offenbarten Verfahren auf ein Bestimmen und/oder Analysieren von Nukleinsäuren in einer Probe, einschließlich Genome, insbesondere des gesamten Genoms, einer Probe. Die hier beschriebenen Verfahren bieten die Möglichkeit, die Verteilung, Lokalisierung oder Expression von Nukleinsäuresequenzen (einschließlich genomischer Sequenzen) in einer Probe quantitativ oder qualitativ zu analysieren, wobei der räumliche Kontext innerhalb der Probe beibehalten wird. Die hier offenbarten Verfahren bieten einen Vorteil gegenüber herkömmlichen Verfahren zur geografischen Codierung von Nukleinsäuren in einer Probe, da Informationen über strukturellen Kontext in einem Hochdurchsatz-Verarbeitungsverfahren beibehalten werden, ohne dass eine Identifizierung bestimmter molekularer Ziele (wie etwa bestimmter Gene oder anderer Nukleinsäuresequenzen) vor der Verarbeitung der Probe für Sequenz-Reads erforderlich ist. Außerdem werden geringe Nukleinsäuremengen benötigt, was insbesondere bei Proben wie etwa FFPE-Proben von Vorteil ist, bei denen die eingesetzten Nukleinsäuren, insbesondere DNA, häufig fragmentiert oder in geringen Konzentrationen vorhanden sind.

**[0087]** Obwohl sich ein Großteil der Erörterung hier auf die Analyse von Nukleinsäuren bezieht, ist zu erkennen, dass die hier erörterten Verfahren und Systeme so angepasst werden können, dass sie auf andere Komponenten einer Probe angewendet werden können, einschließlich Proteine und andere Moleküle.

**[0088]** Wie oben erörtert bedeutet Aufrechterhalten von strukturellem Kontext, hier auch als Aufrechterhalten von geografischem Kontext und Codieren von Geografie bezeichnet, ein Verwenden von Verfahren, die das Gewinnen mehrerer Sequenz-Reads oder mehrerer Teile von Sequenz-Reads ermöglichen, die dem ursprünglichen dreidimensionalen relativen Ort dieser Sequenz-Reads innerhalb einer Probe zugewiesen werden können. Anders ausgedrückt können die Sequenz-Reads einem relativen Ort

innerhalb der Probe in Bezug auf benachbarte Nukleinsäuren (und in einigen Situationen assoziierte Proteine) in dieser Probe zugewiesen werden. Diese räumlichen Informationen sind selbst dann verfügbar, wenn sich diese benachbarten Nukleinsäuren physisch nicht innerhalb der linearen Sequenz eines einzelnen Stammnukleinsäuremoleküls befinden.

**[0089]** Im Allgemeinen beinhalten die hier beschriebenen Verfahren Analysen, bei denen eine Nukleinsäure enthaltende Probe bereitgestellt wird, wobei die Nukleinsäuren dreidimensionale Strukturen enthalten. Teile der Probe werden in diskrete Partitionen partitioniert, sodass Teile der dreidimensionalen Strukturen der Nukleinsäuren ebenfalls in die diskreten Partitionen partitioniert werden - Nukleinsäuresequenzen, die in räumlicher Nähe zueinander sind, werden daher tendenziell in die gleiche Partition partitioniert, wodurch also die dreidimensionalen Informationen dieser räumlichen Nähe beibehalten werden, selbst wenn später gewonnene Sequenz-Reads von Sequenzen stammen, die ursprünglich nicht auf demselben individuellen Stammnukleinsäuremolekül waren. Bezug nehmend auf **Fig. 1**: Wenn die Probe 101, die die Nukleinsäuremoleküle 102 und 103 und 106 enthält, in diskrete Partitionen aufgetrennt wird, sodass Teilmengen der Probe verschiedenen diskreten Partitionen zugeteilt werden, ist es wahrscheinlicher, dass aufgrund des physikalischen Abstands zwischen dem Nukleinsäuremolekül 106 und 102 und 103 die Nukleinsäuremoleküle 102 und 103 miteinander anstatt mit dem Nukleinsäuremolekül 106 in die gleiche Partition platziert werden. Von daher sind Nukleinsäuremoleküle innerhalb der gleichen diskreten Partitionen diejenigen, die in der ursprünglichen Probe in räumlicher Nähe zueinander waren. Von Nukleinsäuren innerhalb der diskreten Partitionen gewonnene Sequenzinformationen stellen somit eine Möglichkeit bereit, die Nukleinsäuren zu analysieren, beispielsweise durch Nukleinsäuresequenzierung, und diese Sequenz-Reads dem strukturellen Kontext der Stammnukleinsäuremoleküle zuzuweisen.

**[0090]** In einigen Beispielen wird eine Bibliothek von Tags auf die Probe zur räumlichen oder geografischen Codierung der Probe aufgebracht. In bestimmten Ausführungsformen sind die Tags Oligonukleotid-Tags (die „Oligonukleotid-Barcodes“ und „DNA-Barcodes“ beinhalten können), aber wie zu erkennen ist, kann jede Art von Tag, die einer Probe hinzugefügt werden kann, verwendet werden, einschließlich, aber nicht beschränkt auf Partikel, Beads, Farbstoffe, Molekularinversionssonden (MIPs) und dergleichen. Die Bibliothek von Tags kann durch einfache Diffusion oder durch aktive Prozesse, wie etwa zelluläre Prozesse innerhalb von Gewebekultur- oder Zellkulturproben, auf die Probe aufgebracht werden. Zu Zelltransportprozessen

gehören, ohne darauf beschränkt zu sein, Osmose, erleichterte Diffusion durch die Beteiligung von Zelltransportproteinen, passiver Transport und aktiver Transport durch die Beteiligung von Zelltransportproteinen und Energiezufuhr von Molekülen wie ATP. Im Allgemeinen werden die Tags so aufgebracht, dass verschiedene räumliche/geografische Orte innerhalb der Probe verschiedene Tags und/oder verschiedene Konzentrationen von Tags erhalten. Jede Weiterverarbeitung der Probe und Analyse der Nukleinsäuren innerhalb der Probe kann durch Identifizierung der Tags einem bestimmten räumlichen Kontext zugewiesen werden. Bezug nehmend auf **Fig. 1** würde zum Beispiel ein Hinzufügen einer Bibliothek von Tags zu Probe 101 dazu führen, dass die Nukleinsäuren 102 und 103 eine räumliche Nähe zu einem anderen Teil oder einer anderen Konzentration der Bibliothek von Tags als Nukleinsäure 106 aufweisen. Jegliche weitere Verarbeitung der Probe gemäß den hier beschriebenen Arbeitsabläufen würde dann dazu führen, dass die Nukleinsäuren 102 und 103 dem gleichen Teil/der gleichen Konzentration von Tags zugeordnet sind, und somit würde die Identifizierung dieser Tags anzeigen, dass sich die Nukleinsäuren 102 und 103 in der ursprünglichen Probe 101 in räumlicher Nähe zueinander befanden. Eine Identifizierung von Nukleinsäure 106 mit einem anderen Teil/einer anderen Konzentration von Tags würde zeigen, dass sich die Nukleinsäure 106 in der ursprünglichen Probe an einem anderen räumlichen Ort befand als die Nukleinsäuren 102 und 103.

**[0091]** In weiteren Beispielen werden auch partitionsspezifische Barcodes eingesetzt, sodass alle gewonnenen Sequenz-Reads der Partition zugewiesen werden können, in der sich die Stammnukleinsäuremoleküle befanden. Wie oben erörtert, identifiziert ein Zuweisen von Sequenz-Reads zu einer bestimmten Partition Nukleinsäuremoleküle, die in der Geografie der ursprünglichen Probe in räumlicher Nähe zueinander waren. Eine weitere Verwendung von Arbeitsabläufen, wie sie in **Fig. 2** dargestellt sind, liefert auch Informationen über den molekularen Kontext der Sequenz-Reads, sodass einzelne Sequenz-Reads den einzelnen Nukleinsäuremolekülen zugewiesen werden können, von denen sie stammen.

**[0092]** Um ein Tagging von Proben zu ermöglichen, können die Proben unter Verwendung beliebiger Verfahren verarbeitet werden, die auf dem Fachgebiet bekannt sind, um ein Aufbringen von exogenen Molekülen wie etwa Oligonukleotid-Tags oder anderen Labels zu ermöglichen. Beispielsweise können in Ausführungsformen, in denen FFPE-Proben verwendet werden, Tags auf die Proben aufgebracht werden, indem die Probe erhitzt wird, um ein Einbetten der Tags in die Probe zu ermöglichen, und dann könnte die Probe gekühlt und gemäß einem der hier beschriebenen Verfahren weiterverarbeitet werden,

einschließlich Aufteilung in diskrete Partitionen und weitere Analyse zur Identifizierung von Sequenzen von Nukleinsäuren in der Probe und der Tags, die sich auch in enger räumlicher Nähe zu diesen Sequenz-Reads befinden, wodurch der strukturelle Kontext dieser Sequenz-Reads beibehalten wird. Andere Probenverarbeitungsverfahren beinhalten Gewebeverarbeitungsverfahren, die extrazelluläre Matrix und/oder andere strukturelle Hindernisse unter Beibehaltung von Molekül- und Proteinelementen entfernen. Zu solchen Verfahren gehören in einigen nicht beschränkenden Beispielen das CLARITY-Verfahren sowie die Verwendung anderer Gewebereinigungs- und -markierungsverfahren, einschließlich jener, die zum Beispiel in Tomer et al., Bd. 9, Nr. 7, 2014, Nature Protocols; Kechschull et al., Neuron, Band 91, Ausgabe 5, 7. September 2016, Seiten 975-987; Chung, K. et al. Structural and molecular interrogation of intact biological systems. Nature 497, 332-337 (2013); Susaki, E. A. et al. Wholebrain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell 157, 726-739 (2014); und Lee et al., ACT-PRESTO: Rapid and consistent tissue clearing and labelling method for 3-dimensional (3D) imaging, Scientific Reports, 2016/01/11/online; B. 6, S. 18631, beschrieben sind, die hiermit jeweils in ihrer Gesamtheit für alle Zwecke und insbesondere für alle Lehren in Bezug auf die Verarbeitung von Proben zur Verwendung in strukturellen und molekularen Untersuchungsverfahren durch Bezugnahme aufgenommen werden.

**[0093]** In bestimmten Ausführungsformen werden die hier beschriebenen Verfahren in Kombination mit Bildgebungstechniken zur Identifizierung räumlicher Orte der Tags innerhalb der Probe verwendet, insbesondere für Proben, die auf Objektträgern immobilisiert sind, wie etwa FFPE-Proben. Solche Bildgebungstechniken können eine Korrelation von Sequenz-Reads mit bestimmten Orten auf den Objektträgern ermöglichen, was eine Korrelation mit anderen pathologischen/bildgebenden Studien ermöglicht, die möglicherweise mit diesen Proben durchgeführt wurden. Beispielsweise können Bildgebungstechniken verwendet werden, um eine vorläufige Identifizierung einer Pathologie bereitzustellen. Die hier beschriebenen Sequenzierungstechniken, die ferner Sequenz-Reads unter Beibehaltung von strukturellem Kontext bereitstellen, könnten mit einer solchen Bildgebungsanalyse kombiniert werden, um Sequenz-Reads mit strukturellem Kontext zu korrelieren, um diese vorläufige Identifizierung der Pathologie zu bestätigen oder weitere Informationen darüber bereitzustellen. Außerdem können die Bildgebungstechniken in Kombination mit Tags mit optischen Eigenschaften verwendet werden, sodass bestimmte Tags bestimmten Regionen der abgebildeten Probe zugeordnet werden. Sequenz-Reads, die mit diesen identifizierten Tags korreliert sind,

könnten dann aufgrund ihres Ortes mit diesen Tags weiter mit Regionen der abgebildeten Probe korreliert werden. Es ist jedoch zu erkennen, dass die hier beschriebenen Verfahren unabhängig von solchen Bildgebungstechniken sind und die Möglichkeit, strukturellen Kontext beizubehalten, ist nicht von einer Verwendung einer Bildgebungstechnik zur Bestimmung räumlicher Informationen von Nukleinsäuren in der Probe abhängig.

**[0094]** In einem beispielhaften Aspekt werden Gradienten von Oligonukleotiden in einer Probe erzeugt, um ein Koordinatensystem bereitzustellen, das durch spätere Verarbeitung durch Sequenzierung decodiert werden kann. Ein derartiger Gradient ermöglicht ein Tagging von Zellen und/oder Nukleinsäuren in der Probe mit einem Oligonukleotid oder einer Oligonukleotidkonzentration, die auf einen physischen Ort innerhalb der ursprünglichen Probe abgebildet werden kann. Dieses Koordinatensystem kann entwickelt werden, indem einer Bibliothek von Oligonukleotiden ermöglicht wird, in eine Probe zu diffundieren, und/oder indem Oligonukleotide in bestimmte Regionen der Probe injiziert werden. Bei Verwendung von Diffusion liefern Standardberechnungen der Diffusionskinetik eine Korrelation zwischen der Konzentration der Oligonukleotid-Tags und ihrem räumlichen Ort in der ursprünglichen Probe. Somit können alle anderen Nukleinsäuren, die mit dieser Konzentration von Oligonukleotid-Tags identifiziert wurden, wiederum mit einer bestimmten geografischen Region der Probe korreliert werden.

**[0095]** In weiteren beispielhaften Ausführungsformen beinhalten die Verfahren Prozesse zum Analysieren von Nukleinsäuren unter Aufrechterhaltung von strukturellem Kontext, wobei eine Bibliothek von Tags auf eine Probe aufgebracht wird, sodass verschiedene geografische Regionen der Probe verschiedene Tags erhalten. Teile der Probe, die nun ihre ursprünglichen Nukleinsäuren sowie die hinzugefügten Tags enthalten, werden dann in diskrete Partitionen aufgetrennt, sodass Teile der Bibliothek von Tags und Teile der Nukleinsäuren, die in der Probe geografisch nahe beieinander liegen, in der gleichen diskreten Partition enden. Sequenzierungsprozesse, wie die hier ausführlich beschriebenen, werden verwendet, um Sequenz-Reads von Nukleinsäuren in den diskreten Partitionen bereitzustellen. Die Tags können auch vor, nach oder gleichzeitig mit diesen Sequenzierungsprozessen identifiziert werden. Die Korrelation von Sequenz-Reads mit bestimmten Tags (oder Konzentrationen von Tags in Ausführungsformen, in denen Konzentrationsgradienten von Tags verwendet werden) hilft dabei, den räumlichen Kontext der Sequenz-Reads bereitzustellen. Wie oben erörtert stellen Ausführungsformen, in denen die für eine räumliche Codierung verwendeten Tags in Verbindung mit

partitionsspezifischer Barcodierung verwendet werden, ferner strukturellen und molekularen Kontext für die Sequenz-Reads bereit.

#### IV. Anwendungen von Verfahren und Systemen zur Nukleinsäure Sequenzierung-

**[0096]** Die hier beschriebenen Verfahren, Zusammensetzungen und Systeme sind insbesondere zugänglich für eine Verwendung in Nukleinsäure-Sequenzierungstechnologien. Zu solchen Sequenzierungstechnologien können beliebige auf dem Fachgebiet bekannte Technologien gehören, einschließlich Short-Read- und Long-Read-Sequenzierungstechnologien. In bestimmten Aspekten werden die hier beschriebenen Verfahren, Zusammensetzungen und Systeme in Short-Read-Sequenzierungstechnologien mit hoher Genauigkeit verwendet.

**[0097]** Im Allgemeinen bewerkstelligen die hier beschriebenen Verfahren und Systeme eine genomische Sequenzierung unter Verwendung von Verfahren, die die Vorteile der extrem niedrigen Sequenzierungsfehlerraten und des hohen Durchsatzes von Short-Read-Sequenzierungstechnologien aufweisen. Wie zuvor beschrieben besteht ein Vorteil der hier beschriebenen Verfahren und Systeme darin, dass sie die gewünschten Ergebnisse durch die Verwendung von überall verfügbaren Short-Read-Sequenzierungstechnologien erzielen können. Solche Technologien haben den Vorteil, dass sie innerhalb der Forschungsgemeinschaft ohne Weiteres verfügbar und weit verbreitet sind, mit Protokollen und Reagenziensystemen, die gut charakterisiert und hochwirksam sind. Zu diesen Short-Read-Sequenzierungstechnologien gehören diejenigen, die z. B. von Illumina, Inc. (GAIIx, NextSeq, MiSeq, HiSeq, X10), Ion Torrent Division von Thermo-Fisher (Ion Proton und Ion PGM) erhältlich sind, Pyrosequenzierungsverfahren sowie andere.

**[0098]** Von besonderem Vorteil ist, dass bei den hier beschriebenen Verfahren und Systemen diese Short-Read-Sequenzierungstechnologien eingesetzt werden und dies mit dem damit verbundenen niedrigen Fehlerraten. Insbesondere erzielen die hier beschriebenen Verfahren und Systeme die gewünschten Längen individueller molekularer Reads oder Kontext, wie oben beschrieben, jedoch mit einzelnen Sequenzierungs-Reads, ausgenommen Mate-Pair-Extensionen, die kürzer als 1000 bp, kürzer als 500 bp, kürzer als 300 bp, kürzer als 200 bp, kürzer als 150 bp oder noch kürzer sind; und mit Sequenzierungsfehlerraten für solche Längen individueller molekularer Reads, die weniger als 5 %, weniger als 1 %, weniger als 0,5 %, weniger als 0,1 %, weniger als 0,05 %, weniger als 0,01 %, weniger als 0,005 % oder sogar weniger als 0,001 % betragen.

**[0099]** Verfahren zum Verarbeiten und Sequenzieren von Nukleinsäuren gemäß den in der vorliegenden Anmeldung beschriebenen Verfahren und Systemen sind auch ausführlicher in den USSNs 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; und 14/316,463 beschrieben, die hiermit in ihrer Gesamtheit für alle Zwecke und insbesondere für alle schriftlichen Beschreibungen, Figuren und Arbeitsbeispiele, die auf ein Verarbeiten von Nukleinsäuren und ein Sequenzieren und andere Charakterisierungen von genomischem Material gerichtet sind, durch Bezugnahme aufgenommen werden.

**[0100]** In einigen Ausführungsformen werden die hier zum Gewinnen von Sequenzinformationen unter Beibehaltung von sowohl strukturellem als auch molekularem Kontext beschriebenen Verfahren und Systeme zum Sequenzieren eines gesamten Genoms verwendet. In einigen Ausführungsformen werden die hier beschriebenen Verfahren zum Sequenzieren von Zielregionen des Genoms verwendet. In weiteren Ausführungsformen beinhalten die hier beschriebenen Sequenzierungsverfahren eine Kombination aus tiefer Abdeckung der ausgewählten Regionen mit Linked-Reads auf niedrigerer Ebene über längere Bereiche des Genoms. Wie zu erkennen ist, bietet diese Kombination aus De-novo- und Resequenzierung einen effizienten Weg, um ein gesamtes Genom und/oder große Teile eines Genoms zu sequenzieren. Die gezielte Abdeckung von schlecht charakterisierten und/oder hochgradig polymorphen Regionen stellt ferner die Menge an Nukleinsäurematerial bereit, die für eine De-novo-Sequenzassemblierung erforderlich ist, während eine Linked-Genom-Sequenzierung über andere Regionen des Genoms eine Hochdurchsatz-Sequenzierung des Rests des Genoms aufrechterhält. Die hier beschriebenen Verfahren und Zusammensetzungen ermöglichen diese Kombination aus De-novo- und Linked-Read-Sequenzierung, da die gleiche Sequenzierungsplattform für beide Abdeckungstypen verwendet werden kann. Die Population von Nukleinsäuren und/oder Nukleinsäurefragmenten, die gemäß den hier beschriebenen Verfahren sequenziert werden, kann Sequenzen sowohl aus den genomischen Regionen für eine De-novo-Sequenzierung als auch aus den genomischen Regionen für eine Resequenzierung enthalten.

**[0101]** In bestimmten Fällen beinhalten hier beschriebene Verfahren einen Schritt, in dem das gesamte Genom oder ausgewählte Regionen davon vor der Sequenzierung amplifiziert werden. Diese Amplifikation, die im Allgemeinen unter Verwendung von auf dem Fachgebiet bekannten Verfahren durchgeführt wird (einschließlich, aber nicht beschränkt auf PCR-Amplifikation), stellt mindestens 1-fache, 2-fache, 3-fache, 4-fache, 5-fache, 6-fache,

7-fache, 8-fache, 9-fache, 10-fache, 11-fache, 12-fache, 13-fache, 14-fache, 15-fache, 16-fache, 17-fache, 18-fache, 19-fache oder 20-fache Abdeckung des gesamten Genoms oder ausgewählter Regionen davon bereit. In weiteren Ausführungsformen stellt die Amplifikation eine mindestens 1- bis 30-fache, 2- bis 25-fache, 3- bis 20-fache, 4- bis 15-fache oder 5-bis 10-fache Abdeckung des gesamten Genoms oder ausgewählter Regionen davon bereit.

**[0102]** Amplifikation zur Abdeckung des gesamten Genoms und/oder ausgewählter Zielregionen des Genoms werden im Allgemeinen durch Extension von Primern durchgeführt, die zu Sequenzen innerhalb oder in der Nähe der ausgewählten Regionen des Genoms komplementär sind. In einigen Fällen wird eine Bibliothek von Primern verwendet, die so ausgelegt ist, dass sie genomische Regionen von Interesse verkachelt - anders ausgedrückt ist die Bibliothek von Primern so ausgelegt, dass Regionen in bestimmten Entfernungen entlang des Genoms amplifiziert werden, sei es über ausgewählte Regionen oder über das gesamte Genom. In einigen Fällen werden bei der selektiven Amplifikation Primer eingesetzt, die zu allen 10, 15, 20, 25, 50, 100, 200, 250, 500, 750, 1000 oder 10000 Basen entlang der ausgewählten Regionen des Genoms komplementär sind. In weiteren Beispielen ist die gekachelte Bibliothek von Primern so ausgelegt, dass sie eine Mischung von Entfernungen erfasst - diese Mischung kann eine zufällige Mischung von Entfernungen sein oder intelligent derart ausgelegt sein, dass bestimmte Teile oder prozentuale Anteile der ausgewählten Regionen durch verschiedene Primerpaare amplifiziert werden. In weiteren Ausführungsformen sind die Primerpaare derart ausgelegt, dass jedes Paar etwa 1-5 %, 2-10 %, 3-15 %, 4-20 %, 5-25 %, 6-30 %, 7-35 %, 8-40 %, 9-45 % oder 10-50 % einer zusammenhängenden Region eines ausgewählten Teils des Genoms amplifiziert.

**[0103]** In bestimmten Ausführungsformen und gemäß eines beliebigen Teils der obigen Beschreibungen erfolgt die Amplifikation über eine Region des Genoms, die mindestens 3 Megabasenpaare (Mb) lang ist. In weiteren Ausführungsformen wird eine ausgewählte Region des Genoms gemäß einem der hier beschriebenen Verfahren selektiv amplifiziert, und diese ausgewählte Region ist mindestens 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 oder 10 Mb lang. In weiteren Ausführungsformen ist die ausgewählte Region des Genoms etwa 2-20, 3-18, 4-16, 5-14, 6-12 oder 7-10 Mb lang. Die Amplifikation kann über diese Regionen erfolgen, indem ein einzelnes Primerpaar verwendet wird, das komplementär zu Sequenzen an den Enden oder nahe den Enden dieser Regionen ist. In anderen Ausführungsformen wird die Amplifikation mit einer Bibliothek von Primerpaaren durchgeführt, die die Länge der Region verkacheln, sodass regelmäßige Seg-

mente, zufällige Segmente oder eine Kombination verschiedener Segmententfernungen entlang der Region amplifiziert werden, wobei das Ausmaß der Abdeckung gemäß der obigen Beschreibung ist.

**[0104]** In einigen Ausführungsformen enthalten die Primer, die bei der selektiven Amplifikation ausgewählter Regionen des Genoms verwendet werden, Uracile, sodass die Primer selbst nicht amplifiziert werden.

**[0105]** Unabhängig von der verwendeten Sequenzierungsplattform wird die Sequenzierung von Nukleinsäuren im Allgemeinen und gemäß einem beliebigen der hier beschriebenen Verfahren typischerweise auf eine Weise durchgeführt, die den strukturellen und molekularen Kontext von Sequenz-Reads oder Teilen von Sequenz-Reads bewahrt. Damit ist gemeint, dass mehrere Sequenz-Reads oder mehrere Teile von Sequenz-Reads dem relativen räumlichen Ort innerhalb der ursprünglichen Probe in Bezug auf andere Nukleinsäuren (struktureller Kontext) und/oder dem Ort innerhalb der linearen Sequenz eines einzelnen Stammmoleküls einer Nukleinsäure (molekularer Kontext) zugewiesen werden können.

**[0106]** Wie zu erkennen ist, wird das einzelne Stammmolekül einer Nukleinsäure, wenngleich es eine beliebige einer Vielfalt von Längen haben kann, in bevorzugten Aspekten ein relativ langes Molekül sein, das eine Bewahrung eines weitreichenden molekularen Kontexts ermöglicht. Insbesondere ist das einzelne Stammmolekül vorzugsweise wesentlich länger als die typische Länge eines Short-Reads, z. B. länger als 200 Basen, und ist häufig mindestens 1000 Basen lang oder länger, 5000 Basen lang oder länger, 10.000 Basen lang oder länger, 20.000 Basen lang oder länger, 30.000 Basen lang oder länger, 40.000 Basen lang oder länger, 50.000 Basen lang oder länger, 60.000 Basen oder länger, 70.000 Basen lang oder länger, 80.000 Basen lang oder länger, 90.000 Basen lang oder länger oder 100.000 Basen lang oder länger und in einigen Fällen 1 Megabase lang oder länger.

**[0107]** Im Allgemeinen beinhalten erfindungsgemäße Verfahren Schritte wie in **Fig. 2** veranschaulicht, die einen schematischen Überblick über erfindungsgemäße Verfahren bereitstellt, die hier ausführlicher erörtert werden. Wie zu erkennen ist, ist das in **Fig. 2** skizzierte Verfahren ein Ausführungsbeispiel, das nach Bedarf und wie hier beschrieben geändert oder modifiziert werden kann.

**[0108]** Wie in **Fig. 2** gezeigt beinhalten die hier beschriebenen Verfahren in den meisten Beispielen einen Schritt, in dem Proben partitioniert werden (202). Vor diesem Partitionierungsschritt kann es einen optionalen Schritt (201) geben, in dem Nukle-

insäuren in der Probe verknüpft werden, um Sequenzregionen anzubringen, die sich in enger räumlicher Nähe zueinander befinden. Im Allgemeinen wird jede Partition, die Nukleinsäuren aus genomischen Regionen von Interesse enthält, einer Art Fragmentierungsprozess unterzogen, und der ursprüngliche molekulare Kontext der Fragmente wird im Allgemeinen beibehalten (203), üblicherweise durch Barcodierung der Fragmente, die spezifisch für die Partition sind, in der sie enthalten sind. Jede Partition kann in einigen Beispielen mehr als eine Nukleinsäure beinhalten und wird in einigen Fällen mehrere hundert Nukleinsäuremoleküle enthalten - in Situationen, in denen sich mehrere Nukleinsäuren innerhalb einer Partition befinden, wird jeder bestimmte Locus des Genoms im Allgemeinen durch eine einzelne individuelle Nukleinsäure vor der Barcodierung repräsentiert. Wie oben erörtert, können barcodierte Fragmente von Schritt 203 unter Verwendung beliebiger auf dem Fachgebiet bekannter Verfahren erzeugt werden - in einigen Beispielen sind Oligonukleotide die Proben innerhalb der unterschiedlichen Partitionen. Solche Oligonukleotide können Zufallssequenzen umfassen, die dazu bestimmt sind, zahlreiche verschiedene Regionen der Proben zufällig zu primen, oder sie können eine bestimmte Primersequenz umfassen, die darauf abzielt, stromaufwärts einer Zielregion der Probe zu primen. In weiteren Beispielen enthalten diese Oligonukleotide auch eine Barcode-Sequenz, sodass der Replikationsprozess auch das resultierende replizierte Fragment der ursprünglichen Nukleinsäure einer Probe barcodiert. Extensionsreaktionsreagenzien, z. B. DNA-Polymerase, Nukleosidtriphosphate, Cofaktoren (z. B.  $Mg^{2+}$  oder  $Mn^{2+}$  usw.), die ebenfalls in den Partitionen enthalten sind, verlängern dann die Primersequenz unter Verwendung der Probe als Templat, um ein komplementäres Fragment zu dem Strang des Templates zu erzeugen, an den sich der Primer angelagert hat, und das komplementäre Fragment beinhaltet das Oligonukleotid und seine zugeordnete Barcode-Sequenz. Das Annealing und die Extension mehrerer Primer an verschiedenen Teilen der Probe kann zu einem großen Pool überlappender komplementärer Fragmente der Probe führen, von denen jedes seine eigene Barcode-Sequenz besitzt, die auf die Partition hinweist, in der es erzeugt wurde. In einigen Fällen können diese komplementären Fragmente selbst als Templat verwendet werden, das durch die in der Partition vorhandenen Oligonukleotide geprimt wird, um ein Komplement des Komplements zu erzeugen, das wiederum die Barcode-Sequenz enthält. In weiteren Beispielen ist dieser Replikationsprozess derart konfiguriert, dass er, wenn das erste Komplement dupliziert wird, zwei komplementäre Sequenzen an oder in der Nähe seiner Termini erzeugt, um die Bildung einer Haarnadelstruktur oder partiellen Haarnadelstruktur zu ermöglichen, was die Möglichkeit des

Moleküls verringert, die Basis für die Herstellung weiterer iterativer Kopien zu sein.

**[0109]** Zurückkehrend zu dem in **Fig. 2** beispielhaft dargestellten Verfahren können die barcodierten Fragmente, sobald die partitionsspezifischen Barcodes an die kopierten Fragmente angebracht sind, optional dann gepoolt werden (204). Die gepoolten Fragmente werden dann sequenziert (205), und die Sequenzen der Fragmente werden ihrem ursprünglichen molekularen Kontext zugewiesen (206), sodass die Zielregionen von Interesse sowohl identifiziert als auch mit diesem ursprünglichen molekularen Kontext verknüpft werden. Ein Vorteil der hier beschriebenen Verfahren und Systeme besteht darin, dass ein Anbringen eines partitions- oder probenspezifischen Barcodes an den kopierten Fragmenten vor einem Anreichern der Fragmente für genomische Zielregionen den ursprünglichen molekularen Kontext dieser Zielregionen bewahrt, wodurch sie ihrer ursprünglichen Partition und somit ihrer Stammnukleinsäure einer Probe zugewiesen werden können.

**[0110]** Zusätzlich zu dem obigen Arbeitsablauf können genomische Zielregionen für weitere Analysen, insbesondere Sequenzierung, unter Verwendung von Verfahren, zu denen sowohl Chip-basierte als auch lösungsbasierte Einfangverfahren gehören, angereichert, isoliert oder getrennt, d. h. „pulled down“ werden. Bei solchen Verfahren werden Sonden eingesetzt, die zu den genomischen Regionen von Interesse oder zu Regionen nahe oder benachbart zu den genomischen Regionen von Interesse komplementär sind. Beispielsweise werden beim Hybrid- (oder Chipbasierten) Einfangen Mikroarrays, die Fangsonden (normalerweise einzelsträngige Oligonukleotide) mit Sequenzen enthalten, die zusammengekommen die Region von Interesse abdecken, an einer Oberfläche fixiert. Genomische DNA wird fragmentiert und kann weiter verarbeitet werden, wie etwa Endreparatur, um Blunt-Ends zu erzeugen, und/oder Hinzufügen zusätzlicher Merkmale, wie etwa universeller Priming-Sequenzen. Diese Fragmente werden an die Sonden auf dem Mikroarray hybridisiert. Nichthybridisierte Fragmente werden gewaschen, und die gewünschten Fragmente werden eluiert oder anderweitig auf der Oberfläche zur Sequenzierung oder für andere Analysen verarbeitet, und somit wird die Population von Fragmenten, die auf der Oberfläche verbleibt, mit Fragmenten angereichert, die die Zielregionen von Interesse enthalten (z. B. die Regionen, die die Sequenzen, die zu den in den Fangsonden enthaltenen komplementär sind). Die angereicherte Population von Fragmenten kann ferner unter Verwendung beliebiger auf dem Fachgebiet bekannter Amplifikationstechnologien amplifiziert werden. Beispielhafte Verfahren für solche gezielten Pulldown-Anreicherungsverfahren sind in USSN 62/072,164, eingereicht am 29. Oktober 2014, beschrieben, das hiermit in seiner Gesamt-



heit für alle Zwecke und insbesondere für alle Lehren in Bezug auf gezielte Pulldown-Anreicherungsverfahren und Sequenzierungsverfahren, einschließlich aller schriftlichen Beschreibungen, Figuren und Beispiele durch Bezugnahme aufgenommen wird.

**[0111]** In einigen Beispielen ist es wünschenswert, sich eher auf ausgewählte Regionen des Genoms als auf die Sequenzierung des gesamten Genoms zu konzentrieren. Die hier beschriebenen Verfahren sind für solche Analysen besonders zugänglich, da die Möglichkeit, auf Teilgruppen des Genoms abzuzeilen, selbst wenn diese Teilgruppen sich in großen linearen Entfernungen, aber möglicherweise in unmittelbarer Nähe im dreidimensionalen Kontext der ursprünglichen Probe befinden, ein vorteilhaftes Merkmal dieser Verfahren ist. In einigen Aspekten beinhalten Verfahren zur Abdeckung ausgewählter Regionen des Genoms Verfahren, bei denen die diskreten Partitionen, die Nukleinsäuremoleküle und/oder Fragmente davon aus diesen ausgewählten Regionen enthalten, selbst zur weiteren Verarbeitung ausgelesen werden. Wie zu erkennen ist, kann dieses Auslesen der diskreten Partitionen in beliebiger Kombination mit anderen Verfahren der selektiven Amplifikation und/oder einem gezielten Pulldown von hier beschriebenen genomischen Regionen von Interesse, insbesondere in beliebiger Kombination mit den Schritten des oben beschriebenen Arbeitsablaufs, erfolgen.

**[0112]** Im Allgemeinen beinhalten Verfahren zum Auslesen der diskreten Partitionen Schritte, in denen Partitionen, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms enthalten, von Partitionen getrennt werden, die keine Sequenzen aus diesen Teilen des Genoms enthalten. Diese Verfahren beinhalten die Schritte des Bereitstellens einer Population, die mit Sequenzen der Fragmente angereichert ist, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms innerhalb der diskreten Partitionen umfassen, die Sequenzen aus diesen Teilen des Genoms enthalten. Eine solche Anreicherung wird im Allgemeinen durch die Verwendung einer gerichteten PCR-Amplifikation der Fragmente innerhalb der diskreten Partitionen erreicht, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms beinhalten, um eine Population zu erzeugen. Diese gerichtete PCR-Amplifikation erzeugt somit Amplikons, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms umfassen. In bestimmten Ausführungsformen sind diese Amplikons an ein nachweisbares Label gebunden, das in einigen nicht beschränkenden Ausführungsformen ein fluoreszierendes Molekül enthalten kann. Im Allgemeinen erfolgt eine solche Anbringung derart, dass nur diejenigen Amplikons, die aus den Fragmenten erzeugt wurden, die den einen oder die mehreren ausge-

wählten Teile des Genoms enthalten, an das nachweisbare Label angebracht werden. In einigen Ausführungsformen erfolgt das Anbringen der nachweisbaren Labels während der selektiven Amplifikation des einen oder der mehreren ausgewählten Teile des Genoms. Zu solchen nachweisbaren Labels können in weiteren Ausführungsformen, ohne darauf beschränkt zu sein, fluoreszierende Labels, elektrochemische Labels, magnetische Beads und Nanopartikel gehören. Diese Anbringung des nachweisbaren Labels kann unter Verwendung von auf dem Fachgebiet bekannten Verfahren erzielt werden. In weiteren Ausführungsformen werden diskrete Partitionen, die Fragmente enthalten, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms umfassen, basierend auf Signalen ausgelesen, die von den nachweisbaren Labels emittiert werden, die an den Amplikons innerhalb dieser Partitionen angebracht sind.

**[0113]** In weiteren Ausführungsformen beinhalten die Schritte des Auslesens diskreter Partitionen, die ausgewählte Teile des Genoms enthalten, von denen, die solche Sequenzen nicht enthalten, die folgenden Schritte: (a) Bereitstellen von genomischem Ausgangsmaterial; (b) Verteilen einzelner Nukleinsäuremoleküle aus dem genomischen Ausgangsmaterial in diskrete Partitionen, sodass jede diskrete Partition ein erstes einzelnes Nukleinsäuremolekül enthält; (c) Bereitstellen einer Population innerhalb mindestens einiger der diskreten Partitionen, die mit Sequenzen der Fragmente angereichert ist, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms umfassen; (d) Anbringen einer gemeinsamen Barcode-Sequenz an die Fragmente innerhalb jeder diskreten Partition, sodass jedes der Fragmente der diskreten Partition zugewiesen werden kann, in der es enthalten war; (e) Trennen diskreter Partitionen, die Fragmente enthalten, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms umfassen, von diskreten Partitionen, die keine Fragmente enthalten, die den einen oder die mehreren ausgewählten Teile des Genoms umfassen; (f) Gewinnen von Sequenzinformationen aus den Fragmenten, die mindestens einen Teil des einen oder der mehreren ausgewählten Teile des Genoms umfassen, wodurch ein oder mehrere Zielteile der genomischen Probe unter Beibehaltung von molekularem Kontext sequenziert werden. Wie zu erkennen ist, kann Schritt (a) eines solchen Verfahrens mehr als ein einzelnes Nukleinsäuremolekül beinhalten.

**[0114]** In weiteren Ausführungsformen und gemäß dem Obigen werden die diskreten Partitionen vor Gewinnen von Sequenzinformationen aus den Fragmenten kombiniert, und die Fragmente werden gepoolt. In weiteren Ausführungsformen wird der Gewinnungsschritt von Sequenzinformationen aus den Fragmenten derart durchgeführt, dass der struk-

turelle und molekulare Kontext der Sequenzen der Fragmente aufrechterhalten bleibt, sodass das Identifizieren ferner ein Identifizieren von Fragmenten umfasst, die von Nukleinsäuren abgeleitet sind, die sich physisch in enger Nähe innerhalb der ursprünglichen Probe befinden und/oder sich auf denselben ersten einzelnen Nukleinsäuremolekülen befinden. In weiteren Ausführungsformen beinhaltet dieses Gewinnen von Sequenzinformationen eine Sequenzierungsreaktion, die aus der Gruppe bestehend aus Folgendem ausgewählt ist: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen. In weiteren Ausführungsformen ist die Sequenzierungsreaktion eine Short-Read-Sequenzierungsreaktion mit hoher Genauigkeit.

**[0115]** In weiteren Ausführungsformen und gemäß dem Obigen umfassen die diskreten Partitionen Tröpfchen in einer Emulsion. In weiteren Ausführungsformen entsprechen die barcodierten Fragmente innerhalb der diskreten Partitionen einer etwa 1- bis 10-fachen Abdeckung des einen oder der mehreren ausgewählten Teile des Genoms. In weiteren Ausführungsformen entsprechen die barcodierten Fragmente innerhalb der diskreten Partitionen einer etwa 2- bis 5-fachen Abdeckung des einen oder der mehreren ausgewählten Teile des Genoms. In weiteren Ausführungsformen entsprechen die barcodierten Fragmente der Amplikons innerhalb der diskreten Partitionen mindestens einer 1-fachen Abdeckung des einen oder der mehreren ausgewählten Teile des Genoms. In weiteren Ausführungsformen entsprechen die barcodierten Fragmente innerhalb der diskreten Partitionen einer mindestens 2- oder 5-fachen Abdeckung des einen oder der mehreren ausgewählten Teile des Genoms.

**[0116]** Zusätzlich zu der Bereitstellung der Möglichkeit, Sequenzinformationen aus ausgewählten Regionen des Genoms zu gewinnen, können die hier beschriebenen Verfahren und Systeme auch andere Charakterisierungen von genomischem Material bereitstellen, einschließlich, ohne darauf beschränkt zu sein, Haplotyp-Phasierung, Identifizierung von strukturellen Variationen und Kopienzahlvariationen, wie in den USSNs 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; und 14/316,463 ausführlich beschrieben, die hiermit in ihrer Gesamtheit für alle Zwecke und insbesondere für alle schriftlichen Beschreibungen, Figuren und Arbeitsbeispiele, die auf die Charakterisierung von genomischem Material gerichtet sind, durch Bezugnahme aufgenommen werden.

**[0117]** In einem Aspekt und in Verbindung mit beliebigen der oben und nachfolgend hier beschriebenen Verfahren sorgen die hier beschriebenen Verfahren und Systeme für die Kompartimentierung, das Ablegen oder die Partitionierung von Nukleinsäureproben oder Fragmenten davon in diskrete Kompartimente

oder Partitionen (hier austauschbar als Partitionen bezeichnet), wobei jede Partition ihre eigenen Inhalte von den Inhalten anderer Partitionen getrennt hält. Eindeutige Kennungen, z. B. Barcodes, können zuvor, anschließend oder gleichzeitig an die Partitionen abgegeben werden, die die kompartimentierten oder partitionierten Nukleinsäuren einer Probe enthalten, um die spätere Zuweisung der Eigenschaften, z. B. Nukleinsäuresequenzinformationen, zu den Nukleinsäuren einer Probe zu ermöglichen, die in einem bestimmten Kompartiment enthalten sind, und insbesondere zu relativ langen Abschnitten von zusammenhängenden Nukleinsäuren einer Probe, die ursprünglich in die Partitionen abgelegt worden sind.

**[0118]** Die in den hier beschriebenen Verfahren eingesetzten Nukleinsäuren einer Probe stellen typischerweise eine Anzahl überlappender Teile der zu analysierenden Gesamtprobe dar, z. B. ein ganzes Chromosom, Exom oder einen anderen großen genomischen Teil. Diese Nukleinsäuren einer Probe können ganze Genome, einzelne Chromosomen, Exome, Amplikons oder beliebige einer Vielfalt verschiedener Nukleinsäuren von Interesse beinhalten. Die Nukleinsäuren einer Probe werden typischerweise so partitioniert, dass die Nukleinsäuren in den Partitionen in relativ langen Fragmenten oder Abschnitten zusammenhängender Nukleinsäuremoleküle vorliegen. Typischerweise können diese Fragmente der Nukleinsäuren einer Probe länger als 1 kb, länger als 5 kb, länger als 10 kb, länger als 15 kb, länger als 20 kb, länger als 30 kb, länger als 40 kb, länger als 50 kb, länger als 60 kb, länger als 70 kb, länger als 80 kb, länger als 90 kb oder sogar länger als 100 kb sein, was den oben beschriebenen weitreichenden molekularen Kontext ergibt.

**[0119]** Die Nukleinsäuren einer Probe werden typischerweise auch auf einem Niveau partitioniert, wobei eine gegebene Partition eine sehr geringe Wahrscheinlichkeit hat, dass sie zwei überlappende Fragmente der Nukleinsäure einer Ausgangsprobe enthält. Dies wird typischerweise dadurch erreicht, dass die Nukleinsäure einer Probe während des Auftrennungsprozesses in einer geringen Eingabemenge und/oder -konzentration bereitgestellt wird. Infolgedessen kann eine gegebene Partition in bevorzugten Fällen eine Reihe von langen, jedoch nicht überlappenden Fragmenten der Nukleinsäuren der Ausgangsprobe enthalten. Die Nukleinsäuren einer Probe in den verschiedenen Partitionen werden dann eindeutigen Kennungen zugeordnet, wobei bei jeder gegebenen Partition die darin enthaltenen Nukleinsäuren die gleiche eindeutige Kennung besitzen, wobei jedoch verschiedene Partitionen verschiedene eindeutige Kennungen enthalten können. Da der Partitionierungsschritt die Probenkomponenten sehr kleinvolumigen Partitionen oder Tröpfchen zuteilt, ist außerdem zu erkennen, dass es zum

Erzielen der gewünschten Zuteilung wie oben dargelegt nicht erforderlich ist, eine wesentliche Verdünnung der Probe durchzuführen, wie dies bei höher-volumigen Prozessen, z. B. in Röhrchen oder Vertiefungen einer Multiwell-Platte, erforderlich wäre. Da bei den hier beschriebenen Systemen ferner solch ein hohes Maß an Barcode-Diversität eingesetzt wird, können diverse Barcodes einer größeren Anzahl von genomischen Äquivalenten zugeteilt werden, wie oben angegeben. Insbesondere laufen die zuvor beschriebenen Multiwell-Platten-Ansätze (siehe z. B. die veröffentlichten US-Anmeldungen Nrn. 2013-0079231 und 2013-0157870) typischerweise mit nur hundert bis einigen hundert verschiedenen Barcode-Sequenzen sowie einem Grenzverdünnungsprozess der Proben ab, um Barcodes verschiedenen Zellen/Nukleinsäuren zuweisen zu können. Von daher laufen sie im Allgemeinen mit weit weniger als 100 Zellen ab, was typischerweise ein Verhältnis von Genomen:(Barcode-Typ) in der Größenordnung von 1:10 und sicherlich weit über 1:100 liefert. Die hier beschriebenen Systeme können andererseits wegen des hohen Maßes an Barcode-Diversität, z. B. über 10.000, 100.000, 500.000, 600.000, 700.000 usw. verschiedene Barcode-Typen, mit Genom:(Barcode-Typ)-Verhältnissen, die in der Größenordnung von 1:50 oder weniger, 1:100 oder weniger, 1:1000 oder weniger liegen, oder sogar noch kleineren Verhältnissen ablaufen, während auch eine höhere Anzahl von Genomen (z. B. in der Größenordnung von mehr als 100 Genome pro Assay, mehr als 500 Genome pro Assay, 1000 Genome pro Assay oder sogar mehr) bei gleichzeitiger weitaus verbesserter Barcode-Diversität pro Genom geladen werden kann.

**[0120]** Häufig wird die Probe mit einem Satz von Oligonukleotid-Tags kombiniert, die vor dem Partitionierungsschritt lösbar an Beads angebracht sind. Verfahren zur Barcodierung von Nukleinsäuren sind auf dem Fachgebiet bekannt und hier beschrieben. In einigen Beispielen werden Verfahren eingesetzt, wie in Amini et al., 2014, Nature Genetics, Advance Online Publication) beschrieben, das hiermit in seiner Gesamtheit für alle Zwecke und insbesondere für alle Lehren in Bezug auf das Anbringen von Barcodes oder anderen Oligonukleotid-Tags an Nukleinsäuren durch Bezugnahme aufgenommen wird. In weiteren Beispielen können die Oligonukleotide mindestens eine erste und eine zweite Region umfassen. Die erste Region kann eine Barcode-Region sein, die zwischen Oligonukleotiden innerhalb einer gegebenen Partition im Wesentlichen die gleiche Barcode-Sequenz sein kann, jedoch zwischen verschiedenen Partitionen eine unterschiedliche Barcode-Sequenz sein kann und in den meisten Fällen ist. Die zweite Region kann ein N-mer sein (entweder ein zufälliges N-mer oder ein N-mer, das dazu ausgelegt ist, auf eine bestimmte Sequenz abzielen), das verwendet werden kann, um die Nukleinsäuren

innerhalb der Probe innerhalb der Partitionen zu primen. In einigen Fällen, in denen das N-mer dazu ausgelegt ist, auf eine bestimmte Sequenz abzielen, kann es dazu ausgelegt sein, auf ein bestimmtes Chromosom (z. B. Chromosom 1, 13, 18 oder 21) oder eine Region eines Chromosoms, z.B. ein Exom, oder eine andere Zielregion abzielen. Wie hier erörtert, kann das N-mer auch für ausgewählte Regionen des Genoms ausgelegt sein, die eher schlecht charakterisiert oder hochgradig polymorph sind oder von der Referenzsequenz divergieren. In einigen Fällen kann das N-mer dazu ausgelegt sein, dass es auf ein bestimmtes Gen oder eine genetische Region abzielt, wie etwa ein Gen oder eine Region, die mit einer Krankheit oder Störung (z. B. Krebs) assoziiert ist. Innerhalb der Partitionen kann eine Amplifikationsreaktion unter Verwendung des zweiten N-mers durchgeführt werden, um die Nukleinsäureprobe an verschiedenen Stellen entlang der Länge der Nukleinsäure zu primen. Infolge der Amplifikation kann jede Partition amplifizierte Produkte der Nukleinsäure enthalten, die an einem identischen oder nahezu identischen Barcode angebracht sind und die überlappende, kleinere Fragmente der Nukleinsäuren in jeder Partition darstellen können. Der Barcode kann als Marker dienen, der anzeigt, dass ein Satz von Nukleinsäuren aus derselben Partition stammt und somit möglicherweise auch aus demselben Nukleinsäurestrang stammt. Nach der Amplifikation können die Nukleinsäuren gepoolt, sequenziert und unter Verwendung eines Sequenzierungsalgorithmus ausgerichtet werden. Da kürzere Sequenz-Reads aufgrund ihrer zugeordneten Barcode-Sequenzen ausgerichtet und einem einzelnen langen Fragment der Nukleinsäure einer Probe zugewiesen werden können, können alle identifizierten Varianten dieser Sequenz einem einzelnen Stammfragment und einem einzigen Stammchromosom zugewiesen werden. Ferner kann durch Ausrichten mehrerer colokalisierter Varianten über mehrere lange Fragmente dieser chromosomale Beitrag weiter charakterisiert werden. Dementsprechend können dann Rückschlüsse auf die Phasierung bestimmter genetischer Varianten gezogen werden, ebenso wie Analysen über große Bereiche genomischer Sequenzen - beispielsweise die Identifizierung von Sequenzinformationen über Abschnitte schlecht charakterisierter Regionen des Genoms hinweg. Solche Informationen können auch zum Identifizieren von Haplotypen nützlich sein, die im Allgemeinen ein spezifizierter Satz genetischer Varianten sind, die sich auf demselben Nukleinsäurestrang oder auf unterschiedlichen Nukleinsäuresträngen befinden. Variationen der Kopienzahl können auf diese Weise auch identifiziert werden.

**[0121]** Die beschriebenen Verfahren und Systeme bieten deutliche Vorteile gegenüber aktuellen Nukleinsäure-Sequenzierungstechnologien und den damit verbundenen Probenvorbereitungsverfahren.

Ensemble-Probenvorbereitungs- und Sequenzierungsverfahren sind prädestiniert dafür, in erster Linie die Hauptbestandteile in der Probe zu identifizieren und zu charakterisieren, und sind nicht dazu ausgelegt, Minderheitsbestandteile, z. B. genetisches Material, das von einem Chromosom aus einer schlecht charakterisierten oder hochgradig polymorphen Region des Genoms beigetragen wird, oder Material aus einer oder wenigen Zellen oder fragmentierte DNA-Moleküle von Tumorzellen, die im Blutstrom zirkulieren und einen kleinen Prozentsatz der gesamten DNA in der extrahierten Probe ausmachen, zu identifizieren und zu charakterisieren. Zu den hier beschriebenen Verfahren gehören selektive Amplifikationsverfahren, die das genetische Material aus diesen Minderheitsbestandteilen vermehren, und die Möglichkeit, den molekularen Kontext dieses genetischen Materials beizubehalten, stellt ferner eine genetische Charakterisierung dieser Bestandteile bereit. Die beschriebenen Verfahren und Systeme bieten auch einen erheblichen Vorteil beim Nachweis von Populationen, die in einer größeren Probe vorhanden sind. Von daher sind sie besonders nützlich für eine Beurteilung von Haplotyp- und Kopienzahlvariationen - die hier offenbarten Verfahren sind auch nützlich zum Bereitstellen von Sequenzinformationen für Sequenzen, die sich innerhalb des dreidimensionalen Raums der ursprünglichen Probe und der ursprünglichen Nukleinsäuremoleküle, von denen diese Sequenzen abgeleitet wurden, in räumlicher Nähe zueinander befanden.

**[0122]** Die Verwendung der hier offenbarten Barcodierungstechnik verleiht die einzigartige Möglichkeit, individuellen strukturellen und molekularen Kontext für Sequenzen und Regionen des Genoms bereitzustellen. Solche Regionen des Genoms können einen gegebenen Satz von genetischen Markern beinhalten, d. h. Zuweisen eines gegebenen Satzes von genetischen Markern (im Gegensatz zu einem einzelnen Marker) zu einzelnen Nukleinsäuremolekülen einer Probe und durch variantenkoordinierte Assemblierung, um einen breiteren oder sogar weiter reichenden erschlossenen individuellen molekularen Kontext unter mehreren Nukleinsäuremolekülen einer Probe und/oder zu einem spezifischen Chromosom bereitzustellen. Diese genetischen Marker können bestimmte genetische Loci beinhalten, z. B. Varianten, wie etwa SNPs, oder sie können kurze Sequenzen beinhalten. Darüber hinaus bietet die Verwendung von Barcodierung die zusätzlichen Vorteile einer erleichterten Unterscheidbarkeit zwischen Minderheitsbestandteilen und Hauptbestandteilen der gesamten aus der Probe extrahierten Nukleinsäurepopulation, z. B. zum Nachweis und zur Charakterisierung von zirkulierender Tumor-DNA im Blutstrom, und reduziert oder eliminiert auch Amplifikations-Bias während optionaler Amplifikationsschritte. Darüber hinaus verleiht eine Implementierung in einem Mikrofluidikformat die Möglichkeit,

mit extrem kleinen Probenvolumina und geringen Eingabemengen an DNA zu arbeiten, sowie die Möglichkeit, schnell eine große Anzahl von Probenpartitionen (Tröpfchen) zu verarbeiten, um ein genomweites Tagging zu ermöglichen.

**[0123]** Wie oben angemerkt stellen die hier beschriebenen Verfahren und Systeme individuellen strukturellen und molekularen Kontext für kurze Sequenz-Reads von längeren Nukleinsäuren bereit. Im hier verwendeten Sinne bezieht sich struktureller Kontext auf den Ort von Sequenzen innerhalb des dreidimensionalen Raums ihrer Stammnukleinsäuremoleküle innerhalb der ursprünglichen Probe. Wie oben erörtert sind Chromosomen, wenngleich das Genom häufig als linear angesehen wird, nicht starr, und die räumliche Entfernung zwischen zwei genomischen Loci korreliert nicht notwendigerweise mit ihrer Entfernung entlang des Genoms - genomische Regionen, die entlang der linearen Sequenz durch mehrere Megabasen getrennt sind, können im dreidimensionalen Raum unmittelbar nebeneinander liegen. Durch Beibehaltung der Informationen über die ursprüngliche räumliche Nähe von Sequenz-Reads stellen die hier beschriebenen Verfahren und Zusammensetzungen eine Möglichkeit bereit, Sequenz-Reads weitreichenden genomischen Wechselwirkungen zuzuweisen.

**[0124]** Ebenso ist die Erhaltung von individuellem molekularem Kontext mit den hier beschriebenen Verfahren möglich, da Sequenzkontext über den jeweiligen Sequenz-Read hinaus bereitgestellt wird, z. B. in Bezug auf benachbarte oder proximale Sequenzen, die in dem Sequenz-Read selbst nicht enthalten sind, und von daher typischerweise so sind, dass sie weder ganz noch teilweise in einem kurzen Sequenz-Read, z. B. einem Read von etwa 150 Basen oder etwa 300 Basen für gepaarte Reads, enthalten wären. In besonders bevorzugten Aspekten stellen die Verfahren und Systeme einen weitreichenden Sequenzkontext für kurze Sequenz-Reads bereit. Ein solcher weitreichender Kontext beinhaltet eine Beziehung oder Verknüpfung eines gegebenen Sequenz-Reads mit Sequenz-Reads, die innerhalb einer Entfernung von länger als 1 kb, länger als 5 kb, länger als 10 kb, länger als 15 kb, länger als 20 kb, länger als 30 kb, länger als 40 kb, länger als 50 kb, länger als 60 kb, länger als 70 kb, länger als 80 kb, länger als 90 kb oder sogar länger als 100 kb oder länger zueinander liegen. Durch Bereitstellen eines weitreichenden individuellen molekularen Kontexts stellen die erfindungsgemäßen Verfahren und Systeme auch einen viel längeren erschlossenen molekularen Kontext bereit. Sequenzkontext, wie hier beschrieben, kann Kontext mit geringerer Auflösung beinhalten, z. B. von der Abbildung der kurzen Sequenz-Reads auf die einzelnen längeren Moleküle oder Contigs verknüpfter Moleküle, sowie den Sequenzkontext mit höherer

Auflösung, z. B. von weitreichender Sequenzierung von großen Teilen der längeren einzelnen Moleküle, z. B. mit zusammenhängenden bestimmten Sequenzen von einzelnen Molekülen, wobei solche bestimmte Sequenzen länger als 1 kb, länger als 5 kb, länger als 10 kb, länger als 15 kb, länger als 20 kb, länger als 30 kb, länger als 40 kb, länger als 50 kb, länger als 60 kb, länger als 70 kb, länger als 80 kb, länger als 90 kb oder sogar länger als 100 kb sind. Wie beim Sequenzkontext kann die Zuweisung kurzer Sequenzen zu längeren Nukleinsäuren, z. B. sowohl einzelnen langen Nukleinsäuremolekülen als auch Sammlungen verknüpfter Nukleinsäuremoleküle oder Contigs, sowohl die Abbildung von kurzen Sequenzen auf längere Nukleinsäureabschnitte, um Sequenzkontext auf hohem Niveau bereitzustellen, als auch das Bereitstellen assemblierter Sequenzen aus den kurzen Sequenzen durch diese längeren Nukleinsäuren beinhalten.

**[0125]** Die hier beschriebenen Verfahren, Zusammensetzungen und Systeme ermöglichen eine Charakterisierung von weitreichenden Wechselwirkungen über das Genom hinweg sowie eine Charakterisierung assoziierter Proteine und anderer Moleküle in einer Probe. Wie die Organisation von Proteinen auf höherer Ebene erzeugt das Biegen und Falten von DNA und Chromatin funktionell signifikante Strukturen auf unterschiedlichsten Ebenen. Auf niedriger Ebene ist hinreichend bekannt, dass DNA häufig um Proteine wie Histone gewickelt ist, um eine Struktur zu schaffen, die als das Nukleosom bekannt ist. Diese Nukleosomen sind zu größeren Chromatinfasern zusammengepackt, und es wurde die Vermutung angestellt, dass das Packungsmuster von zellulären Prozessen wie der Transkription beeinflusst wird. Funktionelle Strukturen existieren auch auf höherer Ebene: Regionen, die durch viele Megabasen entlang der linearen Sequenz des Genoms getrennt sind, können im dreidimensionalen Raum unmittelbar benachbart sein. Solche weitreichenden Wechselwirkungen zwischen genomischen Loci können eine Rolle bei funktionellen Eigenschaften spielen: Beispielsweise funktionieren Gen-Enhancer-, -Silencer- und -Isolator-Elemente möglicherweise alle über große genomische Entfernungen hinweg, und ihre primäre Wirkungsweise könnte eine direkte physikalische Assoziierung mit Zielgenen, nichtcodierenden RNAs und/oder regulatorischen Elementen beinhalten. Weitreichende Wechselwirkungen sind nicht auf Elemente beschränkt, die in cis vorliegen, d. h. auf demselben Chromosom, sondern können auch zwischen in trans vorliegenden genomischen Loci, d. h. auf unterschiedlichen Chromosomen, auftreten. Das Vorhandensein von weitreichenden Wechselwirkungen kann Bemühungen erschweren, die Wege zu verstehen, die zelluläre Prozesse regulieren, da die wechselwirkenden regulatorischen Elemente in großer genomischer Entfernung von einem Zielgen liegen können, sogar auf

einem anderen Chromosom. Im Fall von Onkogenen und anderen krankheitsassoziierten Genen kann eine Identifizierung von weitreichenden genetischen Regulatoren von großem Nutzen bei der Identifizierung der genomischen Varianten sein, die für den Krankheitszustand und den Prozess, durch den der Krankheitszustand hervorgerufen wird, verantwortlich sind. Somit bietet die Möglichkeit, strukturellen und molekularen Kontext gemäß den hier beschriebenen Verfahren beizubehalten, einen Weg, weitreichende genomische Wechselwirkungen zu identifizieren und auch damit assoziierte Proteine zu charakterisieren.

**[0126]** Die hier beschriebenen Verfahren sind besonders nützlich zur Charakterisierung von Nukleinsäuren aus einer FFPE-Gewebeprobe, einschließlich einer historischen FFPE-Gewebeprobe. FFPE-Proben stellen im Allgemeinen eine Herausforderung für die Charakterisierung von Nukleinsäuren dar, da die Nukleinsäuren häufig fragmentiert oder anderweitig abgebaut sind, was die Menge an Informationen einschränken kann, die mit herkömmlichen Verfahren gewonnen werden kann. Die Informationen zum strukturellen und molekularen Kontext, die in den hier beschriebenen Verfahren beibehalten werden, bieten eine einzigartige Gelegenheit mit solchen Proben, da diese Kontextinformationen Charakterisierungen von weitreichenden genomischen Wechselwirkungen sogar für abgebaute Proben bereitstellen können, da diese weitreichenden Informationen durch Short-Read-Sequenzierungstechnologien zugänglich sind. Anwendungen von FFPE-Nukleinsäurecharakterisierungen beinhalten Vergleiche von Sequenzen aus einer oder mehreren historischen Proben mit Sequenzen aus einer Probe von einem Subjekt, z. B. einem Krebspatienten, um diagnostische oder prognostische Informationen bereitzustellen. Beispielsweise kann der Status eines oder mehrerer molekularer Marker in einer historischen Probe mit einem oder mehreren Behandlungsergebnissen korreliert werden, und die Korrelation eines Behandlungsergebnisses mit dem Status eines molekularen Markers in einer oder mehreren historischen Proben kann verwendet werden, um Behandlungsergebnisse für das Subjekt vorherzusagen, z. B. für einen Krebspatienten. Diese Vorhersagen können die Grundlage für eine Entscheidung sein, ob dem Subjekt eine medikamentöse Behandlungsoption empfohlen werden soll oder nicht.

## V. Proben

**[0127]** Wie zu erkennen ist, können die hier erörterten Verfahren und Systeme verwendet werden, um Sequenzinformationen aus jeder Art von genomischem Material zu gewinnen. Solches genomisches Material kann aus einer einem Patienten entnommenen Probe gewonnen werden. Zu beispielhaften Proben und Arten von genomischem Material zur Ver-

wendung in den hier erörterten Verfahren und Systemen gehören, ohne darauf beschränkt zu sein, Polynukleotide, Nukleinsäuren, Oligonukleotide, zirkulierende zellfreie Nukleinsäure, zirkulierende Tumorzellen (CTC), Nukleinsäurefragmente, Nukleotide, DNA, RNA, Peptidpolynukleotide, komplementäre DNA (cDNA), doppelsträngige DNA (dsDNA), einzelsträngige DNA (ssDNA), Plasmid-DNA, Cosmid-DNA, chromosomale DNA, genomische DNA (gDNA), virale DNA, bakterielle DNA, mtDNA (mitochondriale DNA), ribosomale RNA, zellfreie DNA, zellfreie fötale DNA (cffDNA), mRNA, rRNA, tRNA, nRNA, siRNA, snRNA, snoRNA, scaRNA, microRNA, dsRNA, virale RNA und dergleichen. Zusammenfassend können die verwendeten Proben in Abhängigkeit von den speziellen Verarbeitungsanforderungen variieren.

**[0128]** In bestimmten Aspekten beinhalten Proben zur Verwendung in der vorliegenden Erfindung Formalin-fixierte, Paraffin-eingebettete (FFPE) Zell- und Gewebeproben und dergleichen, einschließlich aller anderen Probentypen, bei denen das Risiko eines Abbaus der Probe hoch ist. Zu anderen Arten von fixierten Proben gehören, ohne darauf beschränkt zu sein, Proben, die unter Verwendung von Folgendem fixiert wurden: Acrolein, Glyoxal, Osmiumtetroxid, Carbodiimid, Quecksilberchlorid, Zinksalzen, Pikrinsäure, Kaliumdichromat, Ethanol, Methanol, Aceton und/oder Essigsäure.

**[0129]** In weiteren Ausführungsformen umfassen die Proben zur Verwendung in den hier beschriebenen Verfahren und Systemen Kernmatrix. „Kernmatrix“ bezieht sich auf eine beliebige Zusammensetzung, die Nukleinsäuren und Protein umfasst. Die Nukleinsäuren können zu Chromosomen organisiert sein, wobei die Proteine (d. h. zum Beispiel Histone) mit den Chromosomen assoziiert werden können, die eine regulatorische Funktion aufweisen.

**[0130]** Die hier bereitgestellten Verfahren und Systeme sind besonders nützlich für Nukleinsäuresequenzierungsanwendungen, bei denen die Ausgangsnukleinsäuren (z. B. DNA, mRNA usw.) - oder Ausgangsziel-nukleinsäuren - in kleinen Mengen vorhanden sind, oder wenn Nukleinsäuren, die zur Analyse bestimmt sind, in einem relativ geringen Anteil der gesamten Nukleinsäuren in einer Probe vorhanden sind. In einem Aspekt stellt die vorliegende Offenbarung ein Verfahren zum Analysieren von Nukleinsäuren bereit, wobei die eingegebenen Nukleinsäuremoleküle in einer Menge von weniger als 50 Nanogramm (ng) vorhanden sind. In weiteren Ausführungsformen liegen die Nukleinsäuremoleküle in einer Eingangsmenge von weniger als 40 ng vor. In einigen Ausführungsformen beträgt die Menge weniger als 20 ng. In einigen Ausführungsformen beträgt die Menge weniger als 10 ng. In einigen Ausführungsformen beträgt die Menge weniger als 5 ng. In

einigen Ausführungsformen beträgt die Menge weniger als 1 ng. In einigen Ausführungsformen beträgt die Menge weniger als 0,1 ng. Verfahren zum Isolieren und Analysieren von Nukleinsäuren, bei denen die anfängliche Eingabemenge eine kleine Menge ist, sind beispielsweise in USSN 14/752,602, eingereicht am 26. Juni 2015, beschrieben, die hiermit in ihrer Gesamtheit für alle Zwecke und insbesondere für alle Lehren, die sich auf die Isolierung und Charakterisierung von Nukleinsäuren beziehen, die aus Proben stammen, in denen die Nukleinsäuren in kleinen Mengen vorhanden sind, durch Bezugnahme aufgenommen wird.

**[0131]** Wie zu erkennen ist, können Proben zu jedem Zeitpunkt während der hier beschriebenen Verfahren unter Verwendung von auf dem Fachgebiet bekannten Verfahren verarbeitet werden. Beispielsweise können Proben vor dem Partitionieren oder nachdem die Probe in diskrete Partitionen partitioniert wurde, verarbeitet werden.

**[0132]** In bestimmten Ausführungsformen werden die Proben verarbeitet, um sicherzustellen, dass längere Nukleinsäurestränge erhalten bleiben. In Ausführungsformen, in denen FFPE-Proben verwendet werden, können solche Proben einer Verarbeitung unterzogen werden, um Formaldehyd-Addukte zu entfernen, um die Nukleinsäureausbeuten zu verbessern. Solche Verarbeitungsverfahren können in einem nicht beschränkenden Beispiel die Verwendung von wasserlöslichen Organokatalysatoren beinhalten, um die Umkehrung von Formaldehyd-Addukten von RNA- und DNA-Basen zu beschleunigen, wie in Karmakar et al., (2015), Nature Chemistry, DOI: 10.1038/NCHEM.2307 beschrieben, die hiermit in ihrer Gesamtheit und insbesondere für alle Lehren in Bezug auf die Behandlung und Verarbeitung von FFPE-Proben durch Bezugnahme aufgenommen wird.

**[0133]** Eine beliebige Substanz, die Nukleinsäure umfasst, kann die Quelle einer Probe sein. Die Substanz kann eine Flüssigkeit sein, z. B. eine biologische Flüssigkeit. Eine flüssige Substanz kann Blut, Nabelschnurblut, Speichel, Urin, Schweiß, Serum, Sperma, Vaginalflüssigkeit, Magen- und Verdauungsflüssigkeit, Spinalflüssigkeit, Plazentaflüssigkeit, Höhlenflüssigkeit, Augenflüssigkeit, Serum, Muttermilch, Lymphflüssigkeit oder Kombinationen davon sein. Die Substanz kann fest sein, beispielsweise ein biologisches Gewebe. Die Substanz kann normales gesundes Gewebe, erkranktes Gewebe oder eine Mischung aus gesundem und erkranktem Gewebe umfassen. In einigen Fällen kann die Substanz Tumore umfassen. Tumore können gutartig (kein Krebs) oder bösartig (Krebs) sein. Zu nicht beschränkenden Beispielen für Tumoren gehören: Fibrosarkom, Myxosarkom, Liposarkom, Chondrosarkom, osteogenes Sarkom, Chordom,

Angiosarkom, Endotheliosarkom, Lymphangiosarkom, Lymphangioendotheliosarkom, Synoviom, Mesotheliom, Ewing-Sarkom, Leiomyosarkom, Rhabdomyosarkom, Karzinome des Magen-Darm-Systems, Darmkrebs, Pankreaskarzinome, Brustkrebs, Karzinome des Urogenitalsystems, Eierstockkrebs, Prostatakrebs, Plattenepithelkarzinom, Basalzellkarzinom, Adenokarzinom, Schweißdrüsenkarzinom, Talgdrüsenkarzinom, Papillenkarcinom, papilläre Adenokarzinome, Zystadenokarzinom, Markkarzinom, Bronchialkarzinom, Nierenzellkarzinom, Hepatom, Gallengangskarzinom, Chorionkarzinom, Seminom, embryonales Karzinom, Wilms-Tumor, Gebärmutterhalskrebs, Karzinome des endokrinen Systems, Hodentumor, Lungenkarzinom, kleinzelliges Lungenkarzinom, nichtkleinzelliges Lungenkarzinom, Blasenkarzinom, Epithelkarzinom, Gliom, Astrozytom, Medulloblastom, Kraniopharyngeom, Ependymom, Pinealom, Hämangioblastom, Akustikusneurinom, Oligodendrogliom, Meningiom, Melanom, Neuroblastom, Retinoblastom oder Kombinationen davon. Die Substanz kann mit verschiedenen Arten von Organen in Verbindung stehen. Zu nicht beschränkenden Beispielen für Organe gehören Gehirn, Leber, Lunge, Niere, Prostata, Eierstock, Milz, Lymphknoten (einschließlich Mandeln), Schilddrüse, Bauchspeicheldrüse, Herz, Skelettmuskel, Darm, Kehlkopf, Speiseröhre, Magen oder Kombinationen davon. In einigen Fällen kann die Substanz eine Vielfalt von Zellen umfassen, einschließlich, ohne darauf beschränkt zu sein: eukaryotische Zellen, prokaryotische Zellen, Pilzzellen, Herzzellen, Lungenzellen, Nierenzellen, Leberzellen, Bauchspeicheldrüsenzellen, Reproduktionszellen, Stammzellen, induzierte pluripotente Stammzellen, Magen-Darm-Zellen, Blutzellen, Krebszellen, Bakterienzellen, aus einer menschlichen Mikrobiomprobe isolierte Bakterienzellen usw. In einigen Fällen kann die Substanz Inhalt einer Zelle umfassen, wie beispielsweise den Inhalt einer einzelnen Zelle oder den Inhalt mehrerer Zellen. Verfahren und Systeme zum Analysieren einzelner Zellen werden z. B. in USSN 14/752,641, eingereicht am 26. Juni 2015, bereitgestellt, deren vollständige Offenbarung hiermit in ihrer Gesamtheit durch Bezugnahme aufgenommen wird.

**[0134]** Proben können von verschiedenen Subjekten gewonnen werden. Ein Subjekt kann ein lebendes Subjekt oder ein totes Subjekt sein. Zu Beispielen für Subjekte gehören, ohne darauf beschränkt zu sein: Menschen, Säuger, nichtmenschliche Säuger, Nagetiere, Amphibien, Reptilien, Hunde, Katzen, Rinder, Pferde, Ziegen, Schafe, Hühner, Schafe, Mäuse, Kaninchen, Insekten, Schnecken, Mikroben, Bakterien, Parasiten oder Fische. In einigen Fällen kann das Subjekt ein Patient sein, der eine Krankheit oder Störung hat, vermutlich hat oder dafür anfällig ist. In einigen Fällen kann das Subjekt eine schwangere Frau sein. In einigen Fällen kann das Subjekt

eine normale gesunde schwangere Frau sein. In einigen Fällen kann das Subjekt eine schwangere Frau sein, die Gefahr läuft, mit einem Baby mit bestimmten Geburtsfehlern schwanger zu sein.

**[0135]** Eine Probe kann von einem Subjekt durch beliebige auf dem Fachgebiet bekannte Mittel gewonnen werden. Beispielsweise kann eine Probe von einem Subjekt durch Zugang zum Kreislaufsystem (z. B. intravenös oder intraarteriell über eine Spritze oder eine andere Vorrichtung), Entnehmen einer ausgeschiedenen biologischen Probe (z. B. Speichel, Sputum, Urin, Stuhl usw.), chirurgisches (z. B. Biopsie) Entnehmen einer biologischen Probe (z. B. intraoperative Proben, postoperative Proben usw.), Abstrich (z. B. Wangenabstrich, Oropharynxabstrich) oder Pipettieren gewonnen werden.

## VI. Ausführungsformen

**[0136]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit. Solche Verfahren beinhalten die folgenden Schritte: (a) Bereitstellen einer Nukleinsäuren enthaltenden Probe, wobei die Nukleinsäuren dreidimensionale Strukturen umfassen; (b) Auftrennen von Teilen der Probe in diskrete Partitionen, sodass Teile der dreidimensionalen Nukleinsäurestrukturen ebenfalls in die diskreten Partitionen aufgetrennt werden; (c) Gewinnen von Sequenzinformationen von den Nukleinsäuren, wodurch Nukleinsäuren unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0137]** In einigen Ausführungsformen beinhalten die Sequenzinformationen aus dem Gewinnungsschritt (c) eine Identifizierung von Nukleinsäuren, die sich in räumlicher Nähe zueinander befinden.

**[0138]** In allen Ausführungsformen stellt der Gewinnungsschritt (c) Informationen über intrachromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereit.

**[0139]** In allen Ausführungsformen stellt der Gewinnungsschritt (c) Informationen über Chromosomenkonformationen bereit.

**[0140]** In allen Ausführungsformen werden vor dem Auftrennungsschritt (b) mindestens einige der dreidimensionalen Strukturen verarbeitet, um verschiedene Teile der Nukleinsäuren zu verknüpfen, die innerhalb der dreidimensionalen Strukturen nahe beieinander liegen.

**[0141]** In allen Ausführungsformen ist die Probe eine Formalin-fixierte Probe in Paraffin.

**[0142]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Auftrennungsschritt (b) nicht aus der Probe isoliert.

**[0143]** In allen Ausführungsformen umfassen die diskreten Partitionen Beads.

**[0144]** In allen Ausführungsformen sind die Beads Gel-Beads.

**[0145]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Gewinnungsschritt (c) innerhalb der diskreten Partitionen zum Bilden einer Vielzahl barcodierter Fragmente barcodiert, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen, sodass die Barcodes Nukleinsäuren aus einer gegebenen Partition identifizieren.

**[0146]** In allen Ausführungsformen umfasst der Gewinnungsschritt (c) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

**[0147]** In allen Ausführungsformen umfasst die Probe eine Tumorseite.

**[0148]** In allen Ausführungsformen umfasst die Probe eine Mischung aus Tumor- und normalen Zellen.

**[0149]** In allen Ausführungsformen umfasst die Probe eine Kernmatrix.

**[0150]** In allen Ausführungsformen umfassen die Nukleinsäuren RNA.

**[0151]** In allen Ausführungsformen beträgt die Menge an Nukleinsäuren in der Probe weniger als 5, 10, 15, 20, 25, 30, 35, 40, 45 oder 50 ng/ml.

**[0152]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit, die die folgenden Schritte beinhalten:

(a) Bilden verknüpfter Nukleinsäuren innerhalb der Probe, sodass räumlich benachbarte Nukleinsäuresegmente verknüpft werden; (b) Verarbeiten der verknüpften Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten; (c) Ablegen der Vielzahl von Ligationsprodukten in diskrete Partitionen; (d) Barcodieren der Ligationsprodukte innerhalb der diskreten Partitionen zum Bilden einer Vielzahl barcodierter Fragmente, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen,

wodurch jedes Fragment der verknüpften Nukleinsäure zugeordnet wird, von der es abgeleitet ist; (e) Gewinnen von Sequenzinformationen aus der Vielzahl barcodierter Fragmente, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0153]** In weiteren Ausführungsformen beinhaltet der Verarbeitungsschritt (b) eine Blunt-End-Ligation unter Bedingungen, die eine intramolekulare Ligation begünstigen, sodass die räumlich benachbarten Nukleinsäuresegmente innerhalb desselben Moleküls ligiert werden.

**[0154]** In allen Ausführungsformen umfassen die Bedingungen, die eine intramolekulare Ligation begünstigen, ein Verdünnen der Probe, um die Konzentration der Nukleinsäuren auf unter 10 ng/μl zu verringern.

**[0155]** In allen Ausführungsformen werden die Nukleinsäuren vor Schritt (a) nicht aus der Probe isoliert.

**[0156]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Schritt Bilden (a) immunpräzipitiert, sodass assoziierte DNA-bindende Proteine an den Nukleinsäuren gebunden bleiben.

**[0157]** In allen Ausführungsformen umfassen die Partitionen Beads.

**[0158]** In allen Ausführungsformen sind die Beads Gel-Beads.

**[0159]** In allen Ausführungsformen umfasst die Probe eine Tumorseite.

**[0160]** In allen Ausführungsformen umfasst die Probe eine Mischung aus Tumor- und normalen Zellen.

**[0161]** In allen Ausführungsformen beinhaltet der Verarbeitungsschritt eine Umkehrung der Verknüpfung nach dem Bilden der Ligationsprodukte.

**[0162]** In allen Ausführungsformen stellt der Gewinnungsschritt (e) Informationen über intrachromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereit.

**[0163]** In allen Ausführungsformen stellt der Gewinnungsschritt (e) Informationen über Chromosomenkonformationen bereit.

**[0164]** In allen Ausführungsformen sind die Chromosomenkonformationen Krankheitszuständen zugeordnet.



**[0165]** In allen Ausführungsformen führt der Verarbeitungsschritt zu Ligationsprodukten, die Nukleinsäuren umfassen, die sich ursprünglich in enger räumlicher Nähe in der Probe befanden.

**[0166]** In allen Ausführungsformen umfasst der Gewinnungsschritt (e) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

**[0167]** In allen Ausführungsformen ist die Sequenzierungsreaktion eine Short-Read-Sequenzierungsreaktion mit hoher Genauigkeit.

**[0168]** In allen Ausführungsformen beinhaltet der Bildungsschritt (a) ein Vernetzen von Nukleinsäuren in der Probe.

**[0169]** In allen Ausführungsformen führt der Bildungsschritt (a) zu kovalenten Verknüpfungen zwischen räumlich benachbarten Nukleinsäuresegmenten.

**[0170]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit, die die folgenden Schritte beinhalten:

(a) Bilden verknüpfter Nukleinsäuren innerhalb der Probe, sodass räumlich benachbarte Nukleinsäuresegmente verknüpft werden; (b) Ablegen der verknüpften Nukleinsäuren in diskrete Partitionen; (c) Verarbeiten der verknüpften Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten; (d) Barcodieren der Ligationsprodukte innerhalb der diskreten Partitionen zum Bilden einer Vielzahl barcodierter Fragmente, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen, wodurch jedes Fragment der verknüpften Nukleinsäure zugeordnet wird, von der es abgeleitet ist; (e) Gewinnen von Sequenzinformationen aus der Vielzahl barcodierter Fragmente, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0171]** In weiteren Ausführungsformen beinhaltet der Verarbeitungsschritt (c) eine Blunt-End-Ligation unter Bedingungen, die eine intramolekulare Ligation begünstigen, sodass die räumlich benachbarten Nukleinsäuresegmente innerhalb desselben Moleküls ligiert werden.

**[0172]** In allen Ausführungsformen ist die Probe eine Formalin-fixierte Probe in Paraffin.

**[0173]** In allen Ausführungsformen umfasst die Probe eine Kernmatrix.

**[0174]** In allen Ausführungsformen umfassen die Nukleinsäuren RNA.

**[0175]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Schritt (a) nicht aus der Probe isoliert.

**[0176]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Bildungsschritt (a) immunpräzipitiert, sodass assoziierte DNA-bindende Proteine an den Nukleinsäuren gebunden bleiben.

**[0177]** In allen Ausführungsformen umfassen die Partitionen Beads.

**[0178]** In allen Ausführungsformen sind die Beads Gel-Beads.

**[0179]** In allen Ausführungsformen umfasst die Probe eine Tumorable.

**[0180]** In allen Ausführungsformen umfasst die Probe eine Mischung aus Tumor- und normalen Zellen.

**[0181]** In allen Ausführungsformen führt der Verarbeitungsschritt (c) zu Ligationsprodukten, die Nukleinsäuren umfassen, die sich ursprünglich in enger räumlicher Nähe in der Probe befanden.

**[0182]** In allen Ausführungsformen stellt der Gewinnungsschritt (e) Informationen über intrachromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereit.

**[0183]** In allen Ausführungsformen umfasst der Gewinnungsschritt (e) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem:

Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

**[0184]** In allen Ausführungsformen ist die Sequenzierungsreaktion eine Short-Read-Sequenzierungsreaktion mit hoher Genauigkeit.

**[0185]** In einigen Aspekten stellt die vorliegende Offenbarung Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext bereit, die die folgenden Schritte beinhalten: (a) Vernetzen von Nukleinsäuren innerhalb der Probe zum Bilden vernetzter Nukleinsäuren, wobei das Vernetzen kovalente Verknüpfungen zwischen räumlich benachbarten Nukleinsäuresegmenten bildet; (b) Ablegen der vernetzten Nukleinsäuren in diskrete Partitionen; (c) Verarbeiten der vernetzten Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu

erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten; (d) Gewinnen von Sequenzinformationen aus der Vielzahl von Ligationsprodukten, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

**[0186]** In weiteren Ausführungsformen beinhaltet der Verarbeitungsschritt (b) eine Blunt-End-Ligation unter Bedingungen, die eine intramolekulare Ligation begünstigen, sodass die räumlich benachbarten Nukleinsäuresegmente innerhalb desselben Moleküls ligiert werden.

**[0187]** In allen Ausführungsformen ist die Probe eine Formalin-fixierte Probe in Paraffin.

**[0188]** In allen Ausführungsformen umfasst die Probe eine Kernmatrix.

**[0189]** In allen Ausführungsformen umfassen die Nukleinsäuren RNA.

**[0190]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Vernetzungsschritt (a) nicht aus der Probe isoliert.

**[0191]** In allen Ausführungsformen beträgt die Menge an Nukleinsäuren in der Probe weniger als 5, 10, 15, 20, 25, 30, 35, 40, 45 oder 50 ng/ml.

**[0192]** In allen Ausführungsformen werden die Nukleinsäuren vor dem Vernetzungsschritt (a) immunpräzipitiert, sodass assoziierte DNA-bindende Proteine an den Nukleinsäuren gebunden bleiben.

**[0193]** In allen Ausführungsformen werden die Ligationsprodukte vor dem Gewinnungsschritt (d) einem Barcode zugeordnet.

**[0194]** In allen Ausführungsformen erhalten Ligationsprodukte innerhalb derselben Partition gemeinsame Barcodes, sodass die Barcodes Ligationsprodukte aus einer gegebenen Partition identifizieren.

**[0195]** In allen Ausführungsformen umfasst der Gewinnungsschritt (d) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

**[0196]** Während bevorzugte Ausführungsformen der vorliegenden Erfindung hier gezeigt und beschrieben wurden, ist es für den Fachmann offensichtlich, dass solche Ausführungsformen lediglich beispielhaft bereitgestellt werden. Dem Fachmann werden nun zahlreiche Abwandlungen, Änderungen und Substitutionen einfallen, ohne von der Erfindung abzuweichen. Es versteht sich, dass verschiedene Alternativen zu den hier beschriebenen Ausführungs-

formen der Erfindung bei der Ausführung der Erfindung eingesetzt werden können. Es ist beabsichtigt, dass die folgenden Ansprüche den Schutzbereich der Erfindung definieren und dass Verfahren und Strukturen innerhalb des Schutzbereichs dieser Ansprüche und ihrer Äquivalente dadurch abgedeckt sind.

## BEISPIELE

### Beispiel 1: Probenvorbereitung

**[0197]** Probenvorbereitungsverfahren wurden modifiziert, um lange DNA-Moleküle aus FFPE-Proben bereitzustellen. **Fig. 7** veranschaulicht einen beispielhaften Arbeitsablauf mit Modifikationen, die für die Vorbereitung von FFPE-Proben sowohl für die Sequenzierung des gesamten Genoms (WGS) als auch für die Sequenzierung des gesamten Exoms (WES) angegeben sind. Beispielsweise wurde nach der DNA-Extraktion ein Standard-Thermocycling-Protokoll bei 701 modifiziert, um den 98-Grad-Denaturierungsschritt vom Ende jedes Zyklus zum Anfang zu verschieben. Zusätzlich wurde am Ende jedes Zyklus ein 70-Grad-Halten für 2 Minuten hinzugefügt.

**[0198]** Während der Säuberung nach dem Zyklus 702 und den WES-Bibliotheksherstellungs- und ZielanreicherungsSchritten 704 und 705 wurden 1,8-fache SPRI-Beads (SPRI: Solid Phase reversible Immobilisation) mit normalen Protokollen verwendet.

**[0199]** Eine weitere Modifikation beinhaltete ein Ändern von Bedingungen während des Scherschritts 703, bei dem ein Ultraschallgerät mit einer Spitzenvorlaufleistung von etwa 450 verwendet wurde, im Gegensatz zu einem Standard-Ultraschallgerät mit einer Spitzenvorlaufleistung von 50.

**[0200]** Eine zusätzliche Modifikation, die in bestimmten Situationen verwendet werden kann, besteht darin, die FFPE-Probe zunächst mit Organokatalysatoren zu verarbeiten, um Formaldehyd-Addukte zu entfernen, wie zum Beispiel in Karmakar et al., (2015), Nature Chemistry, DOI: 10.1038/NCHEM.2307 beschrieben. Solche Protokolle beinhalten die Zugabe von 5 mM Organokatalysatoren in 30 mM pH-7-Tris-Puffer zu den Proben, um eine Adduktumkehr zu bewirken. Zu wirksamen Organokatalysatoren gehören ohne Einschränkung wasserlösliche bifunktionelle Katalysatoren, wie etwa die in Karmakar et al. beschriebenen Anthranilat- und Phosphanilat-katalysatoren. Die Umkehrung der Addukte hat die Wirkung, die Ausbeute an Nukleinsäureausbeuten aus der Probe zu verbessern.

### Beispiel 2: Barcodierung von FFPE-Proben

**[0201]** FFPE-Proben (wofür FFPE-Proben auf einem Objektträger gehören können) können mit

DNA-Barcodes markiert werden, die in einem räumlich wohldefinierten Muster aufgebracht werden, wie etwa denjenigen, die beim Drucken von DNA-Mikroarrays verwendet werden. Der DNA-Barcode (im Folgenden als Barcode-1 bezeichnet) ist entweder lang, damit er in nachfolgenden Schritten nicht herausdifundiert, oder er wird kovalent auf die FFPE-Probe aufgebracht. Damit Barcode-DNA in FFPE-Objektträger eingebettet werden kann, wird die Probe erhitzt und dann werden die Barcodes hinzugefügt. Die Barcodes sind im Allgemeinen eine Bibliothek von Barcodes, sodass verschiedene Barcodes in verschiedenen Teilen des Objektträgers bereitgestellt werden. Die Barcodes können auch in unterschiedlichen Konzentrationen an verschiedenen Stellen des Objektträgers hinzugefügt werden, um die geografische Codierung zu unterstützen - in dieser Situation kann die Barcode-Bibliothek identische oder unterschiedliche Barcodes umfassen. Nachdem die Barcodes hinzugefügt wurden, wird der Objektträger gekühlt und dann im Allgemeinen durch Schneiden in Teile aufgetrennt, beispielsweise unter Verwendung von Laser-Mikrodissektion, mechanischen/akustischen Mitteln und dergleichen. Anstelle von Barcodes können auch Fluorophore oder Qdots verwendet werden, jedoch ermöglicht Barcodierung eine massiv parallele, zufällige Verkapselung von Probenteilen unter Beibehaltung lokaler räumlicher Informationen (z. B. Tumor- vs. normale Zellen).

**[0202]** Die Teile der Probe, die die Barcodes enthalten, können dann in ein Sequenzierungssystem gegeben werden, einschließlich eines Tröpfchen-basierten Systems wie etwa das 10X Genomics Chromium™-System, sodass ein einzelner barcodierter Teil pro Tröpfchen verkapselt wird.

**[0203]** Die Entparaffinierung der Probe kann im Tröpfchen durch Erhitzen erfolgen. Paraffin ist mit Wasser nicht mischbar, jedoch in bestimmten Ölen löslich, und somit kann das Paraffin beim Erhitzen der Tröpfchen auf dem Chip leicht aus dem Tröpfchen entfernt werden. Xylol könnte auch in einem Flüssig-Flüssig-Extraktionsprozess verwendet werden, um die Probenteile zu entparaffinieren und ihre Nukleinsäureinhalte für die weitere Verarbeitung vorzubereiten.

**[0204]** Zu weiteren Schritten gehören das Entnetzen von Methylenbrücken der entparaffinierten Probe. Für diesen Schritt können spezielle chemische Mittel verwendet werden, um die Quervernetzungen zu entfernen und dadurch einen Zugriff auf die enthaltenen Nukleinsäuren für jede nachfolgende Verarbeitung zu ermöglichen, einschließlich der hier erörterten Schritte der Nukleinsäure-Barcodierung, Amplifikation und Bibliothekherstellung (siehe zum Beispiel **Fig. 2**). Es sei darauf hingewiesen, dass die Spatial-Barcoding-DNA auch in dem Tröpfchen

eingekapselt ist. Der zweite Barcodierungsschritt der einzelnen Nukleinsäuren dient der Barcodierung der Nukleinsäuren und als Barcode zur räumlichen Codierung der Probe. Sequenz-Reads können dann zusammengefügt werden, um Informationen bereitzustellen, die dann mit dem ursprünglichen räumlichen Ort in der Probe verglichen und somit mit pathologischen Daten in Beziehung gesetzt werden können.

**[0205]** In alternativen Versionen dieses Arbeitsablaufs zur räumlichen Codierung wird der Schritt des Entnetzens zuerst innerhalb des Tröpfchens durchgeführt und dann werden die Nukleinsäuren in der Probe, einschließlich genomischer DNA sowie die Barcodes zur räumlichen Codierung, an Partikel angebracht oder auf andere Weise aus der Probe isoliert. Die Nukleinsäuren werden dann wieder verkapselt und dem Arbeitsablauf der Barcodierung und Sequenzierung in hier beschriebenen Verfahren unterzogen, einschließlich der in **Fig. 2** abgebildeten.

**[0206]** Die vorliegende Beschreibung stellt eine vollständige Beschreibung der Methodologien, Systeme und/oder Strukturen und Verwendungen davon in beispielhaften Aspekten der vorliegend beschriebenen Technologie bereit. Obwohl verschiedene Aspekte dieser Technologie oben mit einem gewissen Maß an Genauigkeit oder unter Bezugnahme auf einen oder mehrere einzelne Aspekte beschrieben wurden, könnte der Fachmann zahlreiche Änderungen an den offenbarten Aspekten vornehmen, ohne vom Geist oder Umfang der Technologie abzuweichen. Da viele Aspekte ausgeführt werden können, ohne von Geist und Umfang der vorliegend beschriebenen Technologie abzuweichen, liegt der angemessene Umfang in den nachstehend beigefügten Ansprüchen. Es werden daher auch andere Aspekte in Betracht gezogen. Darüber hinaus versteht es sich, dass alle Operationen in beliebiger Reihenfolge durchgeführt werden können, es sei denn, es wird ausdrücklich etwas anderes beansprucht oder eine bestimmte Reihenfolge ist von Natur aus durch den Wortlaut eines Anspruchs erforderlich. Es ist beabsichtigt, dass alle Inhalte, die in der obigen Beschreibung enthalten sind und in den begleitenden Zeichnungen gezeigt sind, nur als veranschaulichend für bestimmte Aspekte ausgelegt werden sollen und nicht auf die gezeigten Ausführungsformen beschränkend sind. Sofern aus dem Kontext nicht klar hervorgeht oder ausdrücklich angegeben wird, sind alle hier bereitgestellten Konzentrationswerte im Allgemeinen in Form von Beimischungswerten oder Prozentsätzen ohne Berücksichtigung einer Umwandlung angegeben, die bei oder nach der Zugabe der jeweiligen Komponente der Mischung auftritt. Alle veröffentlichten Literaturverweise und Patentedokumente, auf die in dieser Offenbarung Bezug genommen wird, werden hiermit in ihrer Gesamtheit für alle Zwecke durch Bezugnahme auf-

genommen, soweit sie nicht bereits ausdrücklich hier aufgenommen sind. Änderungen im Detail oder der Struktur können vorgenommen werden, ohne von den Grundelementen der vorliegenden Technologie abzuweichen, wie sie in den folgenden Ansprüchen definiert ist.

**[0207]** Die folgenden nummerierten Absätze, die Aspekte der Erfindung beschreiben, sind Teil der Beschreibung.

1. Ein Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext, wobei das Verfahren Folgendes umfasst:

(a) Bereitstellen einer Nukleinsäuren umfassenden Probe, wobei die Nukleinsäuren dreidimensionale Strukturen umfassen;

(b) Auftrennen von Teilen der Probe in diskrete Partitionen, sodass Teile der dreidimensionalen Nukleinsäurestrukturen ebenfalls in die diskreten Partitionen aufgetrennt werden;

(c) Gewinnen von Sequenzinformationen von den Nukleinsäuren, wodurch Nukleinsäuren unter Beibehaltung von strukturellem Kontext analysiert werden.

2. Das Verfahren nach Absatz 1, wobei die Sequenzinformationen aus dem Gewinnungsschritt (c) eine Identifizierung von Nukleinsäuren beinhalten, die sich in der Probe in räumlicher Nähe zueinander befanden.

3. Das Verfahren nach Absätzen 1-2, wobei der Gewinnungsschritt (c) Informationen über intra-chromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereitstellt.

4. Das Verfahren nach Absätzen 1-2, wobei der Gewinnungsschritt (c) Informationen über Chromosomenkonformationen bereitstellt.

5. Das Verfahren nach Absätzen 1-4, wobei vor dem Auftrennungsschritt (b) mindestens einige der dreidimensionalen Strukturen verarbeitet werden, um verschiedene Teile der Nukleinsäuren zu verknüpfen, die innerhalb der dreidimensionalen Strukturen nahe beieinander liegen.

6. Das Verfahren nach Absätzen 1-5, wobei die Probe eine Formalin-fixierte Probe in Paraffin ist.

7. Das Verfahren nach Absätzen 1-6, wobei die Nukleinsäuren vor dem Auftrennungsschritt (b) nicht aus der Probe isoliert werden.

8. Das Verfahren nach Absätzen 1-7, wobei die diskreten Partitionen Beads umfassen.

9. Das Verfahren nach Absatz 8, wobei die Beads Gel-Beads sind.

10. Das Verfahren nach Absätzen 1-9, wobei die Nukleinsäuren vor dem Auftrennungsschritt (c) nicht aus der Probe isoliert werden.

11. Das Verfahren nach Absätzen 1-10, wobei der Gewinnungsschritt (c) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem umfasst: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

12. Das Verfahren nach Absätzen 1-11, wobei die Probe eine Tumorseite umfasst.

13. Das Verfahren nach Absätzen 1-11, wobei die Probe eine Mischung aus Tumor- und normalen Zellen umfasst.

14. Das Verfahren nach Absätzen 1-13, wobei die Probe eine Kernmatrix umfasst.

15. Das Verfahren nach Absätzen 1-14, wobei die Nukleinsäuren RNA umfassen.

16. Das Verfahren zum Analysieren von Nukleinsäuren aus einer Probe unter Beibehaltung von strukturellem Kontext von Nukleinsäuren innerhalb der Probe, wobei das Verfahren Folgendes umfasst:

(a) Bilden verknüpfter Nukleinsäuren innerhalb der Probe, sodass räumlich benachbarte Nukleinsäuresegmente verknüpft werden;

(b) Verarbeiten der verknüpften Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten;

(c) Ablegen der Vielzahl von Ligationsprodukten in diskrete Partitionen;

(d) Barcodieren der Ligationsprodukte innerhalb der diskreten Partitionen zum Bilden einer Vielzahl barcodierter Fragmente, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen, wodurch jedes Fragment der verknüpften Nukleinsäure zugeordnet wird, von der es abgeleitet ist;

(e) Gewinnen von Sequenzinformationen aus der Vielzahl barcodierter Fragmente, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

17. Das Verfahren nach Absatz 16, wobei der Verarbeitungsschritt (b) eine Blunt-End-Ligation unter Bedingungen, die eine intramolekulare Ligation begünstigen, beinhaltet, sodass die räumlich benachbarten Nukleinsäuresegmente innerhalb desselben Moleküls ligiert werden.

18. Das Verfahren nach Absatz 17, wobei die Bedingungen, die eine intramolekulare Ligation begünstigen, ein Verdünnen der Probe umfassen, um die Konzentration der Nukleinsäuren auf unter 10 ng/µl zu verringern.

19. Das Verfahren nach Absätzen 16-18, wobei die Probe eine Formalin-fixierte Probe in Paraffin ist.

20. Das Verfahren nach Absätzen 16-18, wobei die Probe eine Kernmatrix umfasst.

21. Das Verfahren nach Absätzen 16-20, wobei die Nukleinsäuren RNA umfassen.

22. Das Verfahren nach Absätzen 16-21, wobei die Nukleinsäuren vor dem Schritt (a) nicht aus der Probe isoliert werden.

23. Verfahren nach den Ansprüchen 16-21, wobei vor Schritt (a) die Nukleinsäuren immunpräzipitiert werden, sodass assoziierte DNA-bindende Proteine an den Nukleinsäuren gebunden bleiben.

24. Das Verfahren nach Absätzen 16-23, wobei die Menge an Nukleinsäuren in der Probe weniger als 50 ng/ml beträgt.

25. Das Verfahren nach Absätzen 16-24, wobei die Partitionen Beads umfassen.

26. Das Verfahren nach Absatz 25, wobei die Beads Gel-Beads sind.

27. Das Verfahren nach Absätzen 16-26, wobei die Probe eine Tumorseite umfasst.

28. Das Verfahren nach Absätzen 16-27, wobei die Probe eine Mischung aus Tumor- und normalen Zellen umfasst.

29. Das Verfahren nach Absätzen 16-28, wobei der Verarbeitungsschritt eine Umkehrung der Verknüpfung nach dem Bilden der Ligationsprodukte beinhaltet.

30. Das Verfahren nach Absätzen 16-29, wobei der Gewinnungsschritt (e) Informationen über intrachromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereitstellt.

31. Das Verfahren nach Absätzen 16-29, wobei der Gewinnungsschritt (e) Informationen über Chromosomenkonformationen bereitstellt.

32. Das Verfahren nach Absatz 31, wobei die Chromosomenkonformationen Krankheitszuständen zugeordnet sind.

33. Das Verfahren nach Absätzen 16-32, wobei der Verarbeitungsschritt zu Ligationsprodukten führt, die Nukleinsäuren umfassen, die sich ursprünglich in enger räumlicher Nähe in der Probe befanden.

34. Das Verfahren nach Absätzen 16-33, wobei der Gewinnungsschritt (e) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem umfasst: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

35. Das Verfahren nach Absatz 34, wobei die Sequenzierungsreaktion eine Short-Read-Sequenzierungsreaktion mit hoher Genauigkeit ist.

36. Das Verfahren nach Absätzen 16-35, wobei der Bildungsschritt (a) ein Vernetzen von Nukleinsäuren in der Probe umfasst.

37. Das Verfahren nach Absätzen 16-36, wobei der Bildungsschritt (a) zu kovalenten Verknüpfungen zwischen räumlich benachbarten Nukleinsäuresegmenten führt.

38. Ein Verfahren zum Analysieren von Nukleinsäuren aus einer Probe unter Beibehaltung von strukturellem Kontext von Nukleinsäuren innerhalb der Probe, wobei das Verfahren Folgendes umfasst:

(a) Bilden verknüpfter Nukleinsäuren innerhalb der Probe, sodass räumlich benachbarte Nukleinsäuresegmente verknüpft werden;

(b) Ablegen der verknüpften Nukleinsäuren in diskrete Partitionen;

(c) Verarbeiten der verknüpften Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten;

(d) Barcodieren der Ligationsprodukte innerhalb der diskreten Partitionen zum Bilden einer Vielzahl barcodierter Fragmente, wobei Fragmente innerhalb einer gegebenen diskreten Partition jeweils einen gemeinsamen Barcode umfassen, wodurch jedes Fragment der verknüpften Nukleinsäure zugeordnet wird, von der es abgeleitet ist;

(e) Gewinnen von Sequenzinformationen aus der Vielzahl barcodierter Fragmente, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

39. Das Verfahren nach Absatz 38, wobei der Verarbeitungsschritt (c) eine Blunt-End-Ligation unter Bedingungen, die eine intramolekulare Ligation begünstigen, beinhaltet, sodass die räumlich benachbarten Nukleinsäuresegmente innerhalb desselben Moleküls ligiert werden.

40. Das Verfahren nach Absätzen 38-39, wobei die Probe eine Formalin-fixierte Probe in Paraffin ist.

41. Das Verfahren nach Absätzen 38-39, wobei die Probe eine Kernmatrix umfasst.

42. Das Verfahren nach Absätzen 38-41, wobei die Nukleinsäuren RNA umfassen.

43. Das Verfahren nach Absätzen 38-42, wobei die Nukleinsäuren vor dem Schritt (a) nicht aus der Probe isoliert werden.

44. Das Verfahren nach Absätzen 38-42, wobei vor Schritt (a) die Nukleinsäuren immunpräzipitiert werden, sodass assoziierte DNA-bindende Proteine an den Nukleinsäuren gebunden bleiben.

45. Das Verfahren nach Absätzen 38-44, wobei die Partitionen Beads umfassen.

46. Das Verfahren nach Absatz 45, wobei die Beads Gel-Beads sind.

47. Das Verfahren nach Absätzen 38-46, wobei die Probe eine Tumorseite umfasst.

48. Das Verfahren nach Absätzen 38-46, wobei die Probe eine Mischung aus Tumor- und normalen Zellen umfasst.

49. Das Verfahren nach Absätzen 38-48, wobei der Verarbeitungsschritt (c) zu Ligationsprodukten führt, die Nukleinsäuren umfassen, die sich ursprünglich in enger räumlicher Nähe in der Probe befanden.

50. Das Verfahren nach Absätzen 38-49, wobei der Gewinnungsschritt (e) Informationen über intrachromosomale und/oder interchromosomale Wechselwirkungen zwischen genomischen Loci bereitstellt.

51. Das Verfahren nach Absätzen 38-50, wobei der Gewinnungsschritt (e) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem umfasst: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

52. Das Verfahren nach Absatz 51, wobei die Sequenzierungsreaktion eine Short-Read-Sequenzierungsreaktion mit hoher Genauigkeit ist.

53. Ein Verfahren zum Analysieren von Nukleinsäuren aus einer Probe unter Beibehaltung von strukturellem Kontext von Nukleinsäuren innerhalb der Probe, wobei das Verfahren Folgendes umfasst:

(a) Vernetzen von Nukleinsäuren innerhalb der Probe zum Bilden vernetzter Nukleinsäuren, wobei das Vernetzen kovalente Verknüpfungen zwischen räumlich benachbarten Nukleinsäuresegmenten bildet;

(b) Ablegen der vernetzten Nukleinsäuren in diskrete Partitionen;

(c) Verarbeiten der vernetzten Nukleinsäuren, um eine Vielzahl von Ligationsprodukten zu erzeugen, wobei die Ligationsprodukte Teile der räumlich benachbarten Nukleinsäuresegmente enthalten;

(d) Gewinnen von Sequenzinformationen aus der Vielzahl von Ligationsprodukten, wodurch Nukleinsäuren aus der Probe unter Beibehaltung von strukturellem Kontext analysiert werden.

54. Das Verfahren nach Absatz 53, wobei der Verarbeitungsschritt (b) eine Blunt-End-Ligation unter Bedingungen, die eine intramolekulare Ligation begünstigen, beinhaltet, sodass die räumlich benachbarten Nukleinsäuresegmente innerhalb desselben Moleküls ligiert werden.

55. Das Verfahren nach Absätzen 53-54, wobei die Probe eine Formalin-fixierte Probe in Paraffin ist.

56. Das Verfahren nach Absätzen 53-54, wobei die Probe eine Kernmatrix umfasst.

57. Das Verfahren nach Absätzen 53-56, wobei die Nukleinsäuren RNA umfassen.

58. Das Verfahren nach Absätzen 53-57, wobei die Nukleinsäuren vor dem Vernetzungsschritt (a) nicht aus der Probe isoliert werden.

59. Das Verfahren nach Absätzen 53-57, wobei vor dem Vernetzungsschritt (a) die Nukleinsäuren immunpräzipitiert werden, sodass assoziierte DNA-bindende Proteine an den Nukleinsäuren gebunden bleiben.

60. Das Verfahren nach Absätzen 53-59, wobei vor dem Gewinnungsschritt (d) die Ligationsprodukte einem Barcode zugeordnet werden.

61. Das Verfahren nach Absatz 60, wobei Ligationsprodukte innerhalb derselben Partition gemeinsame Barcodes erhalten, sodass die Barcodes Ligationsprodukte aus einer gegebenen Partition identifizieren.

62. Das Verfahren nach Absätzen 53-61, wobei der Gewinnungsschritt (d) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem umfasst: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

63. Verfahren zum Analysieren von Nukleinsäuren unter Beibehaltung von strukturellem Kontext, wobei das Verfahren Folgendes umfasst:

(a) Bereitstellen einer Probe, die Nukleinsäuren umfasst;

(b) Aufbringen einer Bibliothek von Tags auf die Probe, sodass unterschiedliche geografische

Regionen der Probe unterschiedliche Tags erhalten;

(c) Auftrennen von Teilen der Probe in diskrete Partitionen, sodass Teile der Bibliothek von Tags und Teile der Nukleinsäuren ebenfalls in die diskreten Partitionen aufgetrennt werden;

(d) Gewinnen von Sequenzinformationen von den Nukleinsäuren und

(e) Identifizieren von Tags in den diskreten Partitionen,

wodurch Nukleinsäuren unter Beibehaltung von strukturellem Kontext analysiert werden.

64. Das Verfahren nach Absatz 63, wobei die Bibliothek von Tags Oligonukleotid-Tags umfasst.

65. Das Verfahren nach Absätzen 63-64, wobei die Bibliothek von Tags mRNA-Tags umfasst.

66. Das Verfahren nach Absätzen 63-65, wobei der Aufbringungsschritt (b) durch Diffusion erreicht wird.

67. Das Verfahren nach Absatz 66, wobei der Aufbringungsschritt (b) zu unterschiedlichen Konzentrationen von Tags an unterschiedlichen räumlichen Orten in der Probe führt.

68. Das Verfahren nach Absatz 67, wobei der Identifizierungsschritt (e) ein Bestimmen der Konzentration von Tags in den diskreten Partitionen umfasst.

69. Das Verfahren nach Absätzen 63-65, wobei der Aufbringungsschritt (b) durch einen oder mehrere aktive Prozesse innerhalb der Probe erreicht wird.

70. Das Verfahren nach Absatz 69, wobei ein Vorhandensein identischer Tags an unterschiedlichen räumlichen Orten in der Probe angibt, dass Elemente der Probe an diesen unterschiedlichen räumlichen Orten durch einen oder mehrere aktive Prozesse verbunden sind.

71. Das Verfahren nach Absätzen 69-70, wobei der eine oder die mehreren aktiven Prozesse zelluläre Transportprozesse beinhalten.

72. Das Verfahren nach Absätzen 63-71, wobei der Aufbringungsschritt (b) ein zufälliges Verteilen unterschiedlicher Teile der Bibliothek von Tags auf unterschiedliche räumliche Orte der Probe umfasst.

73. Das Verfahren nach Absätzen 63-71, wobei der Aufbringungsschritt (b) ein Zugeben eines vorbestimmten Teils oder einer vorbestimmten Konzentration der Bibliothek von Tags zu identifizierten räumlichen Orten der Probe umfasst.

74. Das Verfahren nach Absätzen 63-73, wobei der Gewinnungsschritt (d) und der Identifizierungsschritt (e) gleichzeitig durchgeführt werden.

75. Das Verfahren nach Absätzen 63-74, wobei der Gewinnungsschritt (d) eine Sequenzierungsreaktion ausgewählt aus der Gruppe bestehend aus Folgendem umfasst: Short-Read-Sequenzierungsreaktionen und Long-Read-Sequenzierungsreaktionen.

76. Das Verfahren nach Absatz 75, wobei die Sequenzierungsreaktion eine Short-Read-Sequenzierungsreaktion mit hoher Genauigkeit ist.

77. Das Verfahren nach Absätzen 63-76, wobei die Probe eine Tumorseite umfasst.

78. Das Verfahren nach Absätzen 63-76, wobei die Probe eine Mischung aus Tumor- und normalen Zellen umfasst.

79. Das Verfahren nach Absätzen 63-76, wobei die Probe eine Formalin-fixierte Probe in Paraffin ist.

80. Das Verfahren nach Absätzen 63-76, wobei die Probe eine Kernmatrix umfasst.

81. Das Verfahren nach Absätzen 63-76, wobei die Nukleinsäuren RNA umfassen.

82. Das Verfahren nach Absätzen 63-82, wobei die Partitionen Beads umfassen.

83. Das Verfahren nach Absatz 82, wobei die Beads Gel-Beads sind.

84. Das Verfahren nach Absätzen 63-83, wobei die Menge an Nukleinsäuren in den diskreten Partitionen weniger als 10 ng/ml beträgt.

85. Das Verfahren nach Absätzen 63-83, wobei die Menge an Nukleinsäuren in den diskreten Partitionen weniger als 1 ng/ml beträgt.

**ZITATE ENTHALTEN IN DER BESCHREIBUNG**

**Zitierte Patentliteratur**

- US 62/263532 [0001]
- US 14/927297 [0045]
- US 62/119996 [0045]
- US 62/146834 [0047]
- US 20130157870 [0052, 0055]
- US 20130079231 [0055]
- US 13/966150 [0062]
- US 61/977804 [0062]
- US 62/017580 [0063]
- US 14/752773 [0084]
- US 62/072164 [0110]
- US 14/752602 [0130]
- US 14/752641 [0133]



**Schutzansprüche**

1. System, das dazu konfiguriert ist, Nukleinsäuren, die aus einer Formalin-fixierten in Paraffin eingebetteten (FFPE) Gewebeprobe gewonnen wurden, unter Beibehaltung von räumlichem Kontext zu analysieren, wobei die Analyse Folgendes umfasst:

a) Partitionieren der aus der FFPE-Gewebeprobe erhaltenen Nukleinsäuren in eine Vielzahl von Vertiefungen; wobei Nukleinsäuren in räumlicher Nähe zueinander in der FFPE-Gewebeprobe in dieselbe Vertiefung eingebracht werden;

b) Barcodieren partitionierter Nukleinsäuren, um eine Vielzahl barcodierter Nukleinsäuren zu bilden, wobei barcodierte Nukleinsäuren innerhalb einer gegebenen diskreten Vertiefung jeweils eine gemeinsame partitionsspezifische Barcodesequenz umfassen, sodass die Barcodesequenzen Nukleinsäuren aus einer gegebenen Vertiefung identifizieren;

c) Gewinnen von Sequenzinformationen von der Vielzahl barcodierter Nukleinsäuren, wobei die Sequenzinformationen von der Vielzahl barcodierter Nukleinsäuren Sequenzinformationen der partitionsspezifischen Barcodesequenz umfassen; und

d) Zuschreiben der Vielzahl barcodierter Nukleinsäuren zu einer Region räumlicher Nähe, wobei barcodierte Nukleinsäuren, die von einer Region räumlicher Nähe in der FFPE-Gewebeprobe abgeleitet sind, dieselbe partitionsspezifische Barcodesequenz umfassen.

2. System nach Anspruch 1, wobei das Barcodieren ein Amplifizieren mit einem Primer umfasst, der eine Barcodesequenz umfasst.

3. System nach Anspruch 1, wobei mindestens zwei der Nukleinsäuren, die in dem Partitionierungsschritt a) in dieselbe Vertiefung partitioniert werden, unterschiedliche Sequenzen umfassen.

4. System nach Anspruch 1, wobei die Analyse ferner vor (a) einen vorherigen Schritt des Abbildens der FFPE-Gewebeprobe umfasst.

5. System nach Anspruch 1, wobei die FFPE-Gewebeprobe eine Krebsgewebeprobe ist.

6. System nach Anspruch 1, wobei die barcodierte Nukleinsäuren in verschiedenen Vertiefungen gepoolt werden, bevor Sequenzinformationen gewonnen werden.

7. System nach Anspruch 1, wobei die Sequenzinformationen ferner Informationen umfassen, die sich auf eine aus der FFPE-Gewebeprobe gewonnene Nukleinsäure beziehen.

8. System nach Anspruch 1, wobei die Analyse ferner vor a) einen vorherigen Schritt des Freisetzens der Nukleinsäuren aus der FFPE-Gewebeprobe umfasst.

9. System nach Anspruch 1, wobei die aus der FFPE-Gewebeprobe gewonnenen Nukleinsäuren zuvor auf die Probe aufgebrachte Nukleinsäure-Tags umfassen.

10. System nach Anspruch 1, wobei der Schritt des Gewinnens von Sequenzinformationen eine Hochdurchsatz-Sequenzierung der Vielzahl barcodierter Nukleinsäuren umfasst.

11. System nach Anspruch 1, wobei die partitionsspezifischen Barcodesequenzen zwei oder mehr separate Teilsequenzen umfassen.

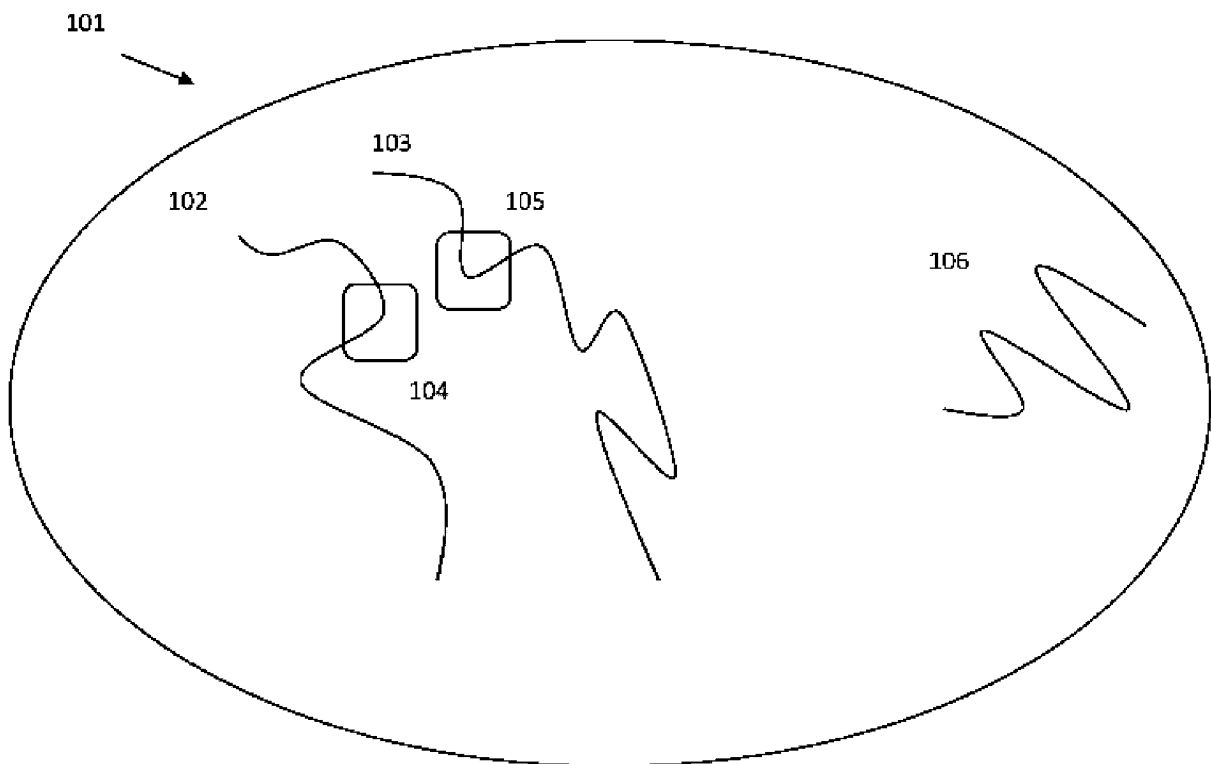
12. System nach Anspruch 4, wobei das Abbilden Abbildungs-Tags mit optischen Eigenschaften umfasst.

13. System nach Anspruch 12, wobei die Analyse ferner ein Korrelieren barcodierter Nukleinsäuren, die aus einer Region räumlicher Nähe in der FFPE-Gewebeprobe abgeleitet sind, mit den abgebildeten Tags umfasst.

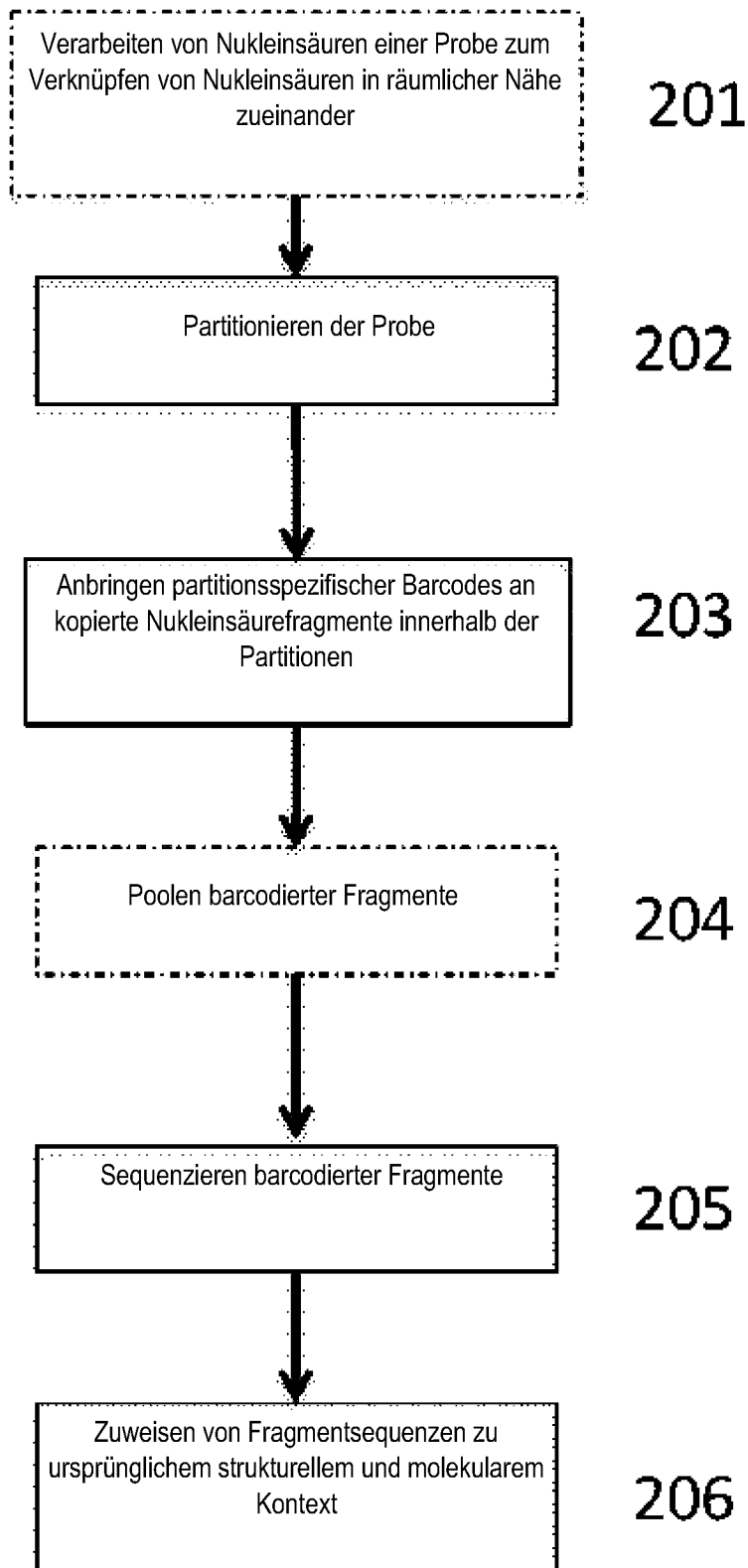
Es folgen 7 Seiten Zeichnungen

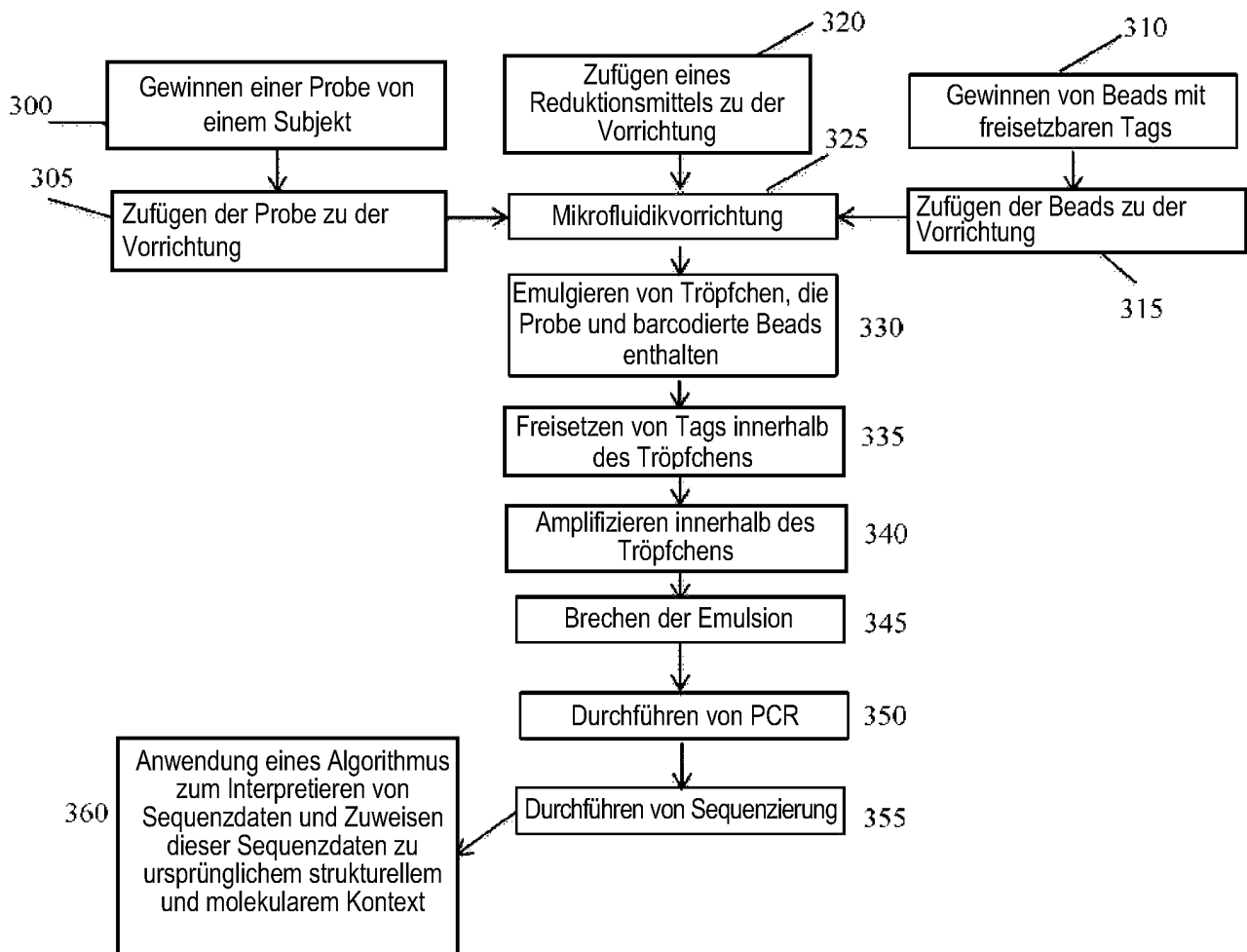
Anhängende Zeichnungen

Figur 1

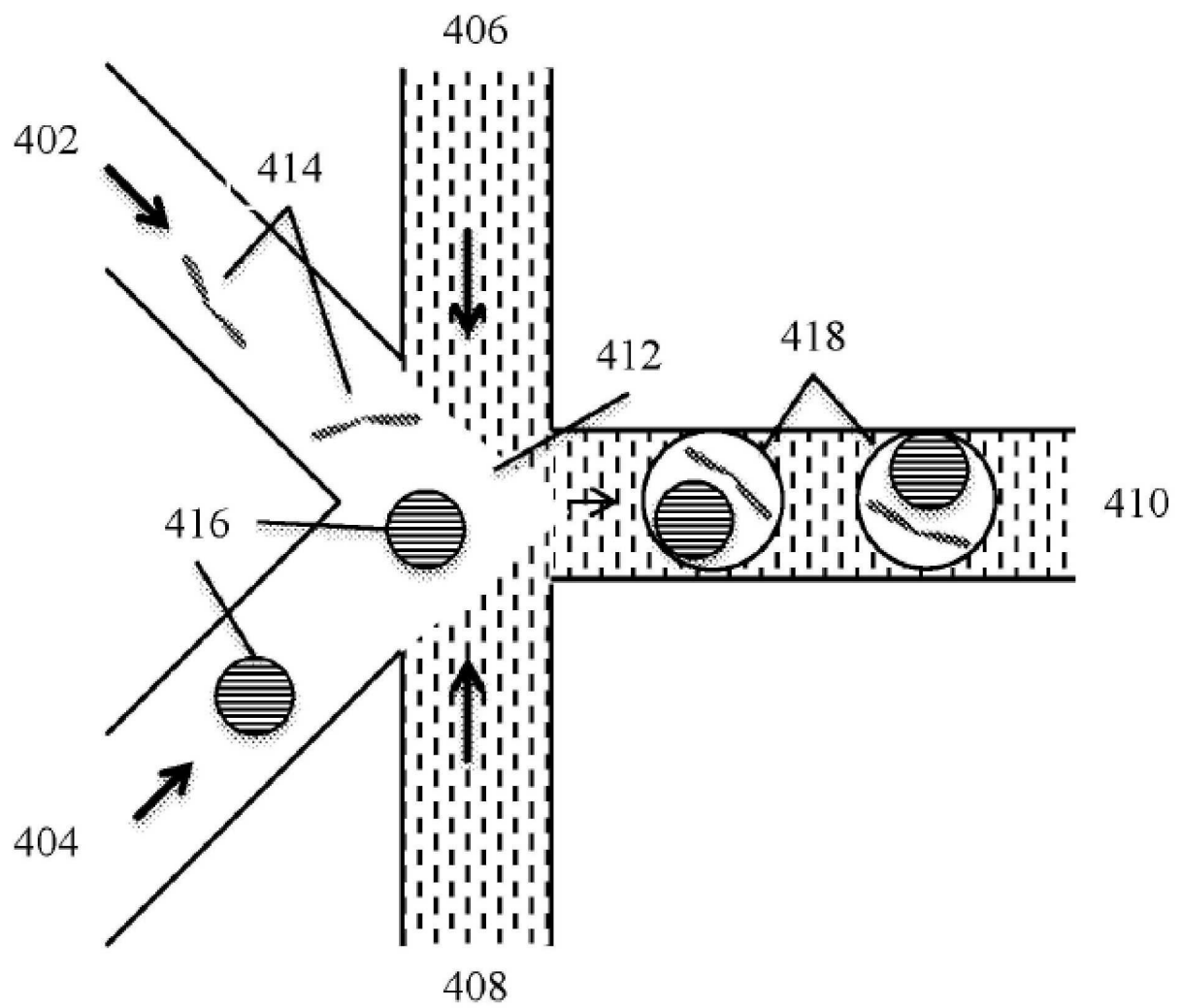


Figur 2

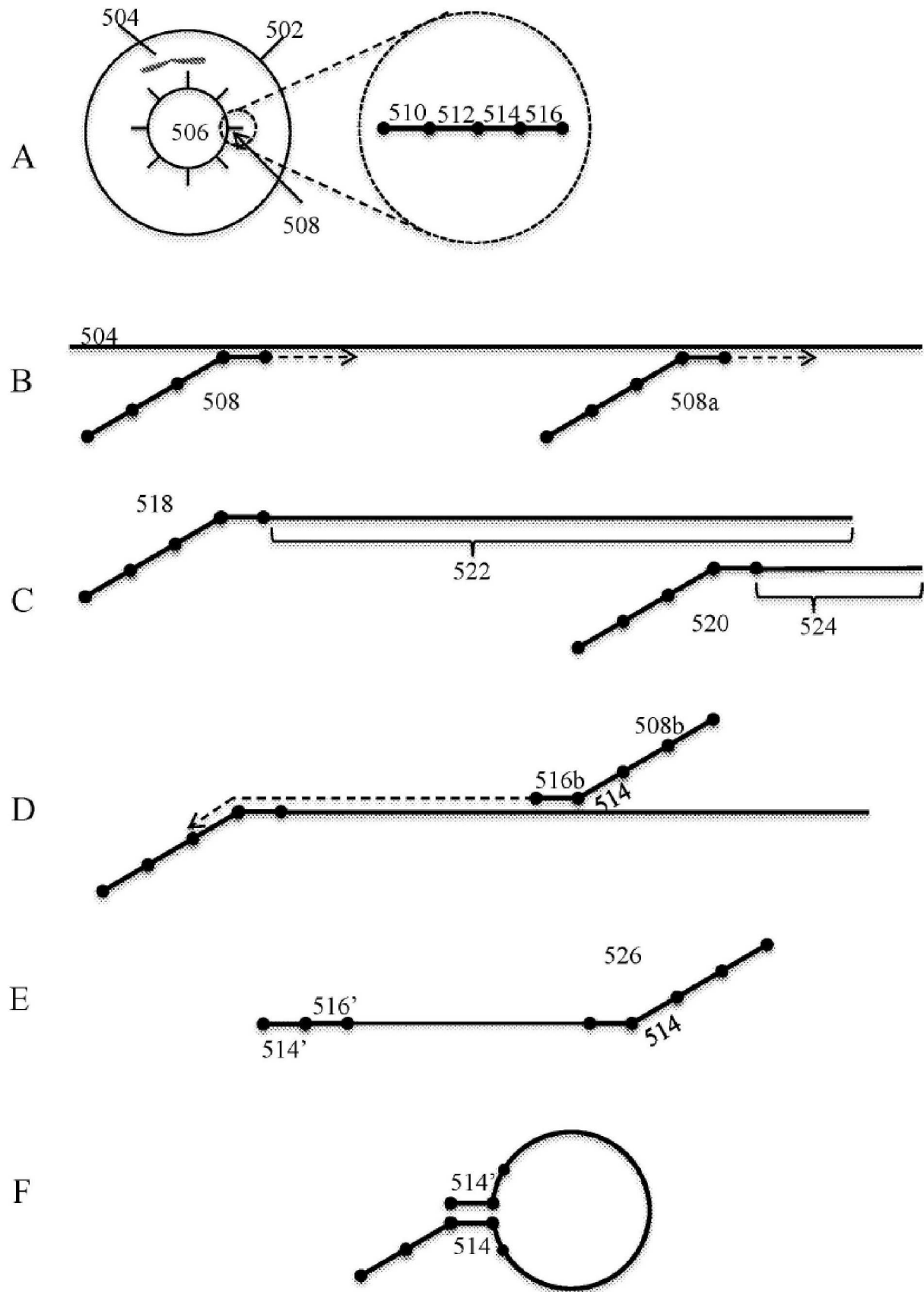




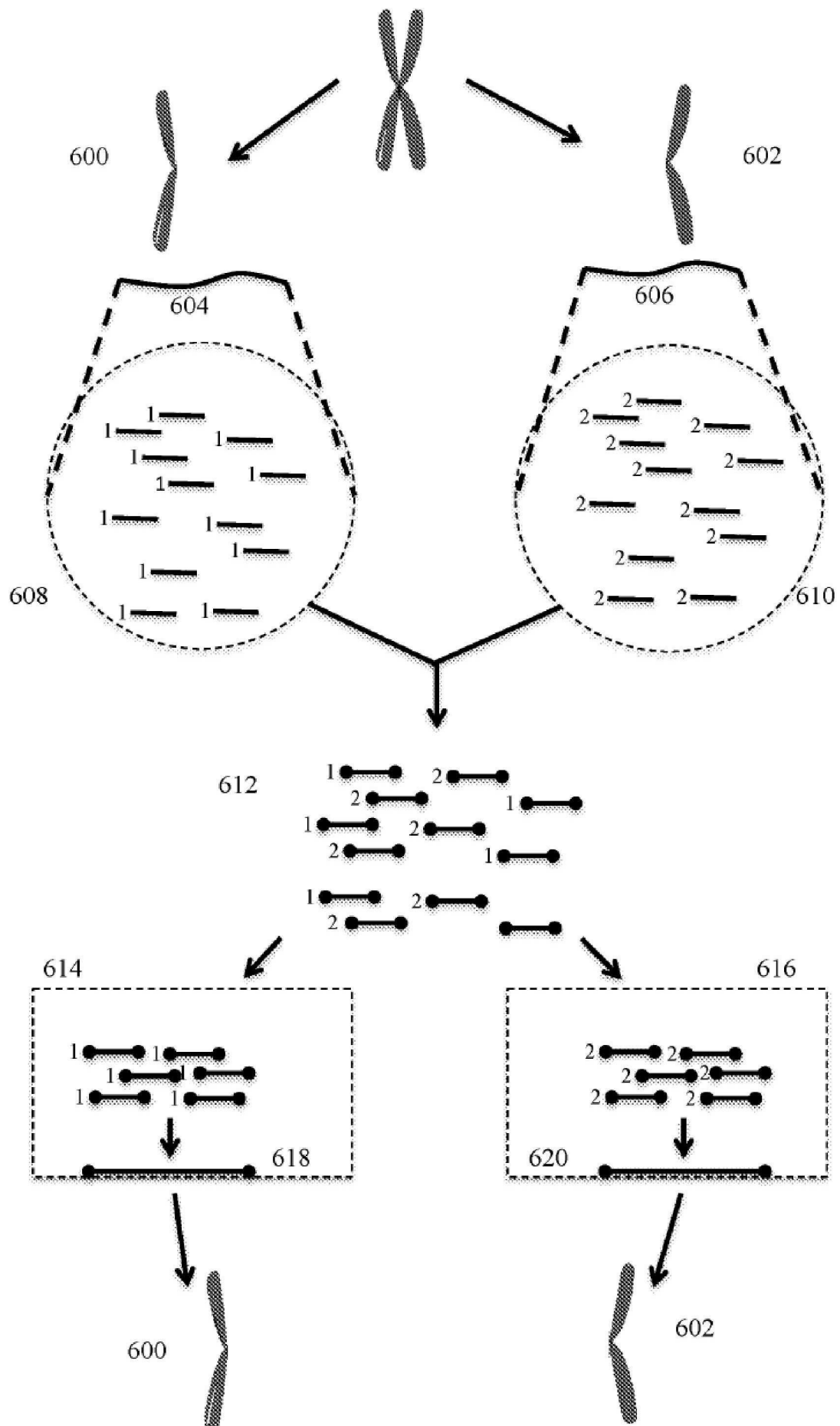
Figur 3



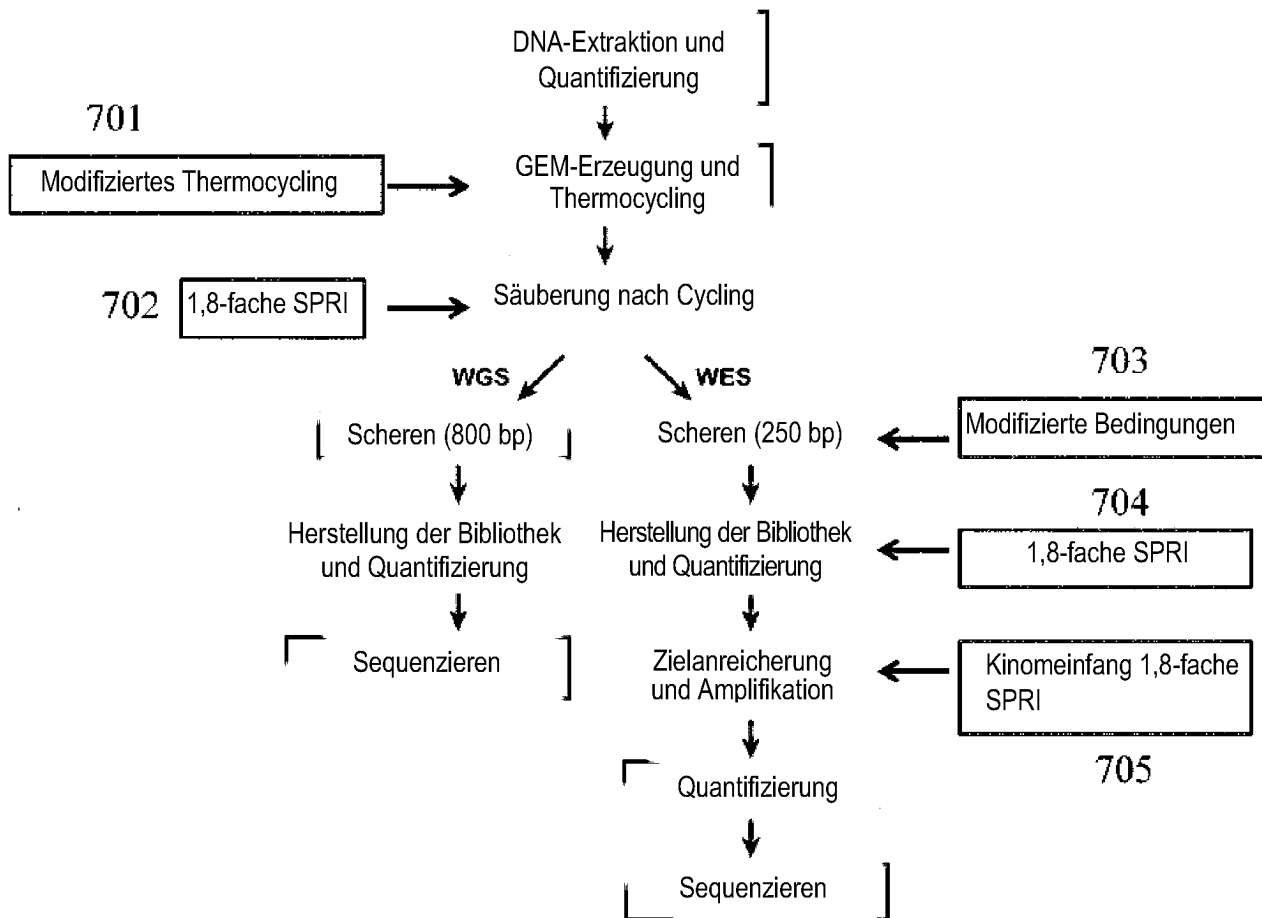
Figur 4



**Figur 5**



Figur 6

**Figur 7**