

Nucleic Acid Barcodes

Field of Invention

The present invention relates to dual modality nucleic acid barcodes that include a
5 programmed identifiable signature and sequence signature allowing for real time identification
of the barcodes to be linked with downstream analysis by linking the identifiable signature with
the sequence signature. Also provided are methods of making said barcodes, libraries of said
barcodes and libraries of supports attached to said barcodes. Further provided are uses and
methods of using said barcodes.

10

Background

Droplet microfluidics has become a cornerstone in biological research due to the low
volumes used. This leads to several advantages such as reduced reagent cost, increased
analytical throughput, automation, and single cell encapsulation. Droplet microfluidics
15 provides a miniaturized "test tube" in which ideally all solutes remain encapsulated in each
individual droplet, analogous to compartmentalization of a solution in plastic or glass tubes.
In this sense each droplet can have its own discrete solution and reaction, and with the high
throughput capabilities, allow for tens of millions of solutions to be analyzed in a single day.
As a result, droplet microfluidics has found applications in various fields, from single-cell
20 analysis and drug discovery to synthetic biology, materials science and diagnostics.

However, one of the foundational challenges in droplet microfluidics is the tracking and
identification of individual droplets, especially when dealing with large-scale experiments as
droplets have similar physical appearances, and do not always travel in single-file.

Furthermore, deciphering the contents of a droplet in real time can be complex and difficult
25 since small differences between droplets cannot be observed. The ability to distinguish
between droplets becomes paramount when each droplet acts as an individual reactor,
potentially containing unique biological or chemical entities. Traditional methods of droplet
identification, such as using fluorescent markers or tracking droplet generation sequences,
have their limitations since the methods allow for tracking of the droplet in the device but do
30 not contain any information about the contents of the droplets.

Moreover, as the field of droplet microfluidics expands and its applications become more
intricate, there is a growing need for a more refined method of droplet identification. For
instance, in single-cell genomic studies, where each droplet might contain genetic material
from a unique cell, the ability to accurately track and identify droplets is crucial, for example
35 in linking assay information to individual cells. Inability to track droplets leads to a plethora of
information being discarded, rendering the results of experiments being understated.

US10829815B2 discloses that plurality of partitions (e.g., droplets) may be generated such that partitions of the plurality of partitions each include a biological particle (e.g., cell) comprising the nucleic acid molecule and a particle (e.g., bead). The partitions can be processed (e.g., imaged) to obtain one or more physical and/or optical properties of their
5 respective biological particles. The nucleic acid molecules included in the partitions can be barcoded and sequenced (e.g., using nucleic acid barcode molecules coupled to the particles of the partitions) to generate nucleic acid sequences of the nucleic acid molecules. The nucleic acid sequences can be electronically associated with the one or more optical properties of the biological particles. However, the barcodes disclosed do not provide a
10 specific and programmable identification system that can be used to simultaneously identify large numbers of different barcodes in real-time.

WO 2020/086510A1 discloses optically readable capture particle (ORCP) including: one or more optically readable particles (ORPs) each including an optical barcode (the emission spectrum of which can be detected) to identify the ORCP; and a plurality of biological
15 capture sites (e.g., a nucleotide strand configured to capture RNA) associated with the one or more ORPs, each of the plurality of biological capture sites including an oligonucleotide-based cellular barcode to identify the ORCP. However, the optical barcode is provided by the capture particle (e.g., an RNA strand attached to a bead) and is not incorporated into the biological capture sites. Therefore, the optical barcodes disclosed are not genetically
20 encoded.

Given this backdrop, there is a need for a reliable, scalable, and non-intrusive method to barcode individual droplets. Such a method would not only streamline experimental processes but also allow more complex and nuanced experiments, pushing the boundaries of what is achievable with droplet microfluidics.

25 There is a need for an improved method of entity (e.g. droplet) tracking.

There is a need for an improved method of entity (e.g. droplet) identification.

There is a need for improved barcode molecules.

There is a need for improved methods of producing barcode molecules.

30 **Brief summary of the disclosure**

The endeavor to ascertain precise and efficient identification of individual droplets, particularly in contexts demanding high-throughput, has perennially posed significant challenges. The invention described herein, offers an innovative solution predicated upon

the utilization of genetically encoded tag binding regions for fluorescent DNA tags and/or DNA tags linked to a specific peptide (e.g. for unique mass identification).

This methodology may include the incorporation of polyacrylamide beads within each microfluidic droplet. These beads are provided with a barcode that can manifest in two distinct modes (dual-barcode). For example, a fluorescent tag, which upon excitation, emits a distinct fluorescent signature, facilitating real-time visual differentiation of each droplet. This immediate fluorescent readout serves as a tool for instantaneous droplet identification. Alternatively, the tag can be covalently bound to a specific molecule possessing a defined mass. This mass-defined tag, when subjected to mass spectrometry, provides a distinct signature, allowing for droplet identification.

Furthermore, irrespective of the identification method employed the dual barcode is designed to be compatible with next-generation sequencing techniques and is as such, a dual-modality system. This ensures a comprehensive identification mechanism, allowing for both real-time differentiation and retrospective analysis based on sequencing. Sequencing can be performed prior or subsequent to the real-time measurements by either incorporating a target of interest with the barcode or by downstream sequencing of captured DNA (such as the InDrops methods (Zilionis *et al.*, Nat Protoc. 2017 Jan;12(1):44-73).

A feature of this system is its capability to track the nucleic acid content (i.e. dual-barcodes) resident within the droplets. Irrespective of the droplet including cells or specifically tagged molecules therein, their inherent nucleic acid content can be indelibly correlated to a signature of the respective droplet. This ensures that any nucleic acid or molecular signature intrinsic to the droplet's contents can be uniquely identified. In addition to this, droplets can be tracked in real time, allowing for combinatorial mixing of droplets through droplet merging and splitting allowing identification of individual dual-barcodes from both parent droplets.

In summation, the present invention embodies a significant advancement in droplet identification methodologies. By proffering a dual-modality identification mechanism—incorporating both real-time readout and DNA-based retrospective analysis to identify dual-barcodes it has potential to become an indispensable asset for novel applications in droplet microfluidics and biology as a whole. For instance, the dual-barcodes may be compatible with method of synthetic biology, enzymology, pharmacology, and DNA-encoded libraries. The dual-barcodes may also be compatible with diverse analytical methods, including fluorescence, absorbance, mass spectrometry, Raman spectroscopy and high-resolution imaging.

Accordingly, in one aspect, the present invention provides a nucleic acid encoded dual modality barcode (dual-barcode) for linking real time identification and subsequent analysis of one or more entities comprising:

- 5 a. one or more distinct tag binding regions each comprising one or more tag binding sequences for binding to a cognate identification tag;
 - wherein each tag binding region comprises a linker region for linking to an adjacent tag binding region or a support; and
 - wherein the tag binding sequences of each tag binding region are distinct for binding to distinct identification tags; and
- 10 b. at least one amplification region for amplification of the nucleic acid molecule.

Suitably, in the dual-barcode of the invention, the nucleic acid molecule comprises C tag binding regions, wherein C is the number of distinct tag binding regions.

- 15 Suitably, in the dual-barcode of the invention, each tag binding region comprises L tag binding sequences, wherein L is the number tag binding sequences of each distinct tag binding region and wherein each tag binding sequence in each individual tag binding region is for binding the same identification tag.

- 20 Suitably, in the dual-barcode of the invention, the number of distinct nucleic acid molecules (N) is equal to L^C .

Suitably, in the dual-barcode of the invention, each one of the tag binding regions is bound to one or more cognate identification tags and wherein each cognate identification tag comprises:

- 25 a tag sequence that specifically binds to its cognate tag binding sequence; and an identifiable moiety.

Suitably, in the dual-barcode of the invention, the one or more cognate identification tags bound to each distinct tag binding region comprises distinct identifiable moieties.

30

Suitably, in the dual-barcode any of the invention, each tag binding region is bound to two or more of its cognate identification tag.

- 35 Suitably, in the dual-barcode of the invention, the identifiable moiety comprises an optical moiety and/or a defined mass moiety.

Suitably, in the dual-barcode of the invention, each tag binding region comprises the same length.

5 Suitably, the dual-barcode of the invention further comprises a unique molecule identifier sequence (UMI).

Suitably, the dual-barcode of the invention further comprises a target region.

Suitably, in the dual-barcode of the invention, the target region:

- 10 a. is for binding to a cellular target;
 b. comprises a sequence encoding a gene of interest; or
 c. is for binding to a chemical moiety.

15 Suitably, in the dual-barcode of the invention, the nucleic acid molecule is attached to a support.

Suitably, in the dual-barcode of the invention, the support comprises a bead.

20 Suitably, in the dual-barcode of the invention, the support comprises a DNA-encoded library (DEL) member.

Suitably, in the dual-barcode of the invention, the support comprises a protein.

25 Suitably, in the dual-barcode of the invention, the support comprises a cell; optionally wherein the support is comprised within the cell and/or on the cell surface.

Suitably, in the dual-barcode of the invention, the support comprises a chemical agent.

30 In another aspect, the invention provides a barcoded support for linking real time identification and subsequent analysis of one or more entities comprising:

- at least one dual-barcode according to the invention; and
- a support comprising one or more attachment moieties for attachment of the dual-barcode to the support.

35 Suitably, in the barcoded support of the invention, each one of the tag binding regions is bound to one or more of its cognate identification tags according to the invention as disclosed herein.

Suitably, in the barcoded support of the invention, the support comprises a bead, a cell, a protein, a nucleic acid or a chemical agent.

- 5 Suitably, in the barcoded support of the invention, the dual-barcode is attached to the bead via a linker region of the dual-barcode.

Suitably, in the barcoded support of the invention, the one or more attachment moieties comprises a nucleic acid.

10

Suitably, in the barcoded support of the invention, the one or more attachment moieties are integral to the support.

- 15 Suitably, in the barcoded support of the invention, each dual-barcode comprises identical tag binding regions.

Suitably, in the barcoded support of the invention, each one of the at least one dual-barcodes comprises a distinct UMI.

- 20 In a further aspect, the present invention provides a plurality of barcoded supports according to the invention, wherein each support of the plurality comprises dual-barcodes comprising distinct target regions bound thereto to the each other support of the plurality.

- 25 In a further aspect, the present invention provides a method of producing a plurality of barcoded supports according to the invention (as disclosed herein); the method comprising:

- a. providing plurality of first tag binding regions for binding to a cognate first identification tag and comprising at least one linking region for linking to a support and/or adjacent tag binding regions;
- 30 wherein one of the plurality of first tag binding regions comprises L tag binding sequences for binding to L cognate first identification tags; and
- wherein each other one of the plurality of tag binding regions comprises L+x tag binding sequences wherein x is an integer of 1 or more and x is not repeated for the plurality of tag binding regions;
- 35 thereby providing a plurality of first tag binding regions comprising a distinct number of tag binding sequences;

b. attaching in separate vessels at least one of each of the plurality of first tag binding regions comprising a distinct number of tag binding sequences to a support;

5 thereby providing a plurality of supports each attached to at least one first binding region comprising a distinct number of tag binding sequences (Support-L1x).

Suitably, the method of the invention further comprises:

10 c. combining the plurality of Support-L1x and separating a plurality of portions of the plurality of Support-L1x into a plurality of separate vessels;

d. adding to each separate vessel a plurality of a second tag binding regions for binding a second cognate identification tag;

15 wherein one of the plurality of the second cognate tag binding region comprises L second tag binding sequences for binding to L cognate second identification tags; and

wherein each other one of the plurality second tag binding regions comprises L+x second tag binding sequences wherein x is an integer of 1 or more and x is not repeated for the plurality of second tag binding regions;

20 thereby providing a plurality of second tag binding regions each comprising a distinct number of second tag binding sequences (L2x);

e. attaching each of the L2x to the first tag binding region;

25 thereby providing a plurality of supports each attached to at least one first binding region comprising a distinct number of tag binding sequences and a second tag binding region comprising a distinct number of second tag binding sequences (Support-L1x-L2x).

Suitably, in the method of producing a plurality of barcoded supports of the invention; the method comprises:

30 a. providing at least one first tag binding region (TBRa) for binding to a cognate first identification tag and at least one second tag binding region (TBRb) for binding to a cognate second identification tag;

wherein each of the first tag binding region comprises L tag binding sequences for binding to L cognate first identification tags and at least one linking region for linking to a support and/or adjacent tag binding regions; and

wherein the second tag binding region comprise Lb tag binding sequences for binding to Lb cognate second identification tags and at least one linking region for linking to adjacent tag binding regions;

b. linking the second tag binding region (TBRa) to the first tag binding region (Support-TBRa(L) – TBRb(Lb))

thereby producing a dual-barcode;

c. repeating steps (a) to (c) with a plurality of distinct tag binding regions to a produce a plurality of distinct dual-barcodes;

d. attaching each distinct one of the plurality of distinct dual-barcodes to a different support or;

wherein step (a) comprises attaching the first binding region to a support thereby providing a plurality of at least one dual-barcodes each attached to a support.

Suitably, in the method of producing a plurality of barcoded supports according to the invention, before step (d):

f. linking one or more further tag binding regions (TBRy) to the second tag binding region wherein each of the further tag binding region comprises Ly tag binding sequences for binding to Ly cognate further identification tags and at least one linking region for linking to adjacent tag binding regions.

Suitably, in the method of producing a plurality of barcoded supports according to the invention, step (f) is repeated for a plurality of further tag binding regions wherein each of the further tag binding region comprises L+x, Lb+x or Ly+x tag binding sequences wherein x is an integer of 1 or more and x is not repeated (Support-TBR1a(L) – TBRb (L) - TBRy(Lyx)).

Suitably, in the method of producing a plurality of barcoded supports according to the invention, at least one further tag binding region comprises a further first tag binding region or a further second tag binding region or wherein at least one further tag binding region comprises a further tag binding region with Lx tag binding sequences (Support-TBRa(L) – TBRb (Lb) – TBRa(Lx) – TBRb(Lbx) or Support-TBRa(L) – TBRb (Lb) – TBRy(Ly) – TBRy(Lyx)).

Suitably, in the method of producing a plurality of barcoded supports according to the invention, step d involving adding to each separate vessel a plurality of a second tag binding regions for binding a second cognate identification tag and step e involving attaching each of the L2x to the first binding region or step, of the invention, or step (c) involving repeating steps

(a)-(c) with a plurality of distinct tag binding regions to produce a plurality of distinct dual-barcodes or step (f) involving repeating linking of one or more further tag binding regions (TBRy) to the second tag binding region are repeated C times, wherein C is the number of distinct tag binding regions thereby producing N distinct barcoded supports, wherein $N = L^C$.

5

Suitably, in the method of producing a plurality of barcoded supports according to the invention, further comprises binding C distinct cognate identification tags to each tag binding sequence of each distinct tag binding region;

wherein the identification tags are according to the invention as disclosed

10 herein.

Suitably, in the method of producing a plurality of barcoded supports according to the invention, the method further comprises, attaching a distinct UMI to each one of the plurality of dual-barcodes attached to each support.

15

Suitably, in the method of producing a plurality of barcoded supports according to the invention, the method further comprises, attaching a distinct target region to at least one of the plurality of dual-barcodes, wherein each dual-barcode comprises at least one target region distinct from other ones of the plurality of dual-barcodes attached to other supports.

20

In a further aspect, the present invention provides a method of linking real time identity of an entity to subsequent analysis, the method comprising:

- a. providing a plurality of barcoded supports according to the invention;
- b. combining one of the plurality of barcoded supports with one entity;
- 25 c. detecting the identification tag of the barcoded support; and
- d. optionally amplifying each dual-barcode of the plurality of barcoded supports and;
- e. further optionally sequencing the amplified dual-barcodes.

30 Suitably, in the method of linking real time identity of an entity to subsequent analysis according to the invention, the identification tag comprises an optical moiety and detection comprises detecting one or more of optical properties.

Suitably, in the method of linking real time identity of an entity to subsequent analysis
35 according to the invention, the identification tag comprises a defined mass moiety and detection comprises detecting at least a mass of each dual-barcode.

Suitably, in the method of linking real time identity of an entity to subsequent analysis according to the invention, the method comprises a method of RNA detection and the dual-barcode comprises a target region for binding to RNA.

- 5 Suitably, in the method of linking real time identity of an entity to subsequent analysis according to the invention, the optical moiety comprises a fluorophore and detection comprises:
- a. simultaneously illuminating two or more of the fluorophores of distinct identification tags;
 - 10 b. capturing fluorescence signals generated from the two or more fluorophores and separating each signal into distinct channels, each channel corresponding to each fluorophore; thereby allowing for simultaneous capture of multi-spectral fluorescence data from single entities.
- 15 Suitably, the method of linking real time identity of an entity to subsequent analysis according to the invention, further comprises extracting fluorescence measurements from the captured fluorescence signals, extracting comprising:
- a. simultaneously receiving signals including information corresponding to the fluorescence measurements;
 - 20 b. processing the signals; the processing comprising:
 - i. Gaussian blurring the signals to reduce noise;
 - ii. thresholding the Gaussian blurred signals to produce binary masks which denote the presence of entities against a background;
 - iii. refining the binary masks and correcting for artefacts using hole-filling;
 - 25 iv. optionally; wherein the refined binary masks comprises two or more entities in close proximity to each other, resolving the two or more entities using morphological erosion of the binary masks to isolate and delineate individual entities;
 - v. for each entity in the binary mask, edge detecting coordinates along each edge of each entity detected; and
 - 30 vi. applying elliptical fitting to the coordinates.

Suitably, in the method of linking real time identity of an entity to subsequent analysis according to the invention, the method further comprises for each entity detected:

- 35 a. calculating the mean of each data point corresponding to each detected entity to estimate the measurement of each fluorophore in each entity.

Suitably, the method of linking real time identity of an entity to subsequent analysis according to the invention, further comprises:

5 before calculating the mean, employing a fitting error metric for assessing accuracy of the elliptical fitting; and
excluding those detected entities exceeding a defined threshold of the fitting error metric thereby excluding detected entities that do not correspond to a single entity.

10 Suitably, in the method of linking real time identity of an entity to subsequent analysis according to the invention, the sequence of the dual-barcodes of each of plurality of barcoded supports is known prior to step (a) or (b) and the method comprises pairing the known sequences to the detected identification tag.

15 In a further aspect, the invention provides the use of a dual-barcode according to the invention, barcoded support according to the invention or a plurality thereof in a method of:

- a. single cell analyses;
- b. genetic material analysis;
- c. epigenetic analysis;
- 20 d. chemical library analysis;
- e. peptide library analysis;
- f. nucleic acid library analysis;
- g. metabolic library analysis;
- h. secretome analysis;
- 25 i. reaction condition analysis;
- j. reaction analysis;
- k. synthetic biology;
- l. therapeutic agent analysis;
- m. enzymology;
- 30 n. drug interaction analysis;
- o. sample analysis;
- p. pathogen analysis;
- q. pathology;
- r. cytology;
- 35 s. DNA encoded library analysis;
- t. gene engineering;
- u. cell engineering; or

v. *in vitro* transcription and/or in vitro transcription and translation.

In another aspect, the invention provides a kit of parts comprising:

- a. at least one dual-barcode according to the invention; and
- 5 b. at least one cognate identification tag according to the invention.

Suitably, the kit of the invention further comprises one or more further tag binding regions each for binding further cognate identification tags; and optionally one or more cognate further identification tags.

10

Suitably, the kit according to the invention further comprises a support comprising one or more attachment moieties for attachment of the dual-barcode to the support.

Suitably, the kit according to the invention further comprises one or more reagents for
15 attaching the at least one dual-barcode to the support.

Suitably, the kit according to the invention further comprises one or more reagents for linking the at least one dual-barcode to one or more further tag binding regions; and optionally for linking two or more further tag binding regions.

20

In another aspect, the invention provides a method for extracting individual measurements from a plurality of simultaneously captured measurements of a plurality of entities comprising a dual-barcode; the method comprising:

- a. simultaneously receiving, from at least one sensor, signals including information
25 corresponding to the plurality of measurements;
- b. processing the signals; the processing comprising:
 - i. Gaussian blurring the signals to reduce noise;
 - ii. thresholding the Gaussian blurred signals to produce binary masks which denote the presence of one or more of the plurality of barcoded
30 entity against a background;
 - iii. refining the binary masks and correcting for artefacts using hole-filling;
 - iv. optionally; wherein the refined binary masks comprises two or more entities in close proximity to each other, resolving the two or more entities using morphological erosion of the binary masks to isolate and
35 delineate individual entities;

- v. for each entity detected in the binary mask, edge detecting coordinates along each edge of each entity detected; and
- vi. applying elliptical fitting to the detected coordinates; and
- vii. for each entity detected, calculating the mean of each data point corresponding to each detected entity to estimate the measurement in each entity,
- thereby extracting measurements for all entities detected; and
- c. outputting one or more signals indicative of the extracted measurements.

10 Suitably, the method for extracting individual measurements according to the invention, further comprises:

- before calculating the mean, employing a fitting error metric for assessing accuracy of the elliptical fitting; and
- excluding those detected entities exceeding a defined threshold of the fitting error metric thereby excluding detected entities that do not correspond to a single entity.

Suitably, in the method for extracting individual measurements according to the invention, the data is analysed using machine learning.

Suitably, the method for extracting individual measurements according to the invention or the method of linking real time identity of an entity to a subsequent analysis (the step of detecting the identification tag of the barcoded support) are implemented using processing circuitry.

25 In another aspect, the invention provides a method for designing a tag sequence for specifically binding to a tag binding sequence for a dual-barcode according to the invention, the method comprising:

- a. receiving data comprising information corresponding to a nucleic acid sequence, and processing the data by:
- 30 i. generating one or more tag sequences of a defined length picking randomly between the nucleotides A, T, C and G wherein each tag sequence comprises a GC content between 30-60%;
- ii. assessing each tag sequence for (a) self hybridisation, (b) hairpin formation, and (c) secondary structure formation using delta G as an
- 35 indication of a likelihood of (a) to (c) occurring;

- iii. selecting the tag sequences with a likelihood of (a) to (c) occurring below a threshold;
- iv. assessing each tag sequences for binding to a complementary sequence using delta G as an indication of binding affinity;
- 5 v. selecting the tag sequences with a binding affinity to a complementary sequence greater than a threshold;
- vi. assessing each of the selected tag sequences for binding affinity to each other selected tag sequences using delta G as an indication of binding affinity;
- 10 vii. selecting the tag sequences that have a binding affinity to other selected tag sequences lower than a threshold and a binding affinity to a complementary sequence greater than a threshold; and
- viii. assessing the tag sequences for binding to the complementary sequence of other tag sequences using delta G as an indication of binding affinity and selecting those tag sequences that specifically bind their complementary sequence and do not bind to the complementary sequence of other tag sequences.
- 15

Suitably, the method for designing a tag sequence for specifically binding to a tag binding sequence according to the invention, further comprises assessing the binding affinity of each tag sequence selected in steps (c), (e), (g) or (f) to one or more further nucleic acid sequences using delta G as an indication of binding affinity and selecting tag sequences that have a binding affinity for the one or more further nucleic acid sequences below a threshold.

25 In another aspect, the invention provides a library of genetically encoded barcodes comprising a plurality of dual-barcodes according to the invention, wherein each one of the plurality is a distinct dual-barcode.

In a further aspect, the invention provides a method of producing a library of dual-barcodes, the method comprising:

- a. designing one or more tag sequences using a method for designing a tag sequence for specifically binding to a tag binding sequence for a dual-barcode according to the invention, as disclosed herein;
- b. producing a plurality of dual-barcodes according to the invention as disclosed
- 35 herein,

wherein each tag binding sequence comprises a complementary nucleic acid sequence to each of the designed tag sequences; and

wherein each of the dual-barcodes of the plurality are distinct.

- 5 Suitably, in the method of producing a library of dual-barcodes according to the invention, producing comprises *in vitro* synthesis of each distinct dual-barcode or recombinant production of each distinct dual-barcode.

Suitably, in the method of producing a library of dual-barcodes according to the invention, the
10 target region of each distinct dual-barcode comprises a sequence of interest, wherein each sequence of interest is distinct for each distinct dual-barcode;

thereby providing a plurality of distinct dual-barcodes each comprising a distinct sequence of interest and wherein the sequence of each target binding region of distinct dual-barcodes and its distinct sequence of interest are known.

15

Suitably, in the method of producing a library of dual-barcodes according to the invention, the sequence of interest comprises:

a nucleic acid sequence encoding a gene of interest; or

a nucleic acid sequence encoding a protein of interest.

20

Suitably, in the method of producing a library of dual-barcodes according to the invention, the method comprises attaching one or more of each distinct dual-barcode to a different support.

Suitably, in the method of producing a library of dual-barcodes according to the invention, the
25 method comprises attaching one or more of each distinct dual-barcodes to a different support; and

producing comprises a method according to the invention, as disclosed herein.

Throughout the description and claims of this specification, the words “comprise” and
30 “contain” and variations of them mean “including but not limited to”, and they are not intended to (and do not) exclude other moieties, additives, components, integers or steps.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used,

the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

Various aspects of the invention are described in further detail below.

Brief description of the Figures

Embodiments of the invention are further described hereinafter with reference to the accompanying drawings, in which:

Figure 1 shows a schematic of a barcoded bead. Bead (A) is covalently linked through copolymerisation to a linker DNA (attachment moiety - B). The linker DNA is then ligated to the dual barcode (C) or genetic optical barcode (GOB) which includes tag binding regions (a, b and c) for different identifiable tags (1a, 1b, 1c). A unique molecular identifier (UMI) shown as (D) gives each dual-barcode a unique identity and the cell marker (target region) shown as (E) allows for binding of cellular mRNA or other targets (or may encode a gene or protein of interest). Primer binding sites (amplification regions - arrows) can then be used for amplification of the dual-barcode following real-time experimentation.

Figure 2 shows histograms plots showing the melting temperature, heterodimer potential, homodimer potential and hairpin potential for millions of tag binding sequence candidates. These are then gated to select tag binding sequences with the desired properties.

Figure 3 shows a heat map showing the hybridisation potential for tag binding sequences and tag sequences. A proprietary algorithm can choose conditions whereby each tag sequence has the highest affinity for its corresponding tag binding sequence and low affinity for all others.

Figure 4 shows an example dual-barcode showing embedded amplification regions (arrows), tag binding regions (B) including one tag binding sequence each (B), and non-binding regions (NBR) to ensure all tag binding regions have the same length.

Figure 5 shows a plots of potential binding affinity to the dual-barcode for different identifiable tags (e.g. colours) with one tag binding sequence in each tag binding region for the complementary strand to the tag sequence on the left and other strand on the right.

Figure 6 shows plots of potential binding affinity to the dual-barcode for different identifiable tags (colours) with ten tag binding sequence in each tag binding region for the

complementary strand the tag sequence on the left and other strand on the right. There is one binding site for a constant identifiable tag on the linker region.

Figure 7 shows three calibration curves representing the relationship between concentration and measurement for droplets containing aqueous DNA-conjugated fluorophores with wavelengths at 488, 561, and 647 nanometres. Each plot shows a linear regression line (the sloped line) depicting the line of best fit, with error bars indicating the mean and standard deviation at each concentration point. R^2 values are 0.9607, 0.9455, and 0.9706 for the 488, 561, and 647 assays, respectively, indicating strong positive correlation between concentration and measurement across the assays.

Figure 8 shows Left: A fluorescent (488nm) image of barcoded beads, showing a mixed population. Right: Histogram of barcoded beads, showing high background and some distinct populations. Loading concentration of dual-barcodes on the beads was 5 μ M. Image taken on an Evos FL microscope.

Figure 9 shows Top Left: A fluorescent (488nm) image of barcoded beads showing a clearer population when beads are loaded with 25 μ M of dual-barcodes. Top Right: Masked fluorescent image showing identifiable barcoded beads that have clustered into 3 populations. Squares represent barcoded beads in cluster 1; circles represent barcoded beads in cluster 3; the remaining barcoded beads represent cluster 2. Bottom: Histogram of the identified barcoded beads showing 3 clear populations. Image taken on an Evos FL microscope.

Figure 10 shows a schematic different tag binding regions (TBRa, b and c). TBRa is bound to 3 identifiable tags (circles) via 3 tag binding sequences (1a, 2a and 3); TBRb is bound to 3 identifiable tags (trapezoids) via 3 tag binding sequences (1b, 2b, and 3b); and TBRc is bound to 1 identifiable tag (triangle) via 1 tag binding sequence (1c).

Figure 11 shows a schematic of a dual-barcode. Tag binding regions and each bound identifiable moiety is shown as in Figure 10. The attachment region is shown by cross hatching; linker regions are shown by left diagonal lines; amplification regions are shown as checkered regions; a UMI is shown by straight horizontal lines; and a target region is shown by zig-zagged lines.

Figure 12 shows schematics of the methods disclosed herein. The upper figure shows a schematic for building a barcode to allow for distinguishability of colours based upon an ordered split and mix method of production as described herein. For example, to create all possible barcodes e.g., tag binding regions a, a, a; a, a, b; a, a, c, etc. the following procedure could be carried out whereby "barcode 1", a dual-barcode containing one tag binding region is pipetted into wells of a 96-well plate (and attached to e.g., a bead), then

“barcode 1” is split into multiple wells of a 96-well plate and to each well a further tag binding region is incorporated to produce “barcode 2” containing now two tag binding regions.

“Barcode 2” is then split into multiple wells of a 96-well plate and a further tag binding region introduced to produce “barcode 3” having 3 tag binding regions. By way of further example,
5 one well could have a dual-barcode having a, a tag binding regions, another well will have a dual-barcode with a, b tag binding regions and a further well will have a dual barcode with a, c tag binding regions. These dual barcodes will then be split again to introduce a further tag binding region to produce a dual-barcode comprising three tag binding regions. The lower figure shows a schematic for emulsion PCR on a bead.

10 **Figure 13** shows maps of the biological sequences disclosed herein, namely the sequences of SEQ ID NOs: 1 to 32.

The patent, scientific and technical literature referred to herein establish knowledge that was available to those skilled in the art at the time of filing. The entire disclosures of the issued patents, published and pending patent applications, and other publications that are cited
15 herein are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference. In the case of any inconsistencies, the present disclosure will prevail.

Various aspects of the invention are described in further detail below.

20 **Detailed Description**

Nucleic acid Molecule– Barcode

Provided herein are nucleic acid molecules that can be used for tracking and identification of entities as described herein (e.g. a bead, cell, cellular component or chemical moiety). The
25 nucleic acid molecules can be used to “barcode” entities both by the sequence of the isolated nucleic acid molecules as well by identifiable properties imparted by the identifiable moieties as described herein. The nucleic acid molecules as described herein therefore allow for entities to be detected, distinguished from each other and tracked in real-time (e.g. via the identifiable moieties attached thereto) and to allow specific entities to be identified
30 (e.g. via the nucleic acid sequence of the isolated nucleic acid molecule). This dual modality of nucleic acid molecules allows for the distinct nucleic acid sequence of each of the nucleic acid molecules to be correlated to the unique code imparted by distinctive identifiable moieties of the respective entity. As used herein the term “dual-barcode” is used interchangeably with “a nucleic acid encoded dual modality barcode” to refer to nucleic acid
35 molecules of the invention.

A barcode in general terms is a nucleic acid tag added to an entity to identify each entity including the same barcode as a group. For example, a barcode may be added to, a bead which may then be incorporated into a droplet to identify a droplet or the contents thereof, a cell to identify an individual cell, and/or a cellular component such as the RNA within a single cell to identify the RNA as from that cell.

In some examples, the dual-barcode is an isolated nucleic acid molecule. An "isolated nucleic acid" refers to a nucleic acid or fragment which has been separated from sequences which flank it in a naturally occurring state, i.e., a DNA fragment which has been removed from the sequences which are normally adjacent to the fragment, i.e., the sequences adjacent to the fragment in a genome in which it naturally occurs. The term also applies to nucleic acids which have been substantially purified from other components which naturally accompany the nucleic acid, i.e., RNA or DNA or proteins, which naturally accompany it in the cell. The term also includes, for example, a recombinant DNA which exists as a separate molecule such as a cDNA or a genomic or cDNA fragment produced by PCR, or a synthetic nucleic acid independent of other sequences.

The nucleic acid and constituent parts thereof (e.g. binding region, linker region, amplification region, attachment region, UMI, target binding region etc.) may be produced by any suitable methods known in the art.

In some examples, the dual-barcode may include or be DNA. Where the dual-barcode is DNA, the dual-barcode may be produced using solid-phase DNA synthesis or DNA printing (such as column-based oligonucleotide synthesis, or microarray-based oligonucleotide synthesis) phosphoramidite-based synthesis, or recombinant technology methods. For example, see Hughes, Randall A., and Andrew D. Ellington. "Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology." Cold Spring Harbor perspectives in biology 9.1 (2017): a023812 which is incorporated herein by reference.

In some examples, dual-barcode comprises single stranded DNA (ssDNA). In some examples, the dual-barcode is ssDNA. In some examples, the dual-barcode comprises double stranded DNA (dsDNA). In some examples, the dual-barcode is double stranded DNA (dsDNA).

Whether the dual-barcodes are ssDNA or dsDNA may be determined by the methods used for producing or assembling the dual-barcodes. For example, when the dual-barcodes are assembled using methods such as Golden Gate or Gibson assembly the dual-barcodes may be dsDNA. When the dual-barcodes are produced using methods such as splint ligation they may be ssDNA (with the expectation of tag binding sequences bound by tag sequences).

Tag binding region

Each dual-barcode includes one or more distinct tag binding regions. Each tag binding region is configured to bind to different distinguishable identification tags. That is to say that each binding region is configured to bind a different identifiable tag to each other tag binding region comprised within a single dual-barcode. Therefore, each tag binding region is configured to bind to at least one of its cognate identifiable tag. As used herein "cognate" refers to components that function or specifically interact together. Thus, in the context of the present invention, a cognate pair refers to a tag binding region that specifically binds to a corresponding identifiable tag.

For example, a dual-barcode may include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 or more tag binding regions.

In some examples, each tag binding region comprises ssDNA. In some examples, each tag binding region is ssDNA.

The sequences of the tag binding region is in part defined by the sequence of the tag binding sequences.

In some examples each tag binding region (of a single dual-barcode or all dual-barcodes in a plurality) may have the same number of nucleotides (i.e. same length).

In some examples, each tag binding region includes one or more non-binding regions (NBRs). NBRs may include arbitrary nucleic acid sequences which are designed to not bind to a tag sequence. The inclusion of NBRs in each tag binding region may reduce or prevent non-specific annealing and mis-annealing of tag sequence as well as self-annealing of the dual-barcodes.

The number of NBRs in each tag binding region may be selected so as to provide each tag binding region with same number of nucleotides. In some examples, the length and/or number of NBRs may be selected to provide each dual-barcode of the invention of plurality of dual-barcodes with the same length (i.e. each nucleic acid molecule has the same total number of nucleotides). Providing each dual-barcode produced in a plurality of dual-barcodes with the same length may help to allow for uniform amplification during analysis.

In some examples, where the dual-barcode includes multiple tag binding sequences, the number of NBRs in each tag binding sequences may differ based on the number of tag binding sequences in each tag binding region. For example, a first tag binding region may include 8 tag binding sequences and 2 NBRs, a second tag binding region in the same dual-barcode may include 5 tag binding sequences and 5 NBRs wherein each tag binding sequence and NBR is the same number of nucleotides, therefore providing tag binding

regions that are the same length. Ergo, if the other regions of the dual-barcode of a plurality of dual-barcodes are the same, the number of NBRs may provide all the dual-barcodes with the same length. NBRs may be dsDNA. In some examples, NBRs are ssDNA.

Tag binding sequence

- 5 Each tag binding region comprises one or more tag binding sequences. Each tag binding sequence in a tag binding region binds to the same cognate identifiable tag. Thus allowing for the binding of one or more of the same cognate identifiable tag to a single tag binding region.

For example, a tag binding region may include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
10 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 or more tag binding sequences.

Therefore, each tag binding region in a dual-barcode may include 1 or more tag binding sequence each of which is configured to bind or is bound to one cognate identifiable tag. The dual-barcode may also include further tag binding regions each including 1 or more tag binding sequence each of which is configured to bind or is bound to one cognate identifiable
15 tag. For example, with reference to Figure 10, a dual-barcode may have a first tag binding region (TBRa) including three tag binding sequences (1a, 2a and 3a) each bound to a first identifiable tag (circles); a second tag binding region (TBRb) including three tag binding sequences (1b, 2b, and 3b) each bound to a second identifiable tag (trapezoids); and a third
20 tag binding region (TBRc) including one tag binding sequence (1c) bound to one identifiable tag (triangle).

Using different combinations of the number of tag binding regions and combinations of the number of tag binding sequences in each tag binding region, a multitude of distinct dual-barcodes can be produced.

For example, the dual-barcode may include C tag binding regions, where C is the number of
25 distinct or different tag binding regions. Each tag binding region can include L tag binding sequences, where L is the number of tag binding sequences of each distinct tag binding region. Each tag binding sequence in each individual tag binding region may be configured or be bound to the same identification tag. Therefore, the number of distinct or different dual-barcodes that can be produced (N) is defined by:

$$30 \quad N = L^C$$

For example, for a dual-barcode including three tag binding regions, each with 10 tag binding sequences, the number of distinct dual-barcodes is 10^3 or 1000. For a dual-barcode including 4 tag binding regions each having 40 tag binding sequences, the number is 40^4 or 2.56 million.

In some examples, a dual-barcode may have only one tag binding region which has a defined number of tag binding sequences. In such cases, different dual-barcodes may be distinct or distinguishable from each other by the number of tag binding sequences in each dual-barcode and ergo the number of identifiable tags bound thereto in use. For example, if
5 the identifiable tag includes a fluorescent identifiable moiety as described herein, each subsequent identifiable tag bound to a binding sequence may provide a plurality of dual-barcodes that each emit a different intensity based on the number of identifiable tags bound thereto (i.e. the number of tag binding sequences). In another example, the identifiable tag may include an identifiable moiety that has a detectable mass. As such, with increasing
10 number of tag binding sequences (and ergo bound identifiable tags) the mass of the each dual-barcode may be different (i.e. increase with the number of bound identifiable tags).

The tag binding sequences may be different or the same within a single binding region. In the case that the tag binding sequences are the same, a first tag binding region may be attached to a tag binding sequence to, for example, a bead, and a first tag binding region
15 that has 2 tag binding sequences that are the same may be attached to a different bead. Alternatively, the tag binding regions may be distinct from one another.

Each tag binding sequence may be comprised of ssDNA. The nucleic acid sequence of the tag binding sequence may be designed using a method as described herein. For example, the tag binding sequence is configured to selectively and specifically bind to its cognate
20 identifiable tag. For example, by binding to a cognate tag sequence of an identifiable tag.

The term “specifically bind” refers to a molecule (e.g., such as a tag sequence of an identifiable moiety) that binds to a target (e.g. a cognate tag binding sequence) with at least 2-fold greater affinity than non-target compounds (e.g. non-cognate tag binding sequences), e.g., at least 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 20-fold, 25-fold, 50-fold or
25 more. The term “selectively bind” refers to the preferential binding of a tag binding region to its cognate identifiable moiety or vice versa. Generally, the tag binding region will possess little or no binding to other tag sequences of non-cognate identifiable moieties.

As such, each binding tag binding sequence is configured to only bind to its cognate tag sequence of an identifiable tag. The tag sequence of the identifiable tag may include a
30 nucleic acid sequence that is complimentary to the nucleic acid sequence of the tag binding sequence. Thus allowing the specific hybridization of each tag binding region to its cognate tag sequence. “Specific hybridization” refers to the binding of a nucleic acid (e.g. tag sequence) to a target nucleotide sequence (e.g. tag binding sequence) in the absence of substantial binding to other nucleotide sequences present in the hybridization mixture (e.g.
35 other regions of dual-barcode such as other tag binding regions, linker regions, attachment

regions, and/or amplification region), under defined stringency conditions. Those of skill in the art recognize that relaxing the stringency of the hybridization conditions allows sequence mismatches to be tolerated. In particular examples, hybridizations are carried out under stringent hybridization conditions, thus only allowing for cognate tag sequence hybridization to its cognate tag binding sequence.

The nucleic acid sequence of each tag binding sequence or its cognate tag sequence may be selected using the methods described herein. For example, each tag binding sequence or tag sequence may comprise a defined number of randomly selected nucleotides that have a Gibbs free energy (ΔG) of hybridization at a given temperature (i.e. ΔG_T) that is relatively low. For example, a Gibbs free energy for hybridization of at most -14000 J. Thus providing a tag binding sequence and complementary tag sequence that selectively and specifically bind to each other. For example, having minimised non-specific binding, minimised mis-annealing, and/or maximised complementary sequence binding.

The nucleic acid sequence of each tag binding sequence or its cognate tag sequence may also be configured to minimise self forming secondary structures, such as hairpins (i.e. a relatively high ΔG for secondary structure formation).

The nucleic acid sequence of each tag binding sequence or its cognate tag sequence may also be configured to minimise self-annealing (i.e. a relatively high ΔG for annealing to itself).

The nucleic acid sequence of each tag binding sequence or its cognate tag sequence may also be configured to minimise non-specific binding (i.e. a relatively high ΔG for annealing to other tag binding regions and/or non-cognate tag sequences, as well as other regions of the dual-barcode such as linker regions, attachment regions, and/or amplification region).

For example, a Gibbs free energy for self forming secondary structures, non-specific binding, homodimerization, self-annealing, mis-annealing, of at least -2000 J.

In some examples, each tag binding sequence and its complementary tag sequence comprise a GC content of between 30 to 50%.

In some examples, the each tag binding sequence may be at least 8 bp in length. For example, at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 bp or more. In some examples, each tag binding sequence is from 8 to 50 bp. In some examples, each tag binding sequence is from 8 to 40 bp. In some examples, each tag binding sequence is from 8 to 30 bp. In some examples, each tag binding sequence is from 8 to 20 bp. In some examples, each tag binding sequence is from 8 to 15 bp.

Methods and systems for designing and selecting suitable tag binding sequences and tag sequences are described herein. Alternative methods, will be known by those skilled in the

art, for example, see Matveeva, O V et al. "Thermodynamic calculations and statistical correlations for oligo-probes design." Nucleic acids research vol. 31,14 (2003): 4211-7. doi:10.1093/nar/gkg476 and Breslauer, K.J.; Frank, R; Blöcker, H; Marky, LA; et al. (1986). "Predicting DNA Duplex Stability from the Base Sequence". Proc. Natl. Acad. Sci. USA. 83
 5 (11): 3746–3750. Bibcode:1986PNAS...83.3746B which are incorporated herein by reference.

In some examples, each tag binding sequence in a single tag binding region comprise the same nucleic acid sequence. In some examples, each tag binding sequence in a single tag binding region comprise a distinct nucleic acid sequence for binding to its cognate tag
 10 sequence of an identifiable tag.

Linker region

Each tag binding region includes at least one linker region. The linker region is configured to allow adjacent tag binding regions to be bound to each other and/or for attachment of a tag binding region to a support (which is referred to herein as an attachment region and is
 15 described below).

In some examples, the first tag binding region may include one linker region for binding to a second tag binding region. The second tag binding region having a first linker region for linking to the first tag binding region and a second linker region at an opposing end of the second tag binding region for linking to a further tag binding region.

20 In some examples, the linker region is a nucleic acid sequence comprised within one or more tag binding regions. For example, at a distal end of a tag binding region, e.g. at the 3' or 5' end of a tag binding region.

A final tag binding region (i.e. located at the opposing end to the first tag binding region) may have only one linker region for linking to the penultimate tag binding region. For example,
 25 see Figure 11. In some examples, the linker region is compatible with a corresponding linker region of a UMI and/or target region as described herein which may be attached to the final tag binding region.

In some examples, the linker region may be a nucleic acid sequence encoding a restriction enzyme recognition site. Each linker region may comprise compatible restriction enzyme
 30 recognitions sites. "Compatible restriction enzyme recognition sites" refers to different restriction sites that, when cleaved, yield nucleotide ends that can be ligated without any additional modification.

In some examples, the linker region is compatible with Golden Gate assembly methods. "Golden Gate assembly" refers to a molecular method comprising assembly of multiple DNA

fragments into a single template using Type IIS restriction enzymes and T4 DNA ligase. The assembly is performed *in vitro*, and common enzymes including BsaI, BsmBI, and BbsI. The methods, techniques, and optimizations involved with performing a Golden Gate assembly reaction are publicly available and are well known in the art. As is known in the art,

- 5 destination vectors are used in the Golden Gate assembly processes. Such vectors are publicly available and known in the art, and the skilled artisan is well capable of identifying, making, or purchasing an appropriate destination vector. For example, see US10865407B2.

In some examples, each linker region comprises a IIS type restriction enzyme recognition site and cleavage site.

- 10 When linker regions include restriction enzyme recognition sites the target binding regions do not include restriction enzyme recognition sites thereby avoiding restriction digestion of target binding regions when linking adjacent target binding regions.

- In some examples, the linker region is compatible with Gibson assembly. "Gibson assembly" refers to a molecular cloning method which joins multiple DNA fragments into a single
15 template in a single, isothermal reaction. In Gibson assembly, the DNA fragments include a ~20-40 base pair overlap with adjacent DNA fragments. These DNA fragments are mixed with a cocktail of three enzymes, along with other buffer components. The three required enzyme activities are exonuclease, DNA polymerase, and DNA ligase. Briefly, the exonuclease chews back DNA from the 5' end, thus not inhibiting polymerase activity and
20 allowing the reaction to occur in one single process. The resulting single-stranded regions on adjacent DNA fragments can anneal. The DNA polymerase incorporates nucleotides to fill in any gaps. The DNA ligase covalently joins the DNA of adjacent segments, thereby removing any nicks in the DNA. The resulting product is different DNA fragments joined into one. As is known in the art, destination vectors are used in the Gibson assembly processes. Such
25 vectors are publicly available and known in the art, and the skilled artisan is well capable of identifying, making, or purchasing an appropriate destination vector. By way of example, but not by way of limitation, pJL1 vectors can be used. There are two approaches to Gibson assembly. A one-step method and a two-step method. Both methods can be performed in a single reaction vessel. The Gibson assembly 1-step method allows for the assembly of up to
30 5 different fragments using a single step isothermal process. In this method, fragments and a master mix of enzymes are combined and the entire mixture is incubated at 50 °C for up to one hour. For the creation of more complex constructs with up to 15 fragments, or for constructs incorporating fragments from 100 bp to 10 kb, the Gibson assembly two-step approach is used. The two-step reaction requires two separate additions of master mix. One
35 of the reactions is for the exonuclease and annealing step while the other is for DNA polymerase and ligation steps. For the two-step approach, different incubation temperatures

are used to carry out the assembly process. For example see <https://synbio.org.uk/dna-assembly/guidetogibsonassembly.html> the contents of which are incorporated herein by reference.

Other methods of linking nucleic acids will be known by the skilled person and therefore the
5 linker regions may be compatible with such methods.

In some examples, when the dual-barcodes are ssDNA, the linker region may be compatible with splint ligation methods. Splint ligation refers to a method of ligating a 5' end of one ssDNA to a 3' end of another ssDNA using a splint strand and a ligase enzyme such as DNA
10 Ligase. For example, see Kershaw CJ, O'Keefe RT. Splint ligation of RNA with T4 SplintR®
Ligase. Methods Mol Biol. 2012;941:257-269. doi:10.1007/978-1-62703-113-4_19.

In some examples, where the dual-barcodes are produced by nucleic acid synthesis the linker region may be an arbitrary sequence which is located between two adjacent tag binding regions. A specific linker region would not be required for dual-barcodes produced by synthesis as the component parts of the dual-barcode would not require linking.
15 In some examples, the linker region comprises dsDNA. In some examples, the linker regions comprise ssDNA.

Amplification region

The dual-barcodes of the invention include at least one amplification region. A amplification region may be a primer binding site, for example, for binding to an amplification primer. For
20 example, for binding to a PCR primer that allows for amplification of the dual-barcode by PCR. Each amplification region in a dual-barcode may have a nucleic acid sequence configured to bind to a different primer. Therefore allowing amplification of different parts of the dual-barcode.

As used herein, the term "primer" refers to an oligonucleotide which is capable of annealing
25 to a polynucleotide target and serving as a point of initiation of DNA synthesis when placed under conditions in which synthesis of a primer extension product is induced (e.g., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH). A primer (in some examples an extension primer and in some examples an amplification primer) may be single stranded for maximum efficiency in
30 extension and/or amplification. A primer is typically sufficiently long to prime the synthesis of extension and/or amplification products in the presence of the agent for polymerization. The minimum length of the primer can depend on many factors, including, but not limited to temperature and composition (A/T vs. G/C content) of the primer.

As such, each amplification region may comprise a nucleic acid sequence designed in a similar manner to a primer as described above that is complimentary to the primer. In some examples, each amplification region is configured to selectively and specifically bind to its complementary primer sequences. For example, the amplification region is configured to not
 5 bind to a tag sequence of any one of the identification tags and is configured to only bind to its complimentary primer sequence.

Amplification of the dual-barcode may comprise non - PCR based methods . Examples of non – PCR based methods include, but are not limited to, multiple displacement amplification (MDA), a ligase chain reaction (LCR), a QB replicase (QB) method, use of
 10 palindromic probes, strand displacement amplification, oligonucleotide-driven amplification using a restriction endonuclease, an amplification method in which a primer is hybridized to a nucleic acid sequence and the resulting duplex is cleaved prior to the extension reaction and amplification, and strand displacement amplification using a nucleic acid polymerase lacking 5 ' exonuclease activity. In such cases, the skilled person would be able to readily
 15 design and select suitable amplification regions depending on the method of amplification to be used.

At least one amplification region may be located at a distal end of the dual-barcode. For example, at or in close proximity to the 3' or 5' end of the dual-barcode. In some examples, at least one amplification region is located adjacent to the start of the first target binding
 20 region and/or adjacent to the end of the last target binding region. Positioning of the at least one amplification region at one or more ends of the one or more tag binding regions allows for the amplification and subsequent sequencing of all of the tag binding regions within each dual-barcode.

For example, the dual-barcode may have a structure as shown below (from 5' to 3'):

25

Attachment region – optional Amp – q(TBRs)– optional Amp –

where “Amp” is an amplification region as described herein and “TBR” is a tag binding region and q is an integer of 1 or more.

30 In some examples, when the dual-barcode includes a target region as described herein at least one amplification region may be positioned at a distal end of the target region. For example, the dual-barcode may have a structure as shown below (from 5' to 3'):

Attachment region – optional Amp – q(TBRs) – optional Amp – target region – optional Amp

Positioning of at least one amplification region at a distal end (i.e. 3' end) of the target region provides for amplification of target region and the other components of the dual-barcode (e.g. the target binding regions).

In some examples, each one of the amplification regions may be configured for use with different methods of amplification. For example, when the target region is for binding to an RNA target the dual-barcode may include at least one amplification region adjacent to the target region that is configured to allow amplification of the RNA target. In some examples, the dual-barcode may include at least one amplification region adjacent to the target region that is configured to allow amplification of the RNA target and the dual-barcode. For example, suitable amplification regions include those configured for amplification methods such as transcription-mediated amplification (TMA), nucleic acid sequence-based amplification (NASBA), multiple cycles of DNA-dependent RNA polymerase-driven RNA transcription amplification or RNA-directed DNA synthesis and transcription to amplify DNA or RNA targets.

For example, when the target region is for binding to a DNA target the dual-barcode may include at least one amplification region adjacent to the target region that is configured to allow amplification of the DNA target. In some examples, the dual-barcode may include at least one amplification region adjacent to the target region that is configured to allow amplification of the DNA target and the dual-barcode.

Amplification of both a target molecule and the nucleic acid may allow for the production of a barcoded target sequence. For example, the amplification may produce a nucleic acid sequence comprising the nucleic acid sequence or complement thereof of the target and the dual-barcode of the invention. The amplified barcoded target sequence may then be analysed, for example sequenced and be identifiable by virtue of the sequence of the distinct dual-barcode of the invention or complement thereof which is incorporated by the amplification into the sequence of the barcoded target sequence.

In some examples, the dual-barcodes of the invention include amplification regions located at a distal end of each constituent part of the dual-barcode. For example, an application region may be located at a distal end of the attachment region, at a distal end of each target binding region, a distal end of each linker region, at a distal end of a target region, and/or a distal end of a unique molecule identifier (UMI). Each amplification region may include a distinct nucleic acid sequence allowing selection of which parts of the dual-barcode may be amplified by selecting a corresponding primer to each amplification region.

The provision of multiple amplification sites as described above may allow for testing and quality control of the dual-barcodes throughout the production of the dual-barcodes as well as after production. For example, amplification may be carried out at different time points of production or at different steps of production of the dual-barcodes in order to assess the structure of dual-barcodes. For example, after addition of each constituent region or part in order to check that each constituent part is linked or attached correctly.

Each amplification region may be ssDNA or dsDNA.

Attachment region

The dual-barcodes of the invention include a linker region for attachment of the dual-barcode to a support. The linker region for attaching to a support may be referred to herein as an attachment region. The attachment region is configured for binding of the dual-barcode to a support as described herein. "Attachment" refers to immobilization of dual-barcodes on supports by either a covalent attachment or via irreversible passive adsorption or via affinity between molecules.

In some examples, the attachment region is a nucleic acid sequence comprised within one or more tag binding regions. For example, at a distal end of a tag binding region, e.g. at the 3' or 5' end of a tag binding region. In some examples, the attachment region may be the same as a linker region. This may allow for attachment of any tag binding region to the support (i.e. as the first tag binding region in a series of tag binding regions). For example, the linker region and attachment region may both be nucleic acid sequences configured for use with restriction enzyme based assembly methods such as Golden Gate assembly or configured for use with Gibson assembly methods. For example, the attachment region may include a type IIS restriction enzyme recognition site that can provide an end of the attachment region that is compatible with a corresponding type IIS restriction site of a support.

The attachment region may bind to a support by any type of binding interaction. For example, binding of the attachment region (and ergo the dual-barcode) may be by covalent binding, ionic binding, hydrogen binding, and/or Van Der Waals interactions. In some examples, the attachment region is bound to a support by covalent binding.

In some examples, the attachment region comprises a nucleic acid sequence for binding to a support. In some examples, the attachment region may include a functional group for attachment to a support including a compatible functional group (i.e. functional group binding partner). In some examples, the attachment region may include a nucleic acid that includes a functional group, such as chemical group, for binding to a functional group of a support.

The attachment region may bind to a support via a functional group that is included on and/or within a support. For example, a functional group may be a compatible nucleic acid or chemical moiety with which a part of the attachment region or a functional group thereof can interact and bind to.

- 5 Methods of attaching to barcodes such as the dual-barcodes described herein to supports such as microparticles, cells, proteins, cellular components and other nucleic acids are well known in the art.

- The attachment region functional group and the support functional group may include, for example, amines, hydroxylamines, hydrazines, hydrazides, thiols, phosphines,
- 10 isothiocyanates, isocyanates, N-hydroxysuccinimide (NHS) esters, carbodiimides, thioesters, haloacetyl derivatives, sulfonyl chlorides, nitro and dinitrophenyl esters, tosylates, mesylates, triflates, maleimides, disulfides, carboxyl groups, hydroxyl groups, carbonyldiimidazoles, epoxides, aldehydes, acyl-aldehydes, ketones, azides, alkynes, alkenes, nitrones, tetrazines, isonitriles, tetrazoles, and boronates. Examples of reactions that may be carried
- 15 out to allow attachment include the reaction between an amine and an activated carboxy group forming an amide, between a thiol and a maleimide forming a thioether bond, between an azide and an alkyne derivative undergoing a 1,3-dipolar cycloaddition reaction, between an amine and an epoxy group, between an amine and another amine functional group reacting with an added bifunctional linker reagent of the type of activated bis-dicarboxylic
- 20 acid derivative giving rise to two amide bonds, or other combinations known in the art. Other reactions, such as UV-mediated cross-linking can be used for covalent attachment of the attachment region to supports.

- The functional groups may be inherently present in the support or they may be provided by treating or coating the support with a suitable material. The functional group may also be
- 25 introduced by reacting the support surface with an appropriate chemical agent. Activation as used herein means a modification of a functional group on the support surface to enable coupling of an attachment region to the surface.

- In some examples, the attachment region include a restriction enzyme recognition site configured to allow restriction enzyme based attachment to a support. For example, a
- 30 support may include a compatible nucleic acid that includes a compatible restriction enzyme site to allow the attachment region to attach thereto. In some examples, the linker region is compatible with Golden Gate assembly methods as described herein. For example, the attachment region may include a type IIS restriction site and the support may include a nucleic acid including a compatible type IIS.

The attachment region may be located at a distal end of the dual-barcode. For example, at the 5' end or 3' end of the dual-barcode. For example, the dual-barcode may have a structure as shown below (from 5' to 3'):

- 5 Attachment region – optional Amp – q(TBRs) – optional Amp – target region – optional Amp
or

optional Amp – q(TBRs) – optional Amp – target region – optional Amp – Attachment region

where q is the number of tag binding regions, which may be an integer of 1 or more.

- The attachment region may selected depending on the support being used. For example, in
10 examples where the support is a solid support such as a bead the attachment region may be a dual-barcode that binds to a binding partner that is presented on the surface of bead. For example, the bead may include a complimentary nucleic acid sequence with which the attachment region selectively and specifically hybridizes. In some examples, a bead may include a nucleic acid that is configured for use with Golden Gate assembly.

- 15 Additional examples of specific binding pairs allowing covalent binding of dual-barcodes (i.e. an attachment region) to a solid support are e.g. SNAP-tag® (New England Biolabs, Ipswich, Mass.)/AGT and benzylguanine derivatives (U.S. Pat. Nos. 7,939,284; 8,367,361; 7,799,524; 7,888,090; and 8,163,479) or pyrimidine derivatives (U.S. Pat. No. 8,178,314), CLIP-tag™ (New England Biolabs, Ipswich, Mass.)/ACT and benzylcytosine derivatives
20 (U.S. Pat. No. 8,227,602), HaloTag® (Promega, Madison, Wis.) and chloroalkane derivatives (Los, et al. Methods Mol Biol., 356:195-208 (2007)), serine-beta-lactamases and beta-lactam derivatives (International Patent Application Publication No. WO2004/072232). In such as examples, dual-barcodes can be functionalized with benzylguanine, pyrimidine, benzylcytosine, chloroalkane, or beta-lactam derivatives respectively, and subsequently be
25 captured in a solid support modified with SNAP-tag/AGT, CLIP-tag/ACT, HaloTag or serine-beta-lactamases. Alternatively, dual-barcodes can be specifically or nonspecifically attached to SNAP-tag/AGT, CLIP-tag/ACT, HaloTag or serine-beta-lactamases and subsequently be captured in a solid support functionalized with benzylguanine, pyrimidine, benzylcytosine, chloroalkane, or beta-lactam derivatives, respectively. Further examples of specific binding
30 pairs allowing covalent binding of dual-barcodes to a solid support are acyl carrier proteins and modifications thereof (binder proteins), which are coupled to a phosphopanthetheine subunit from Coenzyme A (binder substrate) by a synthase protein (U.S. Pat. No. 7,666,612). Examples of proteins or fragments thereof allowing convenient binding of dual-barcodes to a solid support are e.g. chitin binding domain (CBD), maltose binding protein

(MBP), glycoproteins, transglutaminases, dihydrofolate reductases, glutathione-S-transferase al (GST), FLAG tags, S-tags, His-tags, and others known to those skilled in the art. Typically, a nucleic acid module is modified with a molecule which is one part of a specific binding pair and capable of specifically binding to a partner covalently or non-covalently attached to a solid support.

In some examples, wherein the support is a cell, the attachment region may include a sequence that is configured to bind to a surface marker of the cell. For example, the cell may express a cell surface protein or have a protein attached to its surface which can be specifically bound by the attachment region. For example, a cell may have an antibody or fragment thereof presented on the cell's surface which includes an epitope binding region which specially binds to the attachment region. In other examples, the cell may include one or more of the functional groups described above on its surface.

Methods of determining suitable nucleic acid sequence that bind to surface markers of cells are known in the art. For example, the attachment region may include a sequence determined by methods such as Cell-SELEX. For example, the attachment region may include a sequence that selectively and specifically binds to a pre-determined surface protein. A cell may be modified to overexpress or recombinantly express a surface protein to allow increased attachment or binding of the attachment region to the cell. For example, see Mali, Prashant et al. "Barcoding cells using cell-surface programmable DNA-binding domains." Nature methods vol. 10,5 (2013): 403-6. doi:10.1038/nmeth.2407.

In some examples, the support is or includes a peptide. Attachment of nucleic acids to peptides is well known in the art and the attachment region may be configured for attachment to a peptide support by any suitable method. For example, the attachment region may include a functional group as described above that is capable of interacting with and binding to a side chain of a peptide. For example, using Michael acceptors (ex. maleimides) for cysteine side chain thiol targeting, activated esters (e.g. N-Hydroxysuccinimide (NHS)-esters) for lysine side chain ϵ -amine targeting, diazocarbonyl amides for tyrosine phenol targeting, and N-terminal (α -amine) protein labelling via 2-pyridinecarboxaldehydes.

As used herein, 'peptide' may refer to a molecule comprising at least two amino acid residues linked by peptide (e.g., amide) bonds. The term 'peptide' may refer to amino acid dimers, trimers, oligomers, or polymers. The term 'peptide' may also refer to a protein. A peptide may be linear or branched. A peptide may comprise a natural amino acid. A natural amino acid may be a 'proteinogenic amino acid', which, as used herein, may refer to any one of the 22 known amino acids utilized for translation by natural organisms. A peptide may comprise an isomeric variant of a naturally occurring amino acid, such as an α -carbon

enantiomer, also known as a D-amino acid. A peptide may comprise a non-natural (e.g., synthetically derived) amino acid. A non-natural amino acid may comprise a non-natural side chain, such as a perfluorinated aryl or alkyl moiety. A non-natural amino acid may comprise a non-natural backbone structure, for example a silicon in place of the α -carbon or the amine
5 disposed on a β -carbon. A peptide may also comprise non-amino acid units, such as 4-hydroxybutanoic, in place of amino acid residues.

Peptides can also be produced to include handles for nucleic acid (i.e. attachment region) binding. For example, peptides may be configured for copper-catalyzed and strain promoted azide-alkyne cycloadditions, inverse electron demand Diels-Alder reaction, Staudinger
10 ligation, or oxime/hydrazine ligations.

For examples of methods of attaching dual-barcodes to peptides as supports, see Liszczak G, Muir TW. Nucleic Acid-Barcoding Technologies: Converting DNA Sequencing into a Broad-Spectrum Molecular Counter. *Angew Chem Int Ed Engl.* 2019;58(13):4144-4162. doi:10.1002/anie.201808956 and the references included therein which are incorporated
15 herein by reference.

Peptides that may be used as a support may be any protein or peptide. For example, the peptide may be an antibody, therapeutic peptides

In some examples, the support may be chemical agent. For example, the support may be a chemical agent that is a member of a chemical library. Attachment to chemical agents may
20 be via any of the functional group binding methods described above. In some examples, the chemical agent includes a functional group configured for attachment of corresponding functional group of an attachment region. In some examples, the chemical agent is a therapeutic molecule or molecule being tested as a therapeutic. For example, the chemical agent may be a compound or drug being tested or investigated for potential therapeutic
25 activity.

In some examples, the support may be a member of a DNA encoded library (DEL). DELs are collections of molecules, individually coupled to distinctive DNA tags. As such, the dual-barcodes of the invention may be configured to attach to a chemical moiety, peptide or other molecule that is a member of DEL and already includes a previously attached DNA tag (e.g.
30 previously attached barcode). In some examples, the attachment region may be configured to bind to the DNA component of a DEL member via methods described above. For example, the attachment region may be attached to a DNA component of a DEL member via restriction enzyme based attachment methods.

In some examples, the attachment region may be configured to allow for reversible
35 attachment of the dual-barcode to a support. For example, the attachment region may

include a nucleic acid sequence that includes a restriction enzyme recognition site that allows for a restriction enzyme to digest the attachment region. In some examples, the attachment region may include a functional group (and the corresponding functional group of the support) that is cleavable by an input. For example, cleavable upon application of light,
5 UV radiation or cleavable or cleaved in the presence of a specific chemical or by a chemical reaction.

UMI

In some examples, each dual-barcode includes a unique molecule identifier (UMI). A UMI is a nucleic acid sequence which identifies one particular dual-barcode. That is, the UMIs are
10 different for each dual-barcode.

Unique molecular identifiers (UMIs) are sequences of nucleotides applied to or identified in DNA molecules (i.e. a dual-barcode of the invention) that may be used to distinguish individual dual-barcodes from one another. Since UMIs are used to identify DNA molecules, they are also referred to as unique molecular identifiers. See, e.g., Kivioja, Nature Methods
15 9, 72-74 (2012). UMIs may be sequenced along with the DNA molecules with which they are associated to determine whether the read sequences are those of one source DNA molecule or another. The term "UMI" is used herein to refer to both the sequence information of a polynucleotide and the physical polynucleotide per se.

Commonly, multiple instances of a single source molecule (i.e. dual-barcode of the invention) are sequenced. In the case of sequencing by synthesis using Illumina's sequencing technology, the source molecule may be PCR amplified before delivery to a flow cell. Whether or not PCR amplified, the individual DNA molecules applied to flow cell are bridge amplified or ExAmp amplified to produce a cluster. Each molecule in a cluster derives from the same source DNA molecule but is separately sequenced. For error correction and
20 other purposes, it can be important to determine that all reads from a single cluster are identified as deriving from the same source molecule. UMIs allow this grouping.

UMIs are similar to barcodes, which are commonly used to distinguish reads of one sample from reads of other samples, but UMIs are instead used to distinguish one source DNA molecule from another when many DNA molecules are sequenced together. Because there
30 may be many more DNA molecules in a sample than samples in a sequencing run, there are typically many more distinct UMIs than distinct barcodes.

As mentioned, UMIs may be applied to or identified in individual DNA molecules. In some implementations, the UMIs may be applied to the DNA molecules by methods that physically link or bond the UMIs to the DNA molecules, e.g., by ligation or transposition through
35 polymerase, endonuclease, transposases, etc.

Physical UMIs may be defined in many ways. For example, they may be random, pseudo-random or partially random, or non-random nucleotide sequences that are incorporated in a dual-barcode of the invention. In some examples, the physical UMIs may be so unique that each of them is expected to uniquely identify any given dual-barcode present in a sample.

- 5 A UMI must have a sufficient length to ensure uniqueness for each and every dual-barcode. In some examples, a less UMI can be used in conjunction with other identification techniques to ensure that each source DNA molecule is uniquely identified during the sequencing process. In such examples, multiple dual-barcodes or fragments thereof may have the same physical UMI. Other information such as alignment location or virtual UMIs
- 10 may be combined with the physical UMI to uniquely identify reads as being derived from a single dual-barcode or fragment thereof. In some examples, dual-barcodes include physical UMIs limited to a relatively small number of non-random sequences, e.g., 120 non-random sequences. Such physical UMIs are also referred to as non-random UMIs. In some examples, the non-random UMIs may be combined with sequence position information,
- 15 sequence position, and/or virtual UMIs to identify reads attributable to a same dual-barcode. The identified reads may be combined to obtain a consensus sequence that reflects the sequence of the dual-barcode. Using physical UMIs, virtual UMIs, and/or alignment locations, one can identify reads having the same or related UMIs or locations, which identified reads can then be combined to obtain one or more consensus sequences. The
- 20 process for combining reads to obtain a consensus sequence is also referred to as “collapsing” reads.

- A “virtual unique molecular index” or “virtual UMI” is a unique sub-sequence in dual-barcode. In some examples, virtual UMIs are located at or near the ends of the dual-barcode. One or more such unique end positions may alone or in conjunction with other information uniquely
- 25 identify a dual-barcode. Depending on the number of distinct dual-barcodes and the number of nucleotides in the virtual UMI, one or more virtual UMIs can uniquely identify dual-barcodes in a sample.

- The use of UMIs has a number of advantages. For example, incorporating a UMI may allow the averaging out of sequencing results to account for cDNA molecules which are unevenly
- 30 amplified thus providing a more accurate quantification of gene expression and reduction in signal to noise.

- The UMI of each dual-barcode may be located at a position in the dual-barcode that allow for amplification of the UMI along with the tag binding regions of the dual-barcode. For example, a UMI may be located downstream of an application region as described herein. This may
- 35 allow for the UMI to amplified and sequenced in subsequent sequencing reactions.

In some examples, the UMI is located adjacent to a target region of a dual-barcode. For example, the dual-barcode may have the following structure (5' to 3'):

Attachment region – optional Amp – q(TBRs) – optional Amp – target region – optional Amp – UMI – Target region

5 or

optional Amp – q(TBRs) – optional Amp – target region – UMI – optional Amp – Attachment region

or

10 *Attachment region – optional Amp – q(TBRs) – optional Amp – target region – UMI – optional Amp*

or

optional Amp – q(TBRs) – optional Amp – UMI – target region – optional Amp – Attachment region

15 Examples of UMIs and methods of producing UMIs are well known in the art. For example, see US11447818B2, US8822150B2, and WO2016176091A1 which are incorporated herein by reference.

UMIs may be incorporated into a dual barcode by any suitable method of linking or combining nucleic acids, such as those described herein. For example, UMIs may be linked to a tag binding region, target region or other constituent region of the dual-barcode by 20 Golden Gate assembly methods, Gibson assembly or splint ligation (when ssDNA). In some examples, UMIs may be incorporated into the nucleic acid sequence of one or more of the constituents regions of the dual barcode. For example, at least one of the amplification regions, attachment region, linker region and/or target binding region may be designed to include a UMI nucleic acid sequence.

25 Target Region

The dual-barcodes of the invention may include a target region. A target region may not be required depending on the intended use of the dual-barcode. For example, when the dual barcode is intended for use in DEL production or in conjunction with DEL members, a target region may not be required.

30 A target region is used herein to refer to a region that is configured to bind to any desired target. In some examples, the target region is solely a nucleic acid sequence configured to bind to a target. In some examples, the target region may include additional binding molecules attached to a nucleic acid sequence that binds to a desired target. In some

examples, the target is a cellular target (e.g. a cell or molecule derived therefrom, such as, a gene, an mRNA, protein, organelle etc.).

In some examples, the target is a nucleic acid molecule. For example, the target may be an RNA or DNA molecule. In such cases, the target region may include a nucleic acid sequence
5 that is complementary to the sequence of the nucleic acid target. Thus allowing the target region to bind to the target via nucleic acid hybridization.

In some examples, the target is an RNA such as an mRNA. In some examples, the target region may be configured to bind to any mRNA. For example, the target region may include a plurality of thymine nucleotides – poly(T) sequence. A poly(T) sequence may bind to the
10 poly(A) tail of an mRNA. Thus, such a target region may bind all mRNAs within an entity.

In some examples, the target region may be configured to bind to a specific nucleic acid molecule, such as a specific gene or gene product (e.g. specific mRNA). In such cases, the target region may include a nucleic acid sequence that is at least partially complimentary to the specific gene or mRNA therefore allowing for specific binding to the target nucleic acid.

15 In some examples, the target comprises a cell. In such cases, the target region may include a nucleic acid sequence configured to bind a target presented on the surface of the cell or within the cell. For example, the target may be a surface presented cell marker or membrane protein of a cell. In some examples, the target region may be considered to bind to any target cell. For example, the target region may be configured to bind to a target that is
20 common to all cells in a sample (as such the dual-barcode may be used to detect the presence and/ quantity of cells in a sample). In other examples, the target region may be configured to bind to specific cells in a sample. For example, cells expressing or including a specific marker which the target region selectively and specifically binds to (this may allow detection of cells that express a particular marker and thus allow detection and/or
25 quantification of cells with a specific phenotype). Nucleic acid sequences that specifically bind to cell targets may be designed by methods such as Cell-SELEX.

In some examples, the target is a protein and the target region is configured to bind to a protein. In some examples, the target region is configured to bind to a specific protein. Methods of designing nucleic acid sequences that specifically bind to a proteins are well
30 known in the art. In some examples, the protein may be a nucleic acid binding protein. For example, a transcription factor or polymerase (such as a DNA or RNA polymerase). In such cases, the target region may include a nucleic acid sequence that includes the binding sequence of the nucleic acid binding protein.

In some examples, the target is a biomarker. The term “biomarker” refers to a molecule that
35 is associated either quantitatively or qualitatively with a biological change. Examples of

biomarkers include polypeptides, proteins or fragments of a polypeptide or protein; and polynucleotides, such as a gene product, RNA or RNA fragment; and other body metabolites. In certain embodiments, a "biomarker" means a compound that is differentially present (i.e., increased or decreased) in a biological sample from a subject or a group of subjects having a first phenotype (e.g., having a disease or condition) as compared to a biological sample from a subject or group of subjects having a second phenotype (e.g., not having the disease or condition or having a less severe version of the disease or condition).

In some examples, the target is a chemical moiety. In some examples, the chemical agent is a therapeutic molecule or molecule being tested as a therapeutic. For example, the target may be a chemical agent may be a compound or drug being tested or investigated for potential therapeutic activity. In some examples, the target region is configured for binding to a chemical agent. Methods of designing nucleic acid sequences that bind to specific chemical agents, for example, by binding to specific functional groups or structures are well known in the art.

In some examples, the target region comprises a nucleic acid sequence encoding a molecule of interest. For example, the target region comprises nucleic acid sequence encoding a gene of interest. In some examples, the target region may include one or more regulatory elements. "Regulatory elements" refer to sequences involved in controlling the expression of a nucleotide sequence. Regulatory elements comprise a promoter operably linked to the nucleotide sequence of interest and termination signals. They also typically encompass sequences required for proper translation of the nucleotide sequence.

Additional components and properties

In some examples, the dual-barcodes may include one or more markers. For example, the dual-barcode may include or encode one or more fluorescent markers. Such markers may be fluorescent molecules that differ from any fluorescent moieties that may be used for the identifiable moieties.

The markers may be located in one or more of the attachment region, at least one linker region, at least one NBR and or the target region. The inclusion of markers may help with quality control testing of the dual-barcodes.

Identifiable Tags

The dual-barcodes described herein may be bound by at least one identification tag to form a dual modality detectable barcode. The identification tags include at least one identifiable moiety and a tag sequence. The tag sequence being configured to bind to its cognate tag binding sequence.

The tag sequence is partially described above in relation to the tag binding sequence. For example, the tag sequence is the complement of at least one tag binding sequence of one tag binding region.

For example, each binding tag sequence is configured to only bind to its cognate tag binding sequence of a tag binding region. The tag sequence may include a nucleic acid sequence that is complimentary to the nucleic acid sequence of the tag binding sequence. Thus allowing the specific hybridization of each tag binding region to its cognate tag sequence.

The nucleic acid sequence of each tag sequence may be selected using the methods described herein. For example, each tag binding sequence or tag sequence may comprise a defined number of randomly selected nucleotides that have a Gibbs free energy (ΔG) of hybridization at a given temperature (i.e. ΔG_T) that is relatively low. For example, a Gibbs free energy for hybridization of at most -14000 J. Thus providing a tag binding sequence and complementary tag sequence that selectively and specifically bind to each other. For example, having minimised non-specific binding, minimised mis-annealing, and/or maximised complementary sequence binding.

The nucleic acid sequence of each tag sequence may also be configured to minimise self forming secondary structures, such as hairpins (i.e. a relatively high ΔG for secondary structure formation).

The nucleic acid sequence of each tag sequence may also be configured to minimise self-annealing (i.e. a relatively high ΔG for annealing to itself).

The nucleic acid sequence of each tag sequence may also be configured to minimise non-specific binding (i.e. a relatively high ΔG for annealing to other tag binding regions and/or non-cognate tag sequences, as well as other regions of the dual-barcode such as linker regions, attachment regions, and/or amplification region).

For example, a Gibbs free energy for self forming secondary structures, non-specific binding, homodimerization, self-annealing, mis-annealing, of at least -2000 J.

In some examples, each tag sequence comprise a GC content of between 30 to 50%.

In some examples, the each tag sequence may be at least 8 bp in length. For example, at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 bp or more. In some examples, each tag sequence is from 8 to 50 bp. In some examples, each tag sequence is from 8 to 40 bp. In some examples, each tag sequence is from 8 to 30 bp. In some examples, each tag sequence is from 8 to 20 bp. In some examples, each tag sequence is from 8 to 15 bp.

Methods for designing and selecting suitable tag sequences include:

providing one or more tag sequences of a defined length picking randomly between the nucleotides A, T, C and G wherein each the tag sequence comprises a GC content between 30-60%;

5 assessing each tag sequence for (a) self hybridisation, (b) hairpin formation, and (c) secondary structure formation using delta G as an indication of a likelihood of (a) to (c) occurring;

selecting the tag sequences with a likelihood of (a) to (c) occurring below a threshold;

10 assessing each tag sequence for binding to a complementary sequence using delta G as an indication of binding affinity;

selecting the tag sequences with a binding affinity to a complementary sequence greater than a threshold;

assessing each of the selected tag sequences for binding affinity to each other selected tag sequences using delta G as an indication of binding affinity;

15 selecting the tag sequences that have a binding affinity to other selected tag sequences lower than a threshold and a binding affinity to a complementary sequence greater than a threshold; and

20 assessing the tag sequences for binding to the complementary sequence of other tag sequences using delta G as an indication of binding affinity and selecting those tag sequences that specifically bind their complementary sequence and do not bind to the complementary sequence of other tag sequences.

In some examples, the method is a computer implanted method. Therefore, also provided herein is apparatus for designing a tag sequence for specifically binding to a tag binding sequence, the apparatus comprising:

25 communication circuitry configured to receive, from at least one sensor, data comprising information corresponding to a nucleic acid sequence;

a memory configured to store machine-readable instructions; and

processing circuitry configured to operably execute the machine-readable instructions to process the received signals to:

30 generate one or more tag sequences of a defined length picking randomly between the nucleotides A, T, C and G wherein each the tag sequence comprises a GC content between 30-60%;

assess each tag sequence for (a) self hybridisation, (b) hairpin formation, and (c) secondary structure formation using delta G as an indication of a likelihood of (a) to (c) occurring;

5 select the tag sequences with a likelihood of (a) to (c) occurring below a threshold;

assess each tag sequence for binding to a complementary sequence using delta G as an indication of binding affinity;

select the tag sequences with a binding affinity to a complementary sequence greater than a threshold;

10 assess each of the selected tag sequences for binding affinity to each other selected tag sequences using delta G as an indication of binding affinity;

select the tag sequences that have a binding affinity to other selected tag sequences lower than a threshold and a binding affinity to a complementary sequence greater than a threshold; and

15 assess the tag sequences for binding to the complementary sequence of other tag sequences using delta G as an indication of binding affinity and selecting those tag sequences that specifically bind their complementary sequence and do not bind to the complementary sequence of other tag sequences; and

20 control the communication circuitry to output one or more signals indicative of the selected tag sequences.

The apparatus may include communication circuitry, memory circuitry, processing circuitry, and a display. However, the apparatus is not limited to this and may be any suitable computing device. Furthermore, in some examples the apparatus may be in communication with another apparatus having a display, such that the designed tag sequences are output to
25 the display of the another apparatus so as to be remotely displayed.

The communication circuitry may be provided in one or more communication devices, and may be of any suitable type for communicating with the sensor to receive data signals indicative of the plurality of measurements. In the second embodiment, the communication circuitry includes a wireless module to receive the signals from the sensor wirelessly. For
30 example, the communication circuitry may include a Bluetooth®, Wi-Fi, WLAN and/or network data module. The communication circuitry may comprise a receiver in addition to a transmitter or may comprise a transceiver adapted to both transmit information to and receive information from the processing circuitry.

The processing circuitry may be any suitable processing means, and may be provided in one or more processors, such as an electronic processing device or computer processor. The processing circuitry is communicatively coupled to the communication circuitry and the memory. The memory may comprise any suitable memory means and is arranged to store machine-readable instructions. The processing circuitry is arranged to operably execute the machine-readable instructions stored on the memory. The processing circuitry comprises an input means and an output means. The input means may comprise an electrical input of the processing circuitry. The output means may comprise an electrical output of the processing circuitry. In some embodiments, the input means and output means may be unified such as in the form of a network interface which inputs and outputs data, for example to a communication bus of the communication circuitry. The processing circuitry may therefore receive data from the communication bus and output data onto the communication bus.

In some examples, the display may comprise any suitable display means and is in communication with the communication circuitry. As above, the processing circuitry is also in communication with the display and is arranged to control the display via the communication circuitry to output the signals indicative of the extracted measurements by displaying information corresponding to the designed tag sequences. In some examples, the extracted measurements may be displayed in the form of graphs, charts, tables, histograms or graphic depictions of the tag sequences.

In some examples, the processor is arranged to determine tag sequences using machine learning. For example, a neural network based machine learning model may be used. However, the disclosure is not limited to this, and in other examples of the disclosure, tag sequences may be determined using a model, such as a rules-based algorithm.

In some examples, the output data of the designed tag sequences may be stored on a database in a cloud networking environment. In such examples, the apparatus is in communication with the database which may be arranged remotely on a server in the cloud networking environment. The server includes communication circuitry and processing circuitry. However, the server is not limited to this and may be any suitable computing device. The communication circuitry includes a wireless module for communicating remotely with the apparatus by Wi-Fi and/or network data. However, the disclosure is not limited to this, and may include any suitable communication circuitry arranged to receive the extracted measurement data from the apparatus. The communication circuitry may comprise a receiver in addition to a transmitter or may comprise a transceiver adapted to both transmit information to and receive information from the processing circuitry. The apparatus is arranged to store information on the database and also retrieve information from the

database. However, it will be understood that the disclosure is not limited to this, and the database may in other examples be stored on the apparatus itself.

Also provided herein are machine readable instructions, which when executed by processing circuitry, cause the processing circuitry to carry out a method of designing a tag sequence
5 for specifically binding to a tag binding sequence as described herein.

Also provided herein is a machine readable data storage medium having tangibly stored thereon the machine readable instructions of designing a tag sequence for specifically binding to a tag binding sequence as described herein.

The method above, designs tag sequences that only bind to their complementary sequence
10 (i.e. complementary tag binding sequence) and do not bind to other identical tag sequences, other tag sequences of other identifiable tags, and/or tag binding sequences that are not the tag sequence's cognate tag binding sequence. In addition, the method above, designs tag sequences that do not form secondary structures.

The selected tag sequences may also be assessed for binding to other constituents parts of
15 the dual barcodes. For example, each tag sequence may be assessed to check whether the tag sequence is likely to or will bind to the attachment, linker region, UMI, amplification region, and/or target region using delta G as an indication of binding affinity. Depending on how the dual-barcodes are produced, binding to certain regions may not be taken into account. For example, when the dual-barcodes are constructed sequentially there may be
20 no UMI or target region present when the identifiable tags are bound.

Alternative methods, will be known by those skilled in the art, for example, see Matveeva, O V et al. "Thermodynamic calculations and statistical correlations for oligo-probes design." Nucleic acids research vol. 31,14 (2003): 4211-7. doi:10.1093/nar/gkg476 and Breslauer, K.J.; Frank, R; Blöcker, H; Marky, LA; et al. (1986). "Predicting DNA Duplex Stability from
25 the Base Sequence". Proc. Natl. Acad. Sci. USA. 83 (11): 3746–3750.
Bibcode:1986PNAS...83.3746B which are incorporated herein by reference.

In some examples, each tag sequence comprise a GC content of between 30 to 50%.

In some examples, each tag sequence includes one or more locked nucleic acids. Locked nucleic acid (LNA) is the term for oligonucleotides that contain one or more nucleotide
30 building blocks in which an extra methylene bridge fixes the ribose moiety either in the C3'-endo (beta-D-LNA) or C2'-endo (alpha-L-LNA) conformation. The use of one or more locked nucleic acids may provide more stable tag sequences and identifiable tags. The use of locked nucleic acids may also allow for smaller sized tags to be used.

The tag sequences may be ssDNA or dsDNA.

In some examples, each tag sequence in a single set of identification tags the same nucleic tag sequence. In some examples, each tag sequence in a set of identification tags comprise a distinct nucleic acid sequence for binding to its cognate tag binding sequence of tag binding region.

- 5 In some examples, a single nucleic acid molecule encoding all tag sequences (and any NBRs) for all tag binding sequences in a single tag binding region may be provided. In some examples, a single nucleic acid molecule encoding all tag sequences (and any NBRs) for all tag binding sequences of all tag binding regions may be provided. Such nucleic acid molecules may allow for binding of all identification tags of a tag binding region or all tag
10 binding regions in one single round of hybridization.

In other examples, identification tags may each be discrete molecules. For example, an identification tag may be a discrete entity on a single stranded DNA and therefore multiple tags may be needed for multiple binding sites. The identification tags may be double stranded and undergo double stranded hybridisation.

- 15 Each identifiable tag includes at least one identifiable moiety. An identifiable moiety refers to a molecule or material that can produce a detectable (such as visually, electronically or otherwise) signal that indicates the presence (i.e. qualitative analysis) and/or concentration (i.e. quantitative analysis) of the moiety in a sample.

- An "identifiable moiety" is any moiety that may be detected and/or identified by
20 spectroscopic, photochemical, biochemical, immunochemical, chemical and/or other physical means. The terms "identifiable" and "detectable" may be used interchangeably. An identifiable moiety may be coupled either directly and/or indirectly (for example via a linkage, such as, without limitation, a DOTA or NHS linkage) to a tag sequence using methods well known in the art. A wide variety of identifiable moieties may be used, with the choice
25 depending on the sensitivity required, ease of conjugation, stability requirements and available instrumentation. Suitable identifiable moieties include, but are not limited to, fluorescent labels, a radioactive labels (for example, without limitation, ^{125}I , In^{111} , Tc^{99} , I^{131} and including positron emitting isotopes for PET scanner etc), nuclear magnetic resonance active labels, luminescent labels, chemiluminescent labels, chromophore labels, enzyme
30 labels (for example and without limitation horseradish peroxidase, alkaline phosphatase, etc.), quantum dots and/or nanoparticles. Identifiable moieties may cause and/or produce a detectable signal thereby allowing for a signal from the identifiable moiety to be detected and allowing the dual-barcode and ergo entity to be identified. In some examples, identifiable moieties may be moieties configured for detection by any one or more of fluorescence,
35 absorbance, mass spectrometry, Raman spectroscopy and/or high-resolution imaging.

As used herein in reference to monitoring, measurements, detection, identification or observation of dual-barcodes and entities including such dual-barcodes, the term “real-time” refers to measurements performed contemporaneously with the monitored, measured, or observed events, as opposed to measurements taken after an event has occurred. Thus, “real time” detection or identification contains not only the measured and quantitated result, but expresses this at various time points, that is, in hours, minutes, seconds, milliseconds, nanoseconds, etc. “Real time” includes detection of signals, comprising taking a plurality of readings in order to characterize the signal over a period of time.

A signal from the identifiable moiety is detected and/or measured either substantially concurrently with the presence and/or occurrence of an event, response, condition, or agent of interest or after a short lag time. A lag or delay between the timing of an event, response, condition, or agent of interest and the detectable signal is, for example, the result of the time required for the cellular and/or assay steps necessary for detection. Preferably such lag or delay is less than 30 minutes, e.g., 25 minutes, 20 minutes, 15 minutes, 10 minutes, 5 minutes, 4 minutes, 3 minutes, 2 minutes, 1 minute or less, including no lag or delay, or any time period that is short (e.g., <10%, <5%, <2%, <1%) of the time period being monitored. Due to consistent lag times over the time course of an assay, a delayed signal may still be suitable to provide real-time readout of the timing of an event, response, condition or agent of interest.

An identifiable moiety may be positioned at any suitable location of the tag sequence such as, for example, at the 3' end of the tag sequence, at the 5' end of the tag sequence, at a location between the 3' and 5' ends of the tag sequence, on a strand of a double-stranded tag sequence that does not contain a quencher, or at an opposite end of a tag sequence from a quencher.

In some examples, each identification tag includes 1, 2, 3, 4, or more identical identifiable moieties which may each be attached at different locations along the tag sequence. The use of multiple identifiable moieties in each identifiable tag may provide increased signal intensity and therefore improve detection and identification of entities.

An identifiable moiety may be coupled to a tag sequence directly or indirectly via a linker.

Coupling of an identifiable moiety may be achieved via any suitable route including covalent and/or non-covalent routes using methods as described elsewhere herein.

Examples of identifiable and detectable moieties include chromogenic, optical, fluorescent, chemiluminescent, magnetic, plasmonic, mass-based, electrochemical labels, and phenolic substrates.

In some examples, the identifiable moiety is an optical moiety. Optically detectable moieties, may be any moiety that may be detected by optical detection methods. For example absorbance, refraction, emission, fluorescence, luminescence, and/or scattering detection methods. Optically detectable moieties, include luminescent, chemiluminescent, fluorescent, 5 fluorogenic, chromophoric and/or chromogenic moieties. Some examples of suitable optically detectable moieties include fluorescein labels, rhodamine labels, cyanine labels (e.g., Cy3, Cy5, and the like), and the ALEXA® family of fluorescent dyes and other fluorescent and fluorogenic dyes.

Examples of optical moieties such as fluorescent moieties include organic dyes, biological 10 fluorophores, quantum dots, and nanoparticles including carbon dots. In some examples, the identifiable moiety may be a fluorophore. For example, a fluorescent protein. The term "fluorescent protein" refers to a protein that possesses the ability to fluorescence (i.e., to absorb energy at one wavelength and emit it at another wavelength). Examples of fluorescent proteins include GFP, eGFP, eYFP, eYFP Emerald, mApple, mPlum, mCherry, 15 tdTomato, palm-GRET, mStrawberry, J-Red, DsRed-monomer, mOrange, MKO, mCitrine, Venus, YPet, CyPet, mCFPm, Cerulean and T-Sapphire, mKOK, mUKG, Clover, mKate, mKate2, tagRFP, tagGFP, mNEON green, iRFP720, iRFP670 and synthetic non-Aequorea fluorescent proteins. Other fluorescent proteins will be known to those skilled in the art. For example, fluorescent proteins and their respective properties may be obtained from 20 databases such as FPBase (<https://www.fpbases.org/>). Further examples of fluorescent moieties and their selection can be found for example in "Shaner, Nathan C., Paul A. Steinbach, and Roger Y. Tsien. "A guide to choosing fluorescent proteins." *Nature methods* 2.12 (2005): 905-909" which is incorporated herein by reference. In some examples, the reporter is selected from eYFP, mKate2, eGFP, palm-GRET, tdTomato, iRFP720, and/or 25 iRFP670. In some examples, the reporter is selected from eYFP, mKate2, eGFP, tdTomato, iRFP720, and/or iRFP670.

Exemplary chromogenic labels include diaminobenzidine (DAB), nitro blue tetrazolium chloride (NBT), 5-bromo-4-chloro-3-indolyl phosphate (BCIP), and 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal). A label can be a moiety that operates through 30 scattering, either elastic or inelastic scattering, such as nanoparticles and Surface Enhanced Raman Spectroscopy (SERS) reporters (e.g., 4-Mercaptobenzoic acid, 2,7-mercapto-4-methylcoumarin). A label can also be a chemiluminescence/electrochemiluminescence emitter such as ruthenium complexes and luciferases.

In some examples, the identifiable moiety is a defined mass moiety. A defined mass moiety 35 refers to any moiety that has a known and characterised mass. Thus allowing dual-barcodes

to be distinguished from each other by the total mass of defined mass moieties attached thereto. Examples of defined mass moieties include metal ions, proteins,

Support

As described above, the support to which a dual-barcode may be attached may be a cell,
 5 cellular component, peptide, nucleic acid, microparticle, nanoparticle, or chemical moiety or agent.

In some examples, the support is solid support such as a microparticle or nanoparticle. For example, the support may be a bead, quantum dot or janus particle that may be configured to attachment to an attachment region of the dual-barcode.

- 10 In some examples, the support is a bead. The term “bead,” as used herein, generally refers to a particle. The bead may be a solid or semi-solid particle. The bead may be a gel bead. The gel bead may include a polymer matrix (e.g., matrix formed by polymerization or cross-linking). The polymer matrix may include one or more polymers (e.g., polymers having different functional groups or repeat units). Cross-linking can be via covalent, ionic, or
 15 inductive, interactions, or physical entanglement. The bead may be a macromolecule. The bead may be formed of nucleic acid molecules bound together. The bead may be formed via covalent or non-covalent assembly of molecules (e.g., macromolecules), such as monomers or polymers. Such polymers or monomers may be natural or synthetic. Such polymers or monomers may be or include, for example, nucleic acid molecules (e.g., DNA or RNA). The
 20 bead may be formed of a polymeric material. The bead may be magnetic or non-magnetic. The bead may be rigid. The bead may be flexible and/or compressible. The bead may be disruptable or dissolvable. The bead may be a solid particle (e.g., a metal-based particle including but not limited to iron oxide, gold or silver) covered with a coating comprising one or more polymers. Such coating may be disruptable or dissolvable.
- 25 In some examples, the support comprises or is a polyacrylamide bead. For example, the bead may comprise a matrix of cross-linked polyacrylamide (e.g., cross-linked linear polyacrylamide). The cross-linked polyacrylamide (e.g., cross-linked linear polyacrylamide) may be cross-linked by a reducible cross-linkage, such a disulfide linkage. Examples of polyacrylamide beads include those described in e.g., U.S. Patent Publication No.
 30 2014/0378345 and U.S. Provisional Patent Application No. 62/163,238.

The support may include at least one attachment moiety or group for attachment to the dual-barcode. For example, the attachment moiety may be an attachment group or attachment region binding partner as described above in relation to the attachment. For example, the attachment moiety may be selected depending what support is being used.

The support may be a polyacrylamide bead that includes a nucleic acid molecule such a dsDNA molecule incorporated therein. The nucleic acid may include an acryl linker to allow for co-polymerisation of the nucleic acid into the bead. or example, the nucleic acid may be copolymerised with the polyacrylamide. In some examples, the bead comprises a dsDNA molecule that includes a sequence that is configured for attachment to the attachment region of a dual-barcode. For example, the bead may include a DNA molecule that includes a restriction enzyme recognition site that is configured for attachment to the attachment region of a dual-barcode. For example, a nucleic acid sequence that is compatible with Golden Gate assembly. For example, the bead is co-polymerized with a DNA molecule that includes a IIS restriction enzyme recognition site to allow for Golden Gate-mediated attachment of the attachment region to the bead.

A bead is an efficient way to incorporate barcodes in an entity where the bead is a support with barcodes attached. The bead may be any shape. The bead may be a non-dissolvable bead or a dissolvable bead.

Where the bead is non-dissolvable, the dual-barcodes may be attached to the bead via an attachment region which is cleavable to remove the barcode from the bead. For example, the attachment may be cleavable with UV.

In some examples, the support may be comprised within the cell and/or on the cell surface. In some examples, the support may comprise a DEL member.

Each support may have at least one dual-barcode attached. In some examples, the support has a plurality of dual-barcodes attached thereto. The number of dual-barcodes attached to a support may be determined by the number of available attachment region binding groups provided on or in a support. In some examples, each support may have at least 100, 200, 300, 4400, 500, 1000, 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , or more dual barcodes attached thereto.

In some examples, there is provided a support that is attached to at least one dual barcode as described herein via one or more attachment moieties of the support. In some examples, the support may have a plurality of identical dual-barcodes attached thereto. In some examples, the support may have a plurality of dual-barcodes attached thereto where a portion of the dual barcodes include distinct target regions and optionally identical tag binding regions. In some examples, the support may have a plurality of distinct dual-barcodes attached thereto.

In some examples, the support may be a support as described herein, for example, may be a bead, a cell, a protein, a nucleic acid or a chemical agent. In some examples, the support is one member of a library or plurality of supports (e.g. one member of a library of beads,

cells, proteins, nucleic acids or a chemical agents) and each support is a distinct member of the plurality.

In some examples, each dual-barcode is attached via a linker region (attachment region) to the support via an attachment moiety. For example, the attachment moiety is a nucleic acid
5 including a sequence configure for use with an assembly method described herein (e.g. Golden Gate assembly). In some examples, the attachment moiety is at least partially integral to the support. For example, in the case of a bead such as polyacrylamide bead the attachment moiety may be a nucleic acid including an acryl-linker that has been copolymerised with the bead.

10 Each dual-barcode attached to a support may include a distinct UMI.

In some examples, there is provided a plurality of supports each bound to at least one dual-barcode (i.e. a plurality of barcoded supports). Each support may have a plurality of dual-barcodes with identical tag binding regions and/or identical target regions attached thereto. As such, each support may have a dual-barcode having the same number and combination
15 of identification tags and ergo the same identification signature and sequence signature. Each support of the plurality may have a different and distinct dual-barcode attached thereto. For example, each support has a purity of dual-barcodes attached thereto which are distinct to the dual-barcodes attached to other supports of the plurality. Thus there may be provided a library of barcoded supports where each support has a different dual-barcode attached
20 (plurality of distinct barcoded supports).

In some examples, such as when the support comprises a member of a library of supports, each support may be distinct. For example, there may be provided a barcoded library of distinct supports.

In some examples, the dual-barcodes of the invention may not be attached to a support but
25 used free in solution.

Entities

An entity is used to refer interchangeably to any suitable object or article which may be include a support attached to the dual-barcodes described herein or a support to which the dual-barcodes are attached. In general an entity may be any object that may be wished to be
30 identified and/or detected.

For example, the entity may be a droplet, a bead, cell, cell component, a peptide, nucleic acid, chemical agent or any other objects that may wish to be investigated.

In some examples, the entity is a droplet comprising a bead cell, cell component, a peptide, nucleic acid, chemical agent or any other objects that are attached to dual-barcodes of the invention. In some examples, the entity is a microfluidic droplet.

By microfluidic droplet is meant a discrete volume of a first liquid in an immiscible second liquid. In some examples, the droplet may be considered a reaction vessel that includes agents for carrying out a reaction. In such cases, the barcoded support (a support attached to dual-barcodes of the invention) may be used to identify each droplet and the contents thereof as well as tracking each droplet throughout a reaction or assay in real time.

In some examples, the entity may be a droplet that includes reaction reagents that may react with each other in the droplet or react with the support (for example a chemical agent support).

The droplet may include a target of interest which corresponds to a target region of the barcoded support. For example, the target of interest may be a cellular target and the droplet may include lysate from a cell or an intact cell. In some examples, each entity may include lysate from cells with different properties (e.g. phenotype or genotype) or suspected of having different properties. For example, cells that have been exposed to different agents (such as therapeutic agents) or conditions. In some examples, the target region may include a sequence configured to bind to a specific mRNA from a cell or to a plurality of mRNAs produced by a cell. In some examples, the target region may target a specific protein. In some examples, the target region may target a specific metabolite or other cellular component.

Droplets may be generated using any suitable method such as using microfluidic droplet generations methods and systems well known in the art.

In some examples, a barcoded support may be held in one droplet and a droplet including a target or other reagents may be combined, for example, fused with the droplet including the barcoded support.

In some examples, droplets may be formed by injecting a barcoded support. For example, by picoinjection or other injection methods known in the art.

In some examples, the entity may be a cell in a fluid. The cell may be attached to the barcode and act as a support or may be bound by a barcoded support via a target region for binding to a target of the cell.

In some examples, the target region may encode a gene or protein of interest and the entity includes components for *in vitro* transcription. For example, the entity may be a droplet that includes ribonucleotide triphosphates (rNTPs); an RNA polymerase; a buffer solution; a

pyrophosphatase; an RNAase inhibitor; a chaotropic agent; a polyamine; and a source of magnesium ions and optionally a capping analogue. The target region acts a template strand for production of the RNA.

- In some examples, the target region may encode a gene or protein of interest and the entity includes components for *in vitro* transcription and translation. For example, the entity includes components for *in vitro* transcription and components for *in vitro* translation. For example, a suitable buffer, an *in vitro* transcription/replication system and/or an *in vitro* translation system containing all the necessary ingredients, enzymes and cofactors, RNA polymerase, nucleotides, transfer RNAs, ribosomes and amino acids (natural or synthetic).
- In some examples, the entity may include or be a cell and the reaction may be a cell lysis. The entity may then be exposed to conditions to promote binding of a target from the cell lysate to a target region. Exemplary *in vitro* translation systems can include a cell extract, typically from bacteria (Zubay, 1973; Zubay, 1980; Lesley et al., 1991; Lesley, 1995), rabbit reticulocytes (Pelham and Jackson, 1976), or wheat germ (Anderson et al., 1983). Many suitable systems are commercially available (for example from Promega) including some which will allow coupled transcription/translation (all the bacterial systems and the reticulocyte and wheat germ TNT.TM. extract systems from Promega). The mixture of amino acids used may include synthetic amino acids if desired, to increase the possible number or variety of proteins produced in the library. This can be accomplished by charging tRNAs with artificial amino acids and using these tRNAs for the *in vitro* translation of the proteins to be selected (Elman et al., 1991; Benner, 1994; Mendel et al., 1995).

Methods Of Producing

- Also provided herein are methods of producing a plurality of dual-barcodes as described herein. Each one of the plurality of barcodes may have a distinct nucleic acid sequence and a distinct number of each type of distinct tag binding regions (and ergo distinct number of each identification tag bound thereto in use). For example, each dual-barcode may include three distinct tag binding regions. A first dual-barcode of the plurality may have a first tag binding region (TBRa) having one tag binding sequence, a second tag binding region (TBRb) with one tag binding sequence and a third tag binding region (TBRc) with one tag binding sequence. A second dual-barcode of the plurality may have a first tag binding region (TBRa) having two tag binding sequences, a second tag binding region (TBRb) with one tag binding sequence and a third tag binding region (TBRc) with one tag binding sequence. A third dual-barcode of the plurality may have a first tag binding region (TBRa) having three tag binding sequences, a second tag binding region (TBRb) with one tag binding sequence and a third tag binding region (TBRc) with one tag binding sequence. A fourth dual-barcode

of the plurality may have a first tag binding region (TB Ra) having three tag binding sequence, a second tag binding region (TB Rb) with two tag binding sequence and a third tag binding region (TB Rc) with one tag binding sequence.

5 In some examples, the methods may produce one or more of each distinct member of the plurality of distinct dual-barcodes. For example, the methods may produce a plurality of first barcodes, a plurality of second barcodes, and a plurality of third barcodes each of which are distinct from each other by sequence and number of each type of distinct tag binding regions.

10 The method of producing a barcoded support may include providing a plurality of first type tag binding regions each having a distinct number of tag binding sequences. Each tag binding sequence for binding to identifiable tags that include the same identifiable moiety (first identifiable tag). For example, a first one of the first type of tag binding regions may have 1 tag binding sequence (TB Ra1), a second one of the first type tag binding regions may have 2 tag binding sequence (TB Ra2), a third one of the first type tag binding regions may have 3 tag binding sequence (TB Ra3). It will be understood that each tag binding region of the first type may have from 1 to any number greater than 1 but the number of tag binding sequence is not repeated for any of the first type of tag binding sequence. Each first tag binding region may be produced by any suitable nucleic acid synthesis method.

20 Then each first tag binding region is attached to a single support in different vessels. For example, TB Ra1 is bound to a support in a first vessel, TB Ra2 is attached to support in a second vessel and TB Ra3 is attached to a support in a third vessel and so on. Thus providing a plurality of supports each having at least one or more of the first tag binding region attached thereto where each support has a tag binding region with a distinct number of tag binding sequences.

25 Attachment to the support may include attaching via an attachment region and attachment moiety on the support. For example, the support may include a nucleic acid sequence that is configured to bind to an attachment region of the first tag binding region. As described above, in some examples, the attachment region may be a region that is configured for Golden Gate assembly. For example, each first tag binding region may be bound to an individual support by carrying out a Golden Gate assembly protocol. For example, digesting 30 the attachment region and the attachment moiety with a type IIS restriction enzyme and ligating the compatible end of the attachment moiety (of the support) and the attachment region (of each first target binding region) together.

35 As such, each support has a first binding region that differs from other supports by the number of tag binding sequences (i.e. each support comprises $L \times x$ tag binding sequences

wherein L is the number of tag binding sequences and x is an integer of 1 or more that is not repeated).

Each of the supports including each first tag binding region may then be combined and a mixture of all the supports including tag binding regions with different numbers so tag binding
5 sequence is deposited into a plurality of vessels.

A second type of tag binding region is added to each vessel where each second tag binding region has a distinct number of tag binding sequences for a binding to an identification tag having the same identification moiety but a different identification moiety (i.e. second identification tag) to the first identification tag. For example, a first one of the second type of
10 tag binding regions may have 1 binding sequence (TBRb1), a second one of the second type of tag binding regions may have 2 tag binding sequence (TBRa2), a third one of the second type of tag binding sequences may have 3 tag binding sequence (TBRa3).

In each vessel, each second type of tag binding region is linked to the first tag binding region. For example, by use of a Golden Gate assembly protocol or other suitable nucleic acid linking method. Thus providing each support with a first type of tag binding region with a
15 distinct number of tag binding sequences and a second type of tag binding region with a distinct number of tag binding sequences in comparison to other supports. For example providing (each in a separate vessel);

Support – TBRa1 – TBRb1
20 Support – TBRa1 – TBRb2
Support – TBRa1 – TBRb3
Support – TBRa2 – TBRb1
Support – TBRa2 – TBRb2
Support – TBRa2 – TBRb3
25 Support – TBRa3 – TBRb1
Support – TBRa3 – TBRb2
Support – TBRa3 – TBRb3

The support including a first binding region and second binding region may then be mixed
30 together and then the mixture placed into separate vessels. In each separate vessel the support including the first and second type of binding regions each with distinct numbers of tag binding sequences are mixed with a third type of tag binding region is where each third

tag binding region has a distinct number of tag binding sequences for a binding to an identification tag having the same identification moiety but a different identification moiety (i.e. second identification tag) to the first and second identification tags. For example, a first one of the third type of tag binding regions may have 1 binding sequence (TBRc1), a second one of the third type of tag binding regions may have 2 tag binding sequence (TBRc2), a third one of the third type of tag binding sequences may have 3 tag binding sequence (TBRc3). The third type of tag binding region may then be linked to the second tag binding region as described herein. For example, using Golden Gate assembly methods or any other suitable linking method. Thus providing (each in a separate vessel):

- 10 Support – TBRa1 – TBRb1 - TBRc1
- Support – TBRa1 – TBRb1 – TBRc2
- Support – TBRa1 – TBRb1 – TBRc3
- Support – TBRa1 – TBRb2 – TBRc1
- Support – TBRa1 – TBRb2 – TBRc2
- 15 Support – TBRa1 – TBRb2 – TBRc3
- Support – TBRa1 – TBRb3 – TBRc1
- Support – TBRa1 – TBRb3 – TBRc2
- Support – TBRa1 – TBRb3 – TBRc3
- Support – TBRa2 – TBRb1 – TBRc1
- 20 Support – TBRa2 – TBRb1 – TBRc2
- Support – TBRa2 – TBRb1 – TBRc3
- Support – TBRa2 – TBRb2 – TBRc1
- Support – TBRa2 – TBRb2 – TBRc2
- Support – TBRa2 – TBRb2 – TBRc3
- 25 Support – TBRa2 – TBRb3 – TBRc1
- Support – TBRa2 – TBRb3 – TBRc2
- Support – TBRa2 – TBRb3 – TBRc3
- Support – TBRa3 – TBRb1 – TBRc1
- Support – TBRa3 – TBRb1 – TBRc2
- 30 Support – TBRa3 – TBRb1 – TBRc3

Support – TBRa3 – TBRb2 – TBRc1

Support – TBRa3 – TBRb2 – TBRc2

Support – TBRa3 – TBRb2 – TBRc3

Support – TBRa3 – TBRb3 – TBRc1

5 Support – TBRa3 – TBRb3 – TBRc2

Support – TBRa3 – TBRb3 – TBRc3

The method may be carried out with any number of types of binding region until a desired number of distinct or unique dual-barcodes have been produced.

10 Alternatively, the method may include attaching to a plurality of supports a plurality of first type binding regions all having the same number of tag binding sequences for binding an identification tag with a first identifiable moiety.

Then linking a second type of target binding regions all having the same number of tag binding sequences for binding an identification tag with a second identifiable moiety or each of the first type of tag binding regions. For example, if each of the first and second types of tag binding regions have 1 tag binding sequences, providing:

15

Support – TBRa1 – TBRb1

For example, the first type of tag binding region may have L tag binding sequences and the second type of tag binding region may have L_b tag binding sequences wherein L and L_b are any integer.

20 Then at least one further type of tag binding region may be added. For example, a third, fourth and fifth type of tag binding region may be linked sequential to the first and second type of tag binding regions. For example, TBR_y where y is an identity of the type of tag binding region. The further tag binding region may have L_y tag binding sequences where y is any integer of 1 or more when the further tag binding region is a type of tag binding region that has not already been used in the production of the dual-barcode.

25

In some examples, a further tag binding region may be a type of tag binding region that has already been used in the dual-barcode. In such cases, the number of tag binding sequences is L_y+x where x is an integer of 1 or more and is not repeated. For example, the further tag binding region may be a first or second type of tag binding region that has a number of tag binding regions different to any of the other first or second type of tag binding regions already attached used in the dual-barcode.

30

For example, a further first type of tag binding region may have L+x tag binding sequences and a further second type of tag binding region may have L_b+x tag binding sequences and a

further tag binding sequence may have $L_y + x$ tag binding sequences (when the same L_y has been already used in the dual-barcode) wherein L , L_b , L_y are any integer and X is any integer of 1 or more and is not repeated.

For example, linking a further first tag binding region that includes 2 tag binding sequences
5 would provide:

Support – TBRa1 – TBRb1 – TBRa2

This may be repeated as many times as required in order to produce the desired number of distinct dual-barcodes. Thus, a plurality of supports attached to distinct dual barcodes may be produced. As each support is attached to at least one distinct dual-barcode a library of
10 distinct barcoded supports may be provided.

In some examples, the methods may include a wash step between each round of attaching or linking tag binding regions in order to remove any unbound tag binding regions. In some examples, the wash buffer comprises Tris 50 mM, Tween-20 (0.1%).

In some examples, placing partially barcoded supports or barcoded supports in individual
15 vessels may include placing the same amount of each partially barcoded support or barcoded support in each vessel.

In some examples, the design of each distinct dual-barcode may be carried out *in silico*. For example, the design of each distinct barcode is a carried out as a computer implemented method. In such cases, the method may then include a step of producing each one of the
20 distinct dual barcodes. For example, using nucleic acid synthesis methods. In some examples, a first one each barcode may be produced by *de novo* nucleic acid synthesis and subsequent copies of each barcode may be produced by amplification methods such as PCR. Thus a library of dual-barcodes may be provided, where each dual barcode is distinct.

In computer-implemented methods disclosed herein, an algorithm can be used that for
25 example, to detect the beads by calculating the mean intensity of each detected bead. The algorithm may then use such information to recover the underlying DNA sequence that was used to produce the barcode. A separate fluorescence readout can be acquired from the contents of the bead's corresponding droplet. The specific output of the algorithm depends on the task, but in every case is a pair of the barcode's DNA sequence and the task-
30 dependent readout corresponding to a single droplet, acquired at the same time as the readout of the barcode: {DNA, task readout}. This output can (but does not necessarily have to) be produced at the same time images are acquired, so the images themselves can be discarded. Such method removes any potential bottleneck on the ability to store the vast amount of data that can be collected with this method.

For example, a quantitative measure of the deterioration of a specific organelle in the cell inside the droplet could be calculated. The output of the algorithm would then be a pair of the bead's DNA sequence and the value of the deterioration metric: {DNA, deterioration}. An example a metric of the deterioration of the Golgi apparatus – its angle around the nucleus – is reported by Scherer et al., in Sci Adv. 2022 Jan 7;8(1):eabl4895. doi:10.1126/sciadv.abl4895.

Alternatively, if the droplets are used for the evaluation of reaction extent, and the reaction extent is being tracked with a fluorescent probe (such as reported by Markin *et al.*, Science . 2021 Jul 23;373(6553):eabf8761. doi: 10.1126/science.abf8761), the output becomes a pair of the barcode's DNA sequence and the mean fluorescence intensity inside the droplet excluding the bead: {DNA, droplet fluorescence}

Another possible type of output only involves processing the barcode, not the rest of the contents of the droplet. The analysis of the latter can be reserved for downstream analysis at a later point, for example, in computation-intensive tasks. For example, droplets can contain cells, in which case the output would be a pair of the barcode's DNA sequence and the image of the cell, as our system can be used for imaging flow cytometry

At least one of each of the plurality of distinct dual-barcodes may be attached to a support using methods as described herein, such as Golden Gate assembly. Thus providing a library of distinct barcoded supports.

In some examples, each support may be different chemical agent or peptide and the method may produce a DNA (i.e. dual barcode) encoded library (e.g. a DNA encoded chemical library or DNA encoded peptide library). As such, there is also provided herein a DNA encoded library comprising a plurality of library members (e.g. chemical agents or peptides) encoded with at least one distinct dual-barcode as described herein.

In some examples, the methods may include linking (or designing *in silico*) one or more of at least one UMI and/or target binding region to the final type of target binding region (i.e. the tag binding region most distal to the attachment region of support). The UMI may be distinct for each one of the dual barcodes attached to a single support. That is to say that every single dual-barcode produced may include a distinct UMI so that each dual barcode that is attached to a single support (or each of a plurality of the same dual-barcodes) each include a distinct UMI.

In some examples, the target region may be a distinct target region for each of the plurality of the distinct barcodes or barcoded supports. That is to say, a plurality of the same dual barcodes that may be attached to a single support each include the same target region.

Thus allowing each barcoded support to target a distinct target. In some examples, the target region may encode a gene of interest or protein of interest as disclosed herein and each of the plurality of barcoded supports or plurality of dual-barcodes may be used to produce a plurality of unique proteins.

- 5 After production of plurality of distinct barcoded supports or a plurality of distinct barcodes, a plurality of identification tags having a number of distinct identification moieties corresponding to the total number of types of target binding regions may be bound to each target binding sequence via its cognate tag sequence.

- Binding of the target sequence may be carried out by any suitable method of annealing
10 complementary nucleic acid sequences. For example, when the dual-barcodes are formed from dsDNA, the plurality of distinct barcoded supports or a plurality of distinct barcodes may be heated to a temperature to that melts the dsDNA therefore unbinding a complementary strand of the dual-barcode and then the dual-barcodes may be cooled to allow hybridization of the tag sequences to their cognate tag binding sequences. For example, the plurality of
15 distinct barcoded supports or a plurality of distinct barcodes may be heated to a temperature of about 95°C and then cooled in the presence of the identification tags and reheated. This may be repeated in cycles to ensure selective and specific binding or identification tags to their corresponding tag sequences. Methods of annealing nucleic acid sequences are well known in the art. For example, the methods described herein may include steps based on
20 methods such as florescence *in situ* hybridization (FISH) techniques.

- In some examples, binding includes washing the plurality of dual-barcodes or plurality of barcoded supports in a denaturation buffer. In some examples, the denaturation buffer may include NaOH, Brij™ 35P, and Tween-20. In some examples, the denaturation buffer comprises NaOH 150 mM, Brij 35P (0.5%), Tween-20 0.05 %. After washing with
25 denaturation buffer binding of the identifiable tags may include washing with a hybridization buffer before incubating with the identifiable tags. In some examples, the hybridization buffer comprises Tris, KCl, EDTA, and Tween 20. In some examples, the hybridization buffer comprises Tris 5 mM, KCl 100 mM, EDTA 5 mM, Tween-20 (0.05 %).

- After washing with hybridization buffer the plurality of dual-barcodes or plurality of barcoded
30 supports may be incubated with the identification tags in the hybridization buffer. Incubating may include heating the plurality of dual-barcodes or plurality of barcoded supports for a period of time. For example, heating to about 95°C for 10 minutes. After heating, the plurality of dual-barcodes or plurality of barcoded supports and identification tags may be cooled over a period of time. For example, cooled to about 25°C over a time period of 2 hours.

After hybridization, the plurality of dual-barcodes or plurality of barcoded supports may then be washed using a wash buffer as described herein to remove any identification tags that are not bound.

Thus providing a plurality of distinct barcoded supports or a plurality of distinct barcodes
5 each bound to a unique combination of distinct identification tags.

The combination of types and number of identifiable tags provides a dual-barcode that has an identifiable signature based on the identifiable tags and a genetic signature based on the sequence of the dual-barcode. Thus providing a link between detection of the identifiable tags and the specific sequence of each dual-barcode.

10

METHOD OF LINKING REAL TIME IDENTITY OF AN ENTITY TO SUBSEQUENT ANALYSIS

The dual-barcodes and/or barcoded supports may be used in methods of identifying an entity in real-time. For example, the entity may be one or more droplets each including a
15 distinct barcoded support as described herein. The droplet may be identified and/or tracked through detection of the identifiable tags bound to the dual-barcode. For example, if the identifiable moiety is an optically detectable moiety, the method may include optically detecting the entity via the optically detectable moiety. As each barcoded support includes at least one dual-barcode that is distinct to other entities, each entity in a plurality of entities
20 can be individually detected and therefore individually tracked and identified. After detection of entities, the sequence of the at least one dual-barcode can be analysed and linked to the real time detection of the entity. Thus providing a method of linking real time identification of an entity to subsequent analysis.

In some examples, the method includes combining one barcoded support with one entity. In
25 some examples, the entity is a droplet. For example, a barcoded support may be inserted into a droplet.

Methods of combining droplets and/or barcoded beads with droplets are well known in the art. For example, the may include droplet fusion systems or picoinjection systems such as those described in WO2022034344A1 and WO2022034345.

30 Once the entity and barcoded support have been combined the combined entity and barcoded support may be incubated in or exposed to predetermined conditions. Predetermined conditions may be any condition that may be of interest. For example, where the support or entity includes a cell, the conditions may be selected to evoke a response from the cell. In the case of chemical agent supports or peptide supports, the entity may be

exposed to conditions to trigger an event such as activate a reaction with one or more reagents within the entity, binding of two or more reagents within an entity, interaction and response of reagents within an entity (i.e. drug interaction with a target such as cell). In the case of a target region that binds to a target the entity may be exposed to conditions to

5 promote binding of the target to the target region.

In some examples, an event includes a reaction and the method includes detection and/or analysis of the reaction kinetics. In some examples, an event includes a response (e.g. from a cell) and the method includes detection and/or analysis of the response. In some examples, an event includes a reaction and a response (e.g. from a cell) and the method

10 includes detection and/or analysis of the reaction kinetics and/or response.

In some examples, subpopulations or phenotypes of targets, supports or entities that have undergone an event may be identified and/or tracked by virtue of the real-time detection of the identifiable tags.

Simultaneously or subsequently to incubation or upon exposure the identifiable tag may be detected. It will be understood, that each barcoded support may include a plurality of the same dual-barcodes which may increase the level of detectable signal. In addition, each entity with a different barcoded support may be simultaneously incubated and exposed and detected. Therefore, a plurality of signals each from a different barcoded support may be detected at the same time.

20 The signal detected will depend on the type of identifiable moiety used. For example, in the case of identifiable moieties that are optical moieties, detection may comprise optical detection such as microscopy based methods. For example, if the identifiable moieties are fluorescent molecules the entities may be detected using fluorescence. In some examples, the method may include exposing the identifiable moieties to a stimuli to induce a signal to
25 produced by the identifiable moieties. For example, in the case of fluorescent moieties each entity may be exposed to electromagnetic radiation induce fluorescence of each fluorescent identifiable moiety.

Each dual-barcode may be identified by its specific identifiable signature produced by virtue of the distinct number of identifiable tags of each dual-barcode.

30 After incubation or exposure, the entities may be collected and subjected to further analysis. For example, analytes of interest within the entity may be analysed using any suitable methods. For example, where incubation or exposure comprises carrying out an *in vitro* transcription and/or translation, the method may include analysing the produced transcript or peptide for each entity.

In some examples, entities such a droplets, may be merged or split during real time detection. As the methods provided herein allow for real time detection and therefore tacking of each entity via its distinct barcoded support, merged and/or split droplets can be tracked.

- 5 In some examples, detection includes tracking entities in real-time. In some examples, tracking allows for a feedback loop that may be utilised for sorting entities.

The barcoded supports may be subjected to an amplification reaction in order to amplify the dual-barcodes. The amplified dual barcodes may then be sequenced.

- 10 In some examples, detection includes quantitative detection of entity contents. For example, using absorbance analysis.

In some examples, the barcoded support is separated from the entity, for example into a fist split entity prior to amplification. In some examples, the dual-barcodes may be reversibly attached to the support and may be cleaved or released from the support prior to amplification.

- 15 Methods of amplification are well known in the art and may be selected depending on the dual-barcode structure and/or the target of the target region. For example, amplification for DNA based barcodes may be carried out using PCR based methods. In some examples, amplification is a method of emulsion amplification. Emulsion PCR refers to a method of PCR where dilution and compartmentalization of template molecules in water droplets
20 occurs in a water-in-oil emulsion. The dilution can be controlled such that a plurality of droplets each contain a single template molecule, and function as micro PCR reaction vessels. Emulsion PCR can prevent the loss of rare templates through non-amplification or competition with more abundant templates.

- 25 In some examples, where the target includes an RNA amplification may be a reverse transcription reaction which may produce a reverse transcript that includes the sequence of the targeted mRNA and the sequence of the dual-barcode.

- The sequence of each dual-barcode may be linked to a identifiable signature detected. For example, the sequence of the dual-barcode provides the sequence of the tag binding regions and therefore provides the number of each type of identifiable moiety that is or was
30 bound to the dual-barcode and therefore can be linked to its corresponding identifiable signature.

Sequencing of the dual-barcodes may be carried out using any suitable sequencing method. For example, using next generation sequencing methods. The sequencing may be based on any known platform for example sequencing by synthesis, semiconductor sequencing (Ion

Torrent), Sequencing by hybridisation (SOLiD), 454 pyrosequencing, nanopore sequencing and/or single molecule real time sequencing, such as techniques available from Pacific BioSciences®.

5 In some examples, where each entity includes a cell or cellular components from a single, the analysis of the dual-barcode and any analytes from the entity may be analysed using single cell analysis methods. For example, using single cell sequencing methods. "Single cell sequencing" refers a single experiment allowing analysis of genetic information (DNA, RNA, epigenome, etc.) at the level of a single biological cell. The main difference between single cell and bulk sequencing is that each sequencing sample represents a single cell, 10 instead of a population of cells. For example, where the target region is configured to bind to non-specific mRNAs, specific RNAs, proteins, metabolites, genes, or any other cellular components, the analysis may provide information about the target or targets from a single cell.

15 In some examples, analytes from each entity may be further analysed using any suitable analysis. The analysis used may depend on the analyte.

For example, where the analyte is a cell or a cell lysate further analysis may include analysis of transcriptome, genome, epigenome, proteome, epitome, secretome or metabolome of single cells either by themselves or in combinations (multi-omics).

20 For example, conventional scRNA-seq, scRNA-seq of specific cell types (based on staining), sequencing of transcriptomes and epitopes (based on CITE-Seq (Cellular Indexing of Transcriptomes and Epitopes by Sequencing), total RNA, analysis of small RNA (miRNA and snRNA, snoRNA) from single-cells, multiomic assays (DNA-seq plus RNA-seq or ATAC-seq plus RNA-seq), RNA-seq of interacting cells, SINC-seq, targeted single-cell sequencing.

25 In some examples, where the identifiable tags include optically detectable moieties, detecting may include:

- a. simultaneously illuminating two or more of the fluorophores of distinct identification tags; and
- b. capturing fluorescence signals generated from the two or more fluorophores and separating each signal into distinct channels, each channel 30 corresponding to each fluorophore; thereby allowing for simultaneous capture of multi-spectral fluorescence data from single entities.

Detecting may also include extracting fluorescence measurements from the captured fluorescence signals. Extracting may include:

- f. simultaneously receiving signals including information corresponding to the fluorescence measurements;
- g. processing the signals; the processing comprising:
 - ix. Gaussian blurring the signals to reduce noise;
 - 5 x. thresholding the Gaussian blurred signals to produce binary masks which denote the presence of entities against a background;
 - xi. refining the binary masks and correcting for artefacts using hole-filling;
 - xii. optionally; wherein the refined binary masks comprises two or more entities in close proximity to each other, resolving the two or more entities using morphological erosion of the binary masks to isolate and delineate individual entities;
 - 10 xiii. for each entity in the binary mask, edge detecting coordinates along each edge of each entity detected; and
 - xiv. applying elliptical fitting to the coordinates.

15

In some examples, extracting may further include calculating the mean of each data point corresponding to each detected entity to estimate the measurement of each fluorophore in each entity.

In some examples, extracting may include:

20

before calculating the mean, employing a fitting error metric for assessing accuracy of the elliptical fitting; and

excluding those detected entities exceeding a defined threshold of the fitting error metric thereby excluding detected entities that do not correspond to a single entity.

25

As each barcoded support has a unique dual-barcode, the expected value of each pixel corresponding to the barcoded support is the same. Therefore, calculating the mean across all pixels corresponding to a single barcoded support results in a precise measurement of the dual-barcode(s) attached thereto.

30

As such, also provided herein is a method for extracting individual measurements from a plurality of simultaneously captured measurements of a plurality of barcoded entities; the method comprising:

simultaneously receiving, from at least one sensor, signals including information corresponding to the plurality of measurements;

processing the signals; the processing comprising:

Gaussian blurring the signals to reduce noise;

thresholding the Gaussian blurred signals to produce binary masks which denote the presence of one or more of the plurality of barcoded entity against a background;

refining the binary masks and correcting for artefacts using hole-filling;

optionally; wherein the refined binary masks comprises two or more entities in close proximity to each other, resolving the two or more entities using morphological erosion of the binary masks to isolate and delineate individual entities;

for each entity detected in the binary mask, edge detecting coordinates along each edge of each entity detected; and

applying elliptical fitting to the detected coordinates; and

for each entity detected, calculating the mean of each data point corresponding to each detected entity to estimate the measurement in each entity,

thereby extracting measurements for all entities detected; and

outputting one or more signals indicative of the extracted measurements.

Is some examples, the method may include:

before calculating the mean, employing a fitting error metric for assessing accuracy of the elliptical fitting; and

excluding those detected entities exceeding a defined threshold of the fitting error metric thereby excluding detected entities that do not correspond to a single entity.

Also provided is apparatus for extracting individual measurements from a plurality of simultaneously captured measurements of a plurality of barcoded entities, the apparatus comprising:

communication circuitry configured to receive, from at least one sensor, signals including information corresponding to the plurality of measurements;

a memory configured to store machine-readable instructions; and

processing circuitry configured to operably execute the machine-readable instructions to process the received signals to:

Gaussian blur the signals to reduce noise;

threshold the Gaussian blurred signals to produce binary masks which denote the presence of one or more of the plurality of barcoded entity against a background;

refine the binary masks and correcting for artefacts using hole-filling;

5 optionally; wherein the refined binary masks comprises two or more entities in close proximity to each other, resolve the two or more entities using morphological erosion of the binary masks to isolate and delineate individual entities;

for each entity detected in the binary mask, edge detect coordinates along each edge of each entity detected; and

10 apply elliptical fitting to the detected coordinates; and

for each entity detected, calculate the mean of each data point corresponding to each detected entity to estimate the measurement in each entity,

thereby extracting measurements for all entities detected; and

output one or more signals indicative of the extracted measurements; and

15 control the communication circuitry to output one or more signals indicative of the extracted measurements.

In some examples, the apparatus includes communication circuitry, memory circuitry, processing circuitry, and a display. However, the apparatus is not limited to this and may be any suitable computing device. Furthermore, in some embodiments the apparatus may be in
20 communication with another apparatus having a display, such that the one or more signals indicative of the extracted measurements are output to the display of the another apparatus so as to be remotely displayed.

In some examples, the communication circuitry may be provided in one or more communication devices, and may be of any suitable type for communicating with the sensor
25 to receive data signals indicative of the plurality of measurements. In some examples, the communication circuitry includes a wireless module to receive the signals from the sensor wirelessly. For example, the communication circuitry may include a Bluetooth®, Wi-Fi, WLAN and/or network data module. The communication circuitry may comprise a receiver in addition to a transmitter or may comprise a transceiver adapted to both transmit information
30 to and receive information from the processing circuitry.

In some examples, the processing circuitry may be any suitable processing means, and may be provided in one or more processors, such as an electronic processing device or computer processor. In some examples, the processing circuitry is communicatively coupled to the

communication circuitry and the memory. In some examples, the memory may comprise any suitable memory means and is arranged to store machine-readable instructions. In some examples, the processing circuitry is arranged to operably execute the machine-readable instructions stored on the memory. In some examples, the processing circuitry comprises an input means and an output means. In some examples, the input means may comprise an electrical input of the processing circuitry. In some examples, the output means may comprise an electrical output of the processing circuitry. In some examples, the input means and output means may be unified such as in the form of a network interface which inputs and outputs data, for example to a communication bus of the communication circuitry. In some examples, the processing circuitry may therefore receive data from the communication bus and output data onto the communication bus.

In some examples, the communication circuitry is configured to communicate with the at least one sensor.

In some examples, the communication circuitry is in communication with a database arranged on a remote server, the database configured to receive the one or more signals and store thereon data indicative of the extracted measurements.

In some examples, the at least one sensor is an image sensor/camera/photodiode

In some examples, the apparatus further comprises a display, wherein the processing circuitry is arranged to control the display to display information indicative of the extracted measurements.

In some examples, the display may comprise any suitable display means and is in communication with the communication circuitry. As above, the processing circuitry is also in communication with the display and is arranged to control the display via the communication circuitry to output the signals indicative of the extracted measurements by displaying information corresponding to the extracted measurements. In some examples, the extracted measurements may be displayed in the form of e.g. graph, chart, table, histogram, etc. For example, a histogram can displays the relevant property of the contents of each droplet, before the contents of the droplets are identified via the dual-barcode; or a graph that links the contents of each droplet with the corresponding measured properties once the dual-barcode has been processed.

In some examples, the sensors comprise image sensors for detecting fluorescence, such as a camera. For instance, a camera can be used to image the droplets. An OptoSplit (emission image splitter) can then be used to split the signal from different fluorophores and direct it to different areas of the same camera chip. This way, the signal from different fluorophores can be received simultaneously. One specific example of a camera that can be used as a sensor

is an Excelitas pco.edge 4.2 bi USB sCMOS camera, and the OptoSplit may be the OptoSplit III from Cairn-Research. An optical spectrometer as a sensor may be used to identify dual-mode barcodes.

5 The processing circuitry is thus arranged to perform image processing on the image data corresponding to the fluorescence signals received from the sensor via the communication circuitry, to extract fluorescence measurements from the imaged droplets.

In some example, output data may be further analyzed using machine learning. For example, a neural network based machine learning model may be used. However, the disclosure is not limited to this, and in other examples of the disclosure, output data may be
10 determined using a model, such as a rules-based algorithm.

For example, artificial intelligence (AI) can be used in multiple ways. First, an AI model may be used to perform processing of signals as described in relation to the method of linking real time identity of an entity to subsequent analysis, for detecting droplets and beads. In this regard, a neural network can be trained to detect the droplets and beads in acquired images.
15 Similarly, AI could be used to map the fluorescent readout from beads to recover the associated DNA sequence.

AI can be used for any downstream analysis in a task-specific way. AI may also be an important component of the uses and methods disclosed herein in connection with imaging flow cytometry. Due to the high speed of droplets, it is likely that motion deblurring will be
20 necessary for downstream analysis. This could be accomplished via deep learning. Similarly, as exposure time will be on the order of milliseconds, denoising will be required. Deep learning outperforms traditional methods in this setting – e.g. as reported by Weigert *et al.*, Nat Methods. 2018 Dec;15(12):1090-1097. doi: 10.1038/s41592-018-0216-7. The use of AI may thus enable the creation of a super-resolution imaging flow cytometry system, by
25 combining structured illumination microscopy (SIM) with a deep learning-based method capable of performing SIM reconstruction from noisy and blurry images. Deep learning is also reported by Christensen *et al.*, Biomed Opt Express. 2021 Apr 15;12(5):2720-2733. doi: 10.1364/BOE.414680. to outperform traditional methods for SIM reconstruction in standard imaging settings.

30 The ability to collect large amounts of data the invention enables the training of custom AI models to predict protein function, biocatalysis reaction kinetics, cell drug response, and more. For example, by assembling datasets of protein function, models capable of *de novo* protein design for arbitrary functions can be created. An example working on a similar task is reported by Wang *et al.*, Science. 2022 Jul 22;377(6604):387-394. doi:0.1126/science.abn2100. So far similar methods are limited by the assumption that
35

protein function is entirely determined by its shape. This is due to the existing datasets containing information on sequence and structure only, as it would not be feasible without a technology such as the invention as disclosed herein to collect a sufficiently large dataset of arbitrary protein properties, such as corresponding biocatalysis kinetics.

- 5 The ability to conduct large numbers of experiments provides synergy with a deep learning-based automatic scientist (King et al., Science. 2009 Apr; vol. 324(5923) pp 85-89 doi: 10.1126/science.116562). AI can thus be used not only to analyse the results of experiments, but also to stage experiments that need to be conducted in order to improve its performance, the understanding of a particular problem, or both. AI may thus be used in
10 connection with the invention as disclosed herein for the purposes described above.

Also provided is a system for extracting individual measurements from a plurality of simultaneously captured measurements of a plurality of barcoded entities, the system comprising:

- the apparatus for extracting individual measurements from a plurality of
15 simultaneously captured measurements of a plurality of barcoded entities; and

a server in communication with the apparatus and arranged to receive the one of more signals indicative of the extracted measurements, the server comprising a database arranged thereon for storing the extracted measurements.

- Also provided are machine readable instructions, which when executed by processing
20 circuitry, cause the processing circuitry to carry out a method for extracting individual measurements from a plurality of simultaneously captured measurements of a plurality of barcoded entities as described herein.

- Also provided is a machine readable data storage medium having tangibly stored thereon the machine readable instructions for extracting individual measurements from a plurality of
25 simultaneously captured measurements of a plurality of barcoded entities as described herein.

- In some examples, the method includes simultaneously receiving signals including information corresponding to the plurality of measurements, and processing the received signals, which includes the following steps: Step i. includes Gaussian blurring the signals to
30 reduce noise. Step ii. includes thresholding Gaussian blurred signals to produce binary masks which denote the presence of one or more of the plurality of barcoded entity against a background. Step iii. includes refining the binary masks and correcting for artefacts using hole-filling. Step v. includes for each entity detected in the binary mask, edge detecting coordinates along each edge of each entity detected. Step vi. includes applying elliptical

fitting to the detected coordinates. Step vii. includes calculating the mean of each data point corresponding to each detected entity to estimate the measurement in each entity. In doing so, the method extracts measurements for all entities detected.

5 In some examples, the refined binary masks comprises two or more entities in close proximity to each other, the processing includes an additional step iv. of resolving the two or more entities using morphological erosion of the binary masks to isolate and delineate individual entities.

10 In the case of defined mass identifiable moieties detection may be carried out using methods such as mass spectroscopy. For example, a support may include a plurality of dual-barcodes each including identical tag binding regions and number identifiable tags but a part of the plurality are reversibly attached to the support and part of the plurality are irreversibly attached to the support. After incubation or exposure, the reversibly attached dual-barcodes may be cleaved or released from the support into the fluid of the entity and the entity (i.e. droplet) may be split to provide a first split entity including the released dual-barcodes and a
15 second split entity including the barcoded support. The first split entity may then be analysed by mass spec to identify the identifiable moieties of the specific dual-barcode and the second split entity may be analysed using subsequent analysis methods such as amplification and sequencing of the dual-barcode. The mass spectroscopy data may then be linked to the specific sequence for each dual barcode or barcoded support analysed.

20 In some examples, multiple first split entities may be analysed simultaneously using different mass spectroscopy channels. In some examples, a further mass spectroscopy channel may be used to detect analytes produced within the entity.

It will be understood that the method described above may be carried out using any suitable detection means that is capable of detecting the identifiable moieties of the identifiable tags.

25 In one aspect there is provided a system for carrying out the methods described. For example, a system may include one or more modules for generation of droplets; incorporation of barcoded supports into a droplet; incubation or exposure of droplets to predetermined conditions; detecting barcoded supports; collection of droplets and/or barcoded supports; and analysis of collected barcoded supports and/or droplets including
30 any analytes.

Suitable systems may be microfluidic systems or modules thereof such as those described in WO2022034344A1 and WO2022034345. Other suitable systems will be known by those skilled in the art.

In some examples, the system may include an imaging module for real-time visualization of entities including a barcoded support or dual-barcode as described herein.

In some examples, the system may include a time-analysis module for dynamic process monitoring in real-time. In some examples, the system may include a time-analysis module
5 that is configured to detect and record events and correlate events with a time metric. In some examples, the system may include a time-analysis module for configured to detect and record drug or molecule kinetics and responses of targets, supports or cells and correlate kinetics and responses with a time metric.

In some examples, the system may include a data processing and storage unit to process
10 and/or store data from analyses of entities in real time, detection and/or sequence analysis.

USES

It will be understood by the skilled person that the dual-barcodes described herein may be used in any method that employs the use of barcodes. One advantage of the uses and methods disclosed herein is that they links the benefits of test tubes with droplets. Test tubes
15 are identifiable. Droplet microfluidics are high-throughput and allow species compartmentalization. In general, compartmentalisation does not work with flow cytometry for example, as the phenotype genotype linkage is lost. Droplets provide this compartment. Additionally data on all droplets whether that is cells, drugs or anything that can be compartmentalised can be collected with the methods and uses disclosed herein.

For example, the dual-barcodes described herein may be used in analysis of single cell
20 analyses (such as single cell RNA analysis or multi-omics analyses), genetic material analysis, epigenetic analysis, chemical library analysis, peptide library analysis, nucleic acid library analysis, metabolic library analysis, reaction condition analysis, reaction analysis (such as reaction kinetics), synthetic biology, drug interaction analysis, enzymology, patient
25 sample analysis, pathogen analysis, cytology, therapeutic agent analysis, DNA encoded library analysis, gene engineering, cell engineering, *in vitro* transcription and/or *in vitro* transcription and translation.

Single cell analyses that can be carried out using the dual barcodes and barcoded supports of the invention include: analysis of transcriptome, genome, epigenome, proteome, epitome,
30 secretome or metabolome of single cells either by themselves or in combinations (multi-omics).

RNA analysis

In some examples, the dual barcodes and barcoded supports may be used in methods of single cell RNA analysis. By “single-cell RNA analysis” is meant the extraction of RNA from a cell or cell structure, for sequencing.

As such, provided herein is a method of single cell RNA analysis comprising combining one or more barcoded supports described herein with one or more cells or cell lysates including one or more RNA molecules or RNAs from one or more cells (such as mRNAs) in one or more entities (e.g. droplets), incubating the one or more barcoded supports and RNAs under conditions to promote binding of the RNAs to a target region of each one of the barcoded supports, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the RNA bound dual-barcodes and/or bound RNA and sequencing the dual-barcodes and/or the bound RNAs.

In some examples, the RNA may be all mRNAs of a cell (e.g. transcriptome analysis). In some examples, the RNA may be a specific RNA and the method may identify cells expressing a specific RNA and the quantity thereof.

As such, the method may be used to detect specific RNA. Therefore, the method may be used to determine whether cells include RNAs. In some examples, cells may be exposed to one or more conditions and/or reagents of interest (such as therapeutic agents) and the method may be used to detect the presence of a specific RNA in response to the conditions and/or reagents of interest. In some examples, the method may be used to analyse the transcriptome of cells each exposed to different conditions and/or reagents and the method may provide details of changes in transcriptome of each cell in response to each condition and/or reagent.

In some examples, a barcoded support may include a support or target region that includes one or more RNAs of interest (such as an mRNA). The method may determine different cellular components or reagents that interact with the RNA of interest. For example, this may allow for determination of cellular components or reagents (such as proteins or therapeutic molecules) that may interact with a specific RNA. For example, a target region may include a RNA of interest that is associated with a disease and the method may detect cellular components (such as proteins or nucleic acids) that bind to the RNA of interest. As such, the dual-barcodes described herein may also be used in pathology.

Genetic material Analysis

In some examples, the dual-barcodes described herein may be used to analyse genetic material. For example, genomes, genes or parts thereof from cells.

In some examples, there is provided a method of genetic material analysis comprising combining one or more barcoded supports described herein with one or more cells, cell lysates including genetic material of one or more cells or genetic material from one or more cell in one or more entities (e.g. droplets), incubating the one or more barcoded supports
5 and genetic material under conditions to promote binding of the genetic material to a target region of the each one of the barcoded supports, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the genetic material bound dual-barcodes and/or bound genetic material and sequencing the one or more dual-barcode and/or the bound genetic material.

10 In some examples, the genetic material is a single gene. In some examples, the target region may bind to a specific gene or variant thereof. As such, the method may be used to detect modifications (such as mutations) of a gene of interest. In some examples, the genetic material may have been obtained from or part of a patient sample and the method may be used to detect genetic modifications associated with one or more diseases or
15 conditions.

In some examples, the method analyses the entire genome of a cell and the method is a method of genome analysis. Therefore, in some examples, there is provided a method of single cell genome analysis comprising combining one or more barcoded supports described herein with one or more cells, cell lysate including the genome of one or more cells or one or
20 more genomes from one or more cells in one or more entities (e.g. droplets), incubating the one or more barcoded supports and genomes or fragments thereof under conditions to promote binding of the one or more genomes or fragments thereof to a target region of each barcoded support, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the genome or fragments thereof bound dual-barcodes and/or
25 bound genome or fragments thereof and sequencing the dual-barcode and/or the bound genome or fragments thereof.

In some examples, the genome or fragments thereof from a single cell may be bound by different barcoded supports.

In some examples, a barcoded support may include a support or target region that includes
30 one or more genes of interest. The method may determine different cellular components or reagents that interact with the gene of interest. For example, this may allow for determination of cellular components or reagents (such as proteins or therapeutic molecules) that may have importance in diseases and conditions. For example, a target region may include a gene of interest that is associated with a disease and the method may

detect cellular components (such as proteins or nucleic acids) that bind to the gene of interest. For example, the methods may detect transcription factors.

The use of the dual-barcodes may advantageously allow for the high throughput analysis of multiple genes or multiple gene variants.

5 Epigenetic Analysis

In some examples, the dual-barcodes described herein may be used for analysis of epigenetics of a cell or of genetic material. Epigenetic analysis refers to determining the state, or condition of DNA, and its interaction with specific proteins and their modified isoforms in a sample, and involves analysing or detecting epigenetic marks in a sample. In some examples, analysis may be of single genes or an entire genome, for example epigenome analysis. "Epigenome" refers to the state or pattern of alteration of genomic DNA due to covalent modifications of the DNA or proteins attached to the DNA. Examples of such alterations include methylation at position 5 of cytosine in CpG dinucleotides, acetylation of histone lysine residues, and other genetic or non-hereditary causes not due to alterations in the underlying DNA sequence.

Methods of analysing epigenetics may be similar to methods of gene or genome analysis above but may include using epigenetic sensitive amplification and/or sequencing methods and/or epigenetic detection methods after amplification of barcoded support bound genetic material. For example, using western blot analysis; Chromatin Immunoprecipitation; Chromatin Immunoprecipitation followed by quantitative PCR (ChIP-qPCR); chromatin immunocleavage (ChIC) methods, cleavage under targets and tagmentation (CUT&Tag) methods, cleavage Under Targets and Release Using Nuclease (CUT&RUN) methods, Directed Methylation with Long-read sequencing (DiMeLo), DNA adenine methylase identification (DamID) methods, chromatin endogenous cleavage (ChEC) methods and/or nanopore-sequencing-based Histone-modification and Methylome joint-profiling methods; and/or Biotin-ChIP.

In some examples, the method may use a dual-barcode that includes a target region for binding to a specific epigenetic marker. For example, the method allow for the detection and subsequent analysis of epigenetic markers of genetic material by barcoding of genetic material that includes a particular epigenetic marker.

In some examples, the method analyses the epigenetics of an entire genome of a cell and the method is a method of epigenome analysis. Therefore, in some examples, there is provided a method of single cell genome analysis comprising combining one or more barcoded supports described herein with one or more cells, cell lysates including the genomes of the cells or one or more genomes from one or more cells in one or more entities

(e.g. droplets), incubating the one or more barcoded supports and genomes or fragments thereof under conditions to promote binding of the genomes or fragments thereof to a target region of each barcoded support, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the genome or fragments thereof bound dual-barcodes and/or bound genome or fragments thereof and sequencing the dual-barcode and/or the bound genome or fragments thereof.

As such, the method may be used to detect a specific epigenetic marker. Therefore, the method may be used to determine whether cells include a specific epigenetic marker. In some examples, cells may be exposed to one or more conditions and/or reagents of interest (such as therapeutic agents) and the method may be used to detect the presence of a specific epigenetic marker in response to the conditions and/or reagents of interest. In some examples, the method may be used to analyse the epigenome of cells each exposed to different conditions and/or reagents and the method may provide details of changes in the epigenome of each cell in response to each condition and/or reagent.

In some examples, a barcoded support may include a support or target region that includes one or more epigenetic markers or nucleic acids including one or more epigenetic markers. The method may determine different cellular components or reagents that interact with a specific epigenetic marker. For example, this may allow for the determination of molecules (such as proteins) that play a role in epigenetics that may have importance in diseases and conditions. For example, a target region may include an epigenetically modified gene of interest with one or more epigenetic markers that is associated with a disease and the method may detect cellular components (such as proteins or nucleic acids) that bind to the epigenetically modified gene of interest.

The use of the dual-barcodes may advantageously allow for the high throughput analysis of the epigenetics of multiple genes or multiple gene variants.

Protein Analysis

In some examples, the dual-barcodes described herein may be used to analyse one or more proteins or variants thereof. For example, proteins or parts thereof from cells such as cells from a patient sample. Accordingly, the dual-barcodes described herein may be used in peptide library analysis.

In some examples, the dual-barcodes described herein may be used to analyse all proteins of cell (e.g. proteome analysis).

In some examples, there is provided a method of single cell proteome analysis comprising combining one or more barcoded supports described herein with one or more cells, cell

lysates including one or more proteins of the cells or proteins from one or more cells in one or more entities (e.g. droplet), incubating the one or more barcoded supports and proteins under conditions to promote binding of the proteins to a target region of each barcoded support, detecting identifiable moieties of the one or more dual-barcodes and subsequently
5 amplifying the dual-barcodes and sequencing the one or more dual-barcodes. The method may further include analysing each bound protein or all proteins from a single cell using well known methods in the art.

In some examples, the method analyses a single protein or variants thereof. In some examples, the target region of each barcoded support may bind to a specific protein or
10 variant thereof. As such, the method may be used to detect modifications (such as mutations) of a protein of interest. In some examples, the protein may have been obtained from or be part of a patient sample and the method may be used to detect protein modifications associated with one or more diseases or conditions.

As such, the method may be used to detect specific protein. Therefore, the method may be
15 used to determine whether cells express or include a specific protein. In some examples, cells may be exposed to one or more conditions and/or reagents of interest (such as therapeutic agents) and the method may be used to detect the presence of a specific protein in response to the conditions and/or reagents of interest. In some examples, the method may be used to analyse the proteome of cells each exposed to different conditions and/or
20 reagents and the method may provide details of changes in proteome of each cell in response to each condition and/or reagent.

In some examples, a barcoded support may include a support or target region that interacts with one or more proteins in a sample. The method may determine proteins that interact with a target region or support. For example, this may allow for determination of protein
25 binding partners that may have therapeutic uses. For example, a target region may include a gene of interest and the method may detect proteins that bind to the gene of interest, a protein of interest and the method detects proteins that bind to the protein of interest or a chemical moiety of interest and the method detects proteins that bind to the chemical moiety of interest.

30 Secretome Analysis

In some examples, the dual-barcodes described herein may be used to analyse one or more secreted molecules secreted from one or more cells.

In some examples, the dual-barcodes described herein may be used to analyse all secreted molecules of one or more cells (e.g. secretome analysis).

In some examples, there is provided a method of single cell secretome analysis comprising combining one or more barcoded supports described herein with one or more cells or one or more culture media conditioned by the one or more cells (i.e. including secreted molecules from the cells) in one or more entities (e.g. droplet), incubating the one or more barcoded supports and secreted molecules under conditions to promote binding of the secreted molecules to a target region of each barcoded support, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the dual-barcodes and sequencing the one or more dual-barcodes. The method may further include analysing each bound secreted molecule or all secreted molecule from a single cell using well known methods in the art.

In some examples, the method analyses a single secreted molecule. In some examples, the target region of each barcoded support may bind to a specific secreted molecule. As such, the method may be used to detect specific secreted molecules. Therefore, the method may be used to determine whether cells secrete a specific molecule. In some examples, cells may be exposed to one or more conditions and/or reagents of interest (such as therapeutic agents) and the method may be used to determine the secretion of certain molecules in response to the conditions and/or reagents of interest. In some examples, the method may be used to analyse the secretome of cells each exposed to different conditions and/or reagents and the method may provide details of changes in secretome of each cell in response to each condition and/or reagent.

Metabolite Analysis

In some examples, the dual-barcodes described herein can be used in metabolic library analysis.

In some examples, the dual-barcodes described herein may be used to analyse one or more metabolites of one or more cells.

In some examples, the dual-barcodes described herein may be used to analyse all metabolites of one or more cells (e.g. metabolome analysis). "Metabolome" as used herein refers to the complete set of small-molecule metabolites to be found within an organism or cell.

In some examples, there is provided a method of single cell metabolome analysis comprising combining one or more barcoded supports described herein with one or more cells, cell lysates including one or more metabolites of the cells or metabolites from one or more cells in one or more entities (e.g. droplet), incubating the one or more barcoded supports and metabolites under conditions to promote binding of the metabolites to a target region of each barcoded support, detecting identifiable moieties of the one or more dual-

barcodes and subsequently amplifying the dual-barcodes and sequencing the one or more dual-barcodes. The method may further include analysing each bound metabolite or all metabolites from a single cell using well known methods in the art. For example, by mass spectroscopy, nuclear magnetic resonance, and the like.

- 5 In some examples, the method analyses a single metabolite. In some examples, the target region of each barcoded support may bind to a specific metabolite. As such, the method may be used to detect specific metabolites. Therefore, the method may be used to determine whether cells include a specific metabolite. In some examples, cells may be exposed to one or more conditions and/or reagents of interest (such as therapeutic agents)
- 10 and the method may be used to determine the presence of specific metabolites in response to the conditions and/or reagents of interest. In some examples, the method may be used to analyse the metabolome of cells each exposed to different conditions and/or reagents and the method may provide details of changes in the secretome of each cell in response to each condition and/or reagent.

15 Sample Testing

In some examples, the dual-barcodes described herein may be used for sample analysis. For example, patient samples, or environmental sample.

- Patient samples may include a biological fluid (also referred to herein as a bodily fluid) sample. "Biological fluid sample" encompasses a blood sample. A blood sample may be a
- 20 whole blood sample, or a processed blood sample e.g. buffy coat. Methods for obtaining biological fluid samples (e.g. whole blood,) from a subject are well known in the art.

- In some examples, the dual-barcodes may include target regions that bind to one or more biomarkers and the methods described herein may be used to detect the presence of the biomarker in a sample. In some examples, the biomarker may be associated with a disease
- 25 or condition. In such examples, the methods described herein may be capable of or help in the diagnosis of a disease or condition. As such, the dual-barcodes described herein may be used in methods of pathology.

- In some examples, the dual-barcodes may include target regions that bind to one or more pathogens and the methods described herein may be used to detect the presence of a
- 30 pathogen in a sample. For example, in a patient sample or environmental sample. As such, the dual-barcodes described herein may be used in methods of pathogen detection. Accordingly, the dual-barcodes described herein may be used in pathogen analysis.

In some examples, the dual-barcodes described herein may include target regions that bind to one or more cells. In some examples, the dual-barcodes may have target regions that

bind to specific cells, for example via specific cell markers. For example, the target regions may be configured for selectively binding cancer cells. For example, the methods described herein may be used to analyse patient samples from patients suspected of having a cell disease such as cancer. As such, the dual-barcodes described herein may be used in

5 methods of cytology and pathology.

In some examples, there is provided of method of sample testing, the method including combining one or more barcoded supports described herein with one or more samples in one or more entities (e.g. droplet), incubating the one or more barcoded supports and samples under conditions to promote binding of a target in the sample to a target region of each barcoded support, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the dual-barcodes and sequencing the one or more dual-barcodes. The method may further include analysing each bound target.

10

The high throughput nature of the methods described herein may allow for testing of greater numbers of samples in a given time period than other barcoding methods.

15 DNA Encoded Libraries (DELs)

The dual-barcodes described herein may be used to form DNA encoded libraries of molecules such as chemical moieties, peptides or therapeutic agents. For example, a distinct barcode may be attached to each chemical moiety, peptide or therapeutic agent in a library thereof to provide a DNA encoded library thereof. DELs can be made using the same split and mix method on beads as described herein but instead adding in the optical barcoded DNA as part of the DNA sequence used. Additionally DELs can be made as normal and then split and mix can be performed afterwards as described herein to ligate the optical barcode DNA. Tagging may happen once the library is fully made, as one of the last steps before screening/analysis.

20

25 In some examples, the dual-barcodes described herein may be attached to the DNA element of an already formed DNA encoded library. Therefore providing an additional identification tag (barcode) that is identifiable in real-time.

For example, a barcoded support or dual-barcode of the invention may be attached to or combined with a member of a DNA encoded library in an entity. This may then be subject to an event which can be tracked and monitored in real time by virtue of the identifiable tags. Tracking and identification may allow for multiple DEL members to be mixed and/or split and tracked. Any products of reactions of DEL members may then be tracked, analysed and linked to specific entities by virtue of the dual-barcode sequence.

30

In some examples, DEL members may be combined with a barcoded support or attached to a dual-barcode in an entity and a DEL target. For example, the DEL target may be a gene, protein, or cell that has therapeutic interest. The methods described herein could allow interactions of the DEL member and DEL target to be identified and tracked and any products of each interaction can be linked to real-time measurements by sequencing of the barcode.

In some examples, the barcoded support may include a DEL target as a support. For example, the support may be peptide of interest or cell of interest. The interaction of the DEL target and the DEL member may then be identified, tracked and monitored in real time and any changes to a cell or the nature of the interactions may be analysed and linked to the real time analysis.

In some examples, the DEL member comprise peptides. In some examples, the DEL members comprise a therapeutic agents. In some examples, the DELs member comprise chemical agents. As such, in some examples, the dual-barcodes may be used in methods of chemical library analysis. In some examples, the dual-barcodes may be used in methods of nucleic acid library analysis. In some examples, the dual-barcodes may be used in methods of peptide library analysis.

Therapeutic Agent Analysis

In some examples, the dual-barcodes provided herein may be used in screening of therapeutic agents. For example, barcoded supports may include a target region that includes or is bound to a therapeutic target of interest. Each barcoded support in a library of barcoded supports may then be combined with a different therapeutic agent. The barcoded support allows identification, tracking and monitoring of the therapeutic target and therapeutic agent interaction and subsequent analysis may allow for any effects or analytes produced due to the interaction to be analysed. In some examples, the dual-barcodes may be used to select therapeutic molecules that bind to the target of interest.

In some examples, there is provided a method of therapeutic agent analysis comprising combining one or more barcoded supports described herein with therapeutic agents, in one or more entities (e.g. droplets), incubating the one or more barcoded supports and therapeutic agents to promote binding of the therapeutic agents to a target region of each barcoded support, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the dual-barcodes and sequencing the one or more dual-barcodes. The method may further include analysing each bound therapeutic agent using well known

methods in the art. For example, by mass spectroscopy, nuclear magnetic resonance, and the like.

In some examples, the barcoded support may not include a target region but be incorporated in an entity that includes a therapeutic agent and a target of interest. In some examples, the support may be the therapeutic target, for example, the support may be a cell.

In some examples, the cell may be pathogen cell such as a bacterial cell and the therapeutic agents may be investigated for anti-pathogenic properties.

In Vitro Transcription and Translation

In some examples, the dual-barcodes described herein may be used in methods of *in vitro* transcription. In some examples, the dual-barcodes described herein may be used in methods of *in vitro* transcription and translation (i.e. protein production).

For example, the target region of each barcoded support may include a gene of interest and regulatory elements for initiation of transcription and optionally translation. Barcoded supports may be combined with entities that include components for *in vitro* transcription and optionally translation. The combined barcoded supports and components for *in vitro* transcription and optionally translation may be incubated under conditions to lead to transcription of the gene of interest and optionally translation of the transcribed mRNA.

The use of the barcoded supports provided herein may provide for the high throughput production of a large number of mRNAs or proteins of interest each of which can be identified in real-time and identified by the dual-barcode sequence. In some examples, each gene of interest may encode a variant of protein and allow production of a large number of variants. For example, this may be useful for generating a library of proteins that may have therapeutic interest or application.

In some examples, there is provided a method of *in vitro* transcription comprising combining one or more barcoded supports comprising a target region encoding a gene of interest and regulatory elements for initiation of transcription with components for *in vitro* transcription, in one or more entities (e.g. droplets), incubating the one or more barcoded supports and components for *in vitro* transcription to promote transcription of the gene of interest each barcoded support and each dual-barcode thereof, detecting identifiable moieties of the one or more dual-barcodes and subsequently amplifying the dual-barcodes and sequencing the one or more dual-barcodes. The method may further include isolating and/or analysing each transcript generated.

In some examples, the entities may also include components for translation of a transcription product and therefore the method may produce one or more proteins. In some examples, components may be combined with the entity in which the transcription took place after transcription. For example, the method may split an entity including transcription products
5 and analyse the contents of a first split entity and translating the transcription products of a second split entity.

Reactions

The dual-barcodes described herein may be used in methods of testing multiple reactions and reaction conditions (i.e., reaction condition analysis and reaction analysis). For
10 example, multiple reaction components such as buffers, substrates, active agents, adjuvants, catalysts, therapeutic agents, or cells may each be encapsulated in individual entities (e.g. droplets). For example, a first droplet includes buffer A, a second droplet includes buffer B, a third droplet includes substrate A, a fourth droplet includes substrate B, a fifth droplet includes active agent A, and a sixth droplet includes active agent B. In some
15 examples, droplets may include the same reaction components at different concentrations thus allowing multiple reaction component concentrations to be tested.

Each of the droplets can include a distinct barcoded support that identifies each reaction component by the identification signature of each barcoded support. Various combination of each reactant component can be made using microfluidic devices by, for example, droplet
20 merging to form a reaction entity. Each reaction component in each reaction entity can be identified by virtue of its distant barcoded support in real-time. The products or results of each combination can then be analysed and linked to the specific combination by sequencing the dual-barcodes of each barcoded support.

In some examples, there is provided a method of testing multiple reactions comprising
25 combining one or more barcoded supports with one or more entities (e.g. droplets) wherein each entity includes a distinct reaction component and distinct barcoded support, combining two or more entities to provide a reaction entity, detecting identifiable moieties of each barcoded support in the reaction entity and subsequently amplifying the dual-barcodes and sequencing the dual-barcodes of each barcoded support in each reaction entity. The method
30 may further include analysing any reaction products produced in the reaction entity.

In some examples, each reaction entity may include an enzyme or variant thereof and the method may screen substrates for the enzyme and/or enzymatic activity of enzyme variants. In some examples, each barcoded support may include a reaction component of interest as a support. For example, the support may be a catalyst of interest and the method may test
35 reaction products and kinetics altered by the catalyst. In some examples, a support is an

enzyme of interest and the method may screen for substrates and reaction compositions that effect the enzymes activity, kinetics and products. As such, the dual-barcodes described herein may be used in methods of enzymology.

- 5 In some examples, at least one barcoded support in the reaction entity may include a target region that binds to a reaction product of interest and the method may also identify reaction compositions that lead to production of a specific reaction product.

In some examples, the method may include production of an identifiable product that the production of can be detected in real-time in combination with the identity of each barcoded support.

- 10 In some examples, each reaction entity may be incubated or exposed to different reaction conditions such as temperatures

Synthetic biology

In some examples, the dual-barcodes described herein by used in methods of synthetic biology.

- 15 Synthetic biology refers to the design and construction of biological elements, devices and systems, and the purposeful redesign of existing natural biological systems, widely used in the fields such as chemical synthesis, medicine, agriculture, and the environment.

- The barcoded supports described herein may be used in a similar manner described above in relation to reactions. For example, each component of a biological element, device or
20 system may be combined with a barcoded support of the invention in an entity (e.g. droplet). The assembly of different combinations of each component may then be tracked and analysed.

- As above, one or more components may be used as a support of at least one of the barcoded supports or may be bound by a target region of one or more of the barcoded
25 supports.

Combinatorial biochemistry

In some examples, the dual-barcodes described herein by used in methods of combinatorial biochemistry.

- Combinatorial biochemistry, also called combinatorial biosynthesis, comprises a series of
30 methods that establish novel enzyme–substrate combinations *in vivo* and, in turn, lead to the biosynthesis of new, natural product-derived compounds that can be used in drug discovery programs. Accordingly, the dual-barcodes described herein can also be used in enzymology.

The barcoded supports described herein may be used in a similar manner described above in relation to reactions and therapeutic agent analysis.

Screening/ quality control

In connection with the invention, screening beads (i.e., the support including dual barcode)
5 for quality control may be performed, which is useful for when the beads are made into a packaged product. The bead could be screened to remove large or small beads and non-constant levels of fluorophore after copolymerisation by attaching a constant tag to the linker DNA. This would be screened via Fluorescence-Activated Cell Sorting (FACS). Then using a microscope setup and droplet microfluidic sorting as disclosed herein, the beads can be
10 screened so that duplicates or non-library members can be removed in uses disclosed herein. Alternatively, during the analysis of beads, the results of any beads that do not 'fit' into the library or are ambiguous could be discarded.

KIT

Also provided herein is a kit of parts.

15 In some examples, the kit of parts includes a plurality of distinct barcoded supports and a plurality of corresponding identification tags as described herein for binding to the tag sequences of each of the at least one dual-barcodes of distinct barcoded supports.

In some examples, the kit of parts includes a plurality of distinct dual-barcodes and a plurality of corresponding identification tags as described herein for binding to the tag sequences of
20 each of the at least one dual-barcodes.

In some examples, the kit of parts includes a plurality of distinct tag binding regions as described herein and a plurality of corresponding identification tags as described herein for binding to the tag sequences of each of the distinct tag binding regions.

In some examples, where the kit includes a plurality of distinct dual-barcodes or plurality of
25 distinct tag binding regions, the kit may further comprise at least one type of support for attaching at least one of each distinct dual-barcode or at least one target binding region to.

In some examples, the kit of parts one or more further tag binding regions for linking to each one of the distinct tag binding regions of the plurality of barcoded supports, plurality of dual-barcodes or plurality of distinct tag binding regions.

30 In some examples, the kit of parts may include one or more reagents for attaching plurality of distinct dual-barcodes or plurality of distinct tag binding regions to a support. For example, the kit may include reagents for carrying out Golden Gate assembly or other methods of nucleic acid assembly.

In some examples, the kit of parts may include one or more reagents for linking one or more distinct tag binding regions each other or for linking at least further tag binding region to a barcoded support or dual-barcode. For example, the kit may include reagents for carrying out Golden Gate assembly or other methods of nucleic acid assembly.

- 5 In some examples, the kit of parts includes one or more further distinct tag binding regions and corresponding identification tags for each distinct tag binding region.

In some examples, the kit of parts includes instructions for producing one or more distinct dual-barcodes and/or barcoded supports. In some examples, the kit includes instructions for binding the identification tags to their corresponding tag binding sequences.

- 10 Each part of the kit of parts may be provided in a separate container. In some examples, each distinct dual-barcode, distinct barcoded support and/or distinct target binding region may be provided in a separate container.

- Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. For example, Singleton and Sainsbury, Dictionary of Microbiology and Molecular Biology, 2d Ed., John Wiley and Sons, NY (1994); and Hale and Marham, The Harper Collins Dictionary of Biology, Harper Perennial, NY (1991) provide those of skill in the art with a general dictionary of many of the terms used in the invention. Although any methods and materials similar or equivalent to those described herein find use in the practice of the present invention, the preferred methods and materials are described herein. Accordingly, the terms defined immediately below are more fully described by reference to the Specification as a whole. Also, as used herein, the singular terms "a", "an," and "the" include the plural reference unless the context clearly indicates otherwise. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art.
- 15
20
25

Aspects of the invention are demonstrated by the following non-limiting examples.

30

EXAMPLES

Method

- The core system consists of polyacrylamide beads (i.e. a support) that are copolymerized with DNA (attachment moiety). The DNA contains binding regions for a DNA barcode (i.e. a
- 35

dual functional barcode of the invention). The barcode was built in such a way as to allow for distinguishability of colors (or mass) based upon an ordered split and mix method of production.

- 5 The bead is made from polyacrylamide with a linker region (attachment moiety) which is a DNA molecule containing an acryl linker to allow copolymerization into the bead. It also contains a binding region for a single fluorophore and a Golden Gate site for restriction digestion by a type IIS nuclease.

Bead Construction

- 10 The bead and buffers were made following a detailed protocol ¹. Briefly, a 150 μ L mixture was made for bead generation which contained a buffering solution (10 mM Tris-HCl, 1 mM EDTA, 20 mM NaCl, pH 7.5), Ammonium persulfate (0.3%), Acrylamide (6.2%), Bis-acrylamide (0.18%), linker DNA (attachment moiety) (between 25 μ M and 100 μ M), and H₂O to reach 150 μ L total volume.
- 15 Beads were generated in a 20 μ m flow focusing microfluidics device at a volume of $\sim$ 4.2 pL or $\sim$ 20 μ m per bead. Briefly, tubing used to inject the bead mixture was washed with DNA Low Binding Buffer (Tris-HCl 10 mM, BSA 0.1%, Tween-20 0.5%), to prevent non-specific annealing of linker DNA(attachment moiety). Typical flow rates are 1.5 μ L/min for the bead mixture and 10 μ L/min for the oil phase (HFE-7500 (3M) with 2% fluorosurfactant 008 (Ran technologies)). Beads were collected in a microcentrifuge tube with 200 μ L Mineral Oil. The beads sat in a phase in-between the carrier oil and the Mineral Oil. After bead generation the carrier oil was removed using a syringe and around 0.4% TEMED (N,N,N',N' - Tetramethylethylenediamine) was pipetted into the bottom of the tube. The tube was then incubated overnight at 37 °C.
- 20 Subsequently the Mineral Oil was removed, 200 μ L of 1H, 1H, 2H, 2H-Perfluoro-1-octanol and 400 μ L of a wash buffer (Tris 50 mM, Tween-20 (0.1%)) was added to demulsify the beads. The beads were then washed three times in the wash buffer by centrifuging the tube at 7000 G for 1 minute and the supernatant discarded. Beads were then filtered using a 10 μ m filter to remove debris. Beads can be stored in the wash buffer at 4 °C indefinitely.

30

Dual-Barcode Assembly

- Beads then underwent split and mix to build up the barcodes. For the first split and mix step, a first tag binding region was added into several wells of a microtiter plate. Each well contained this first tag binding region with a certain number of binding sites (tag binding sequences) for identifiable tags so that each well contained first tag binding regions with only
- 35

that number of binding sites (tag binding sequences). Each first tag binding region was the same length. Beads were then equally split into wells so that tag binding region and beads were mixed to a total volume of around 20 μ L. Golden Gate assembly was carried out according to the NEBridge Golden Gate Assembly Kit (Bsal-HF v2, New England Biolabs) protocol. The tag binding regions had binding regions for identifiable tags, in increasing order e.g. well one has a tag binding region with one binding site, well two has a tag binding region with two binding sites, and so on. The tag binding regions also had a type IIS restriction site complementary to the linker. Golden Gate assembly attached each barcode to the linker, and millions of copies of the tag binding region were ligated to the bead. Beads were then mixed together, washed with Wash Buffer once, and equally split into wells with the second tag binding region type. Again, this second tag binding region type had binding regions for identifiable tags, but these were different from the first identifiable tags by virtue of the identifiable moiety used. Golden Gate assembly then attached this second tag binding region type to the first one. This split and mix procedure can be performed as many times as needed in as many wells as needed in order to build up a combinatorial library of barcodes.

The complexity of the library followed $N = L^C$, where N is the number of distinct dual barcodes, L was the number of different tag binding sequences per tag binding region type, and C was the number of different tag binding region types. As an example, for three tag binding regions each with up to 10 tag binding sequences, the number of distinct dual-barcodes was 10^3 or 1000. For 4 tag binding region and 40 tag binding sequences, the number was 40^4 or 2.56 million.

Each tag binding sequence and tag sequence (or the tag binding sequences of each tag binding region type) was designed with a proprietary algorithm to ensure that the tag sequence of each identifiable tag was both highly specific at its respective tag binding sequence, and highly unspecific to all other regions in the dual-barcode. In this way identifiable tags only bound to the tag binding sequences designated, lowering the chance of miss-annealing. The algorithm worked by looking at the hybridization ability or ΔG between two DNA strands. It used a Monte-Carlo method to generate millions of candidate tag sequences and then gating selected the tightest binders, with a low chance of self-hybridization and hairpin formation. Selected tag sequences were then simulated *in silico* to look at binding potential throughout the barcode.

Real-time identification of entities and/or barcoded supports

The following details the method for identifying fluorescently labelled dual-barcodes or genetic optical barcode sequencing (GOB-seq). Genetic mass barcode sequencing (i.e.

using defined mass identifiable moieties) or GMB-seq follows similar principles, however the detection method read droplet-content masses individually in a mass spectrometer.

Following generation of the barcoded support library (library of barcoded beads), barcoded beads were individually encapsulated in droplets and simultaneously illuminated by lasers emitting at wavelengths of 491 nm, 561 nm, and 647 nm to induce fluorescence. The fluorescence signals captured from the droplets were separated into three distinct channels, corresponding to each fluorophore, using an emission image splitter. This approach allowed for the simultaneous acquisition of multi-spectral fluorescence data from single droplets. The imaging system is detailed in ², but widefield illumination was used, instead of structured illumination.

Image processing was performed in Python to extract fluorescence measurements from the imaged droplets. Initially, a Gaussian blur algorithm was applied to mitigate the impact of image noise on subsequent segmentation. Following noise reduction, Otsu thresholding was applied to convert the blurred grayscale images into binary masks, which denote the presence of droplets against the background.

To refine these binary masks and correct for artefacts, a hole-filling algorithm was employed. Additionally, in order to resolve the masks of droplets in close proximity to each other, a morphological erosion step was carried out. This process facilitated the isolation and precise delineation of individual droplets.

For each object in the resulting binary mask detected by this procedure, edge detection was performed, and the coordinates along the edge were recorded. Elliptical fitting was then applied to these coordinates. A fitting error metric was employed to assess the accuracy of the ellipse representation. Objects with fitting errors exceeding a defined threshold, or those with major and minor axes outside a specified size range, were excluded to ensure the detected objects corresponded exclusively to singular droplets.

Gaussian blurring, Otsu thresholding, morphological erosion, object detection, edge detection, and ellipse fitting used existing implementations in the scikit-image Python library³. The hole-filling algorithm used the implementation from the SciPy Python library⁴.

For each droplet detected and localised in this way, the mean of all of its corresponding pixels was calculated. This allowed for the estimation of the concentration of any specific fluorophore in each droplet. Following the real-time measurement, beads were collected by de-emulsification and then the barcodes were amplified by PCR. Next generation sequencing was applied to the amplified barcodes. In this way, the DNA sequencing results revealed the code from the original split and mix. Barcodes could therefore be linked to the predicted colour that the tag binding regions would facilitate for each bead. Therefore, any

genetic material linked to the genetic colour code can be linked to the droplet as imaged in the real-time experiment ⁵.

Single cell sequencing was carried out simultaneously following the InDrops method in detailed protocol ^{1,6}.

5

Discussion

Figure 1 shows the main schematic of the structure of the dual-barcode attached to a polyacrylamide bead. Each bead contains potentially millions of identical barcode oligonucleotides which are covalently bonded through copolymerisation of acrylamide and an acrylated DNA linker molecule. The oligonucleotide had a region for binding the tag molecules, shown in Figure 1 as "GOB" or "b", and additionally had a region for a UMI (unique molecular identifier), and a cell marker (target region). The UMI acts as a unique identification code for each individual oligonucleotide and is thus distinct from all other UMI codes on other oligonucleotides attached to the bead. The cell marker acts as a region for attachment of cellular contents. Primer sites (amplification regions) are embedded within the DNA molecule to act as amplification points to extract the DNA from the bead for sequencing.

An algorithm was used to identify novel tag binding sequences (and corresponding tag sequences) according to a certain criteria. The tag binding sequences were chosen from a pool of billions of potential DNA oligonucleotides of the same length and are screened *in silico* for low homodimerization, low hairpin formation, low secondary structure formation, high melting temperature, and high heterodimerisation (to their complementary tag sequence). Figure 2 shows histograms and gating criteria used to select highly efficient binding oligonucleotides. Additionally, the tag binding sequences and respective tag sequences were chosen through comparing millions of the selected tag binding sequences to have a set where they bind the highest to their complementary tag sequence and the lowest to all other barcode regions (Figure 3). Additional regions, shown in Figure 4, such as the NBR (non-binding regions), linker regions, and primer regions were selected in a similar manner to prevent homodimerization and secondary structure formation. The tag binding sequences and their tag sequences were then screened against these other regions to identify regions in which they bind weakly; this was to prevent them from mis-annealing. The assembled tag binding regions including the tag binding sequences was then tested *in silico* to simulate points of binding. As shown in Figure 5, there were 3 binding sites for the tag sequences and one binding site for a constant tag. The constant tag was used for quality control. Figure 6 shows the simulated binding points for when there were 10 tag binding

sequences. Delta G was used as a proxy for binding affinity and each tag was compared in a rolling-window to all points on the total assembled barcode oligonucleotide.

Figure 7 shows the calibration curve for fluorophore with excitation wavelengths at 488, 561 and 657 nm. All had high R^2 values showing that the curves are linear within this range.

- 5 Figure 8 shows a falsely coloured widefield image with a population of beads tagged with Alexa 488, and a histogram of identified beads when the beads were loaded with 5 μ M of dual-barcodes. Figure 8 shows a similar widefield image and the beads clustered into groups, when the beads were loaded with 25 μ M of dual-barcodes. Figure 10 Shows violin plots of a small population of barcoded beads imaged using the SIM fluorescence
10 microscope, showing three populations.

The examples provided show a novel way of using a DNA both as a scaffold and as a means of coding information. By controlling the identifiable tags and tag binding regions and synthesizing the DNA in an ordered way, information can be gained in real-time, either optically or through mass measurements.

- 15 The reader's attention is directed to all papers and documents which are filed concurrently with or previous to this specification in connection with this application and which are open to public inspection with this specification, and the contents of all such papers and documents are incorporated herein by reference.

- 20 All of the features disclosed in this specification (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

- Each feature disclosed in this specification (including any accompanying claims, abstract and drawings), may be replaced by alternative features serving the same, equivalent, or
25 similar purpose, unless expressly stated otherwise. Thus, unless expressly stated otherwise, each feature disclosed is one example only of a generic series of equivalent or similar features.

- The invention is not restricted to the details of any foregoing embodiments. The invention extends to any novel one, or any novel combination, of the features disclosed in this
30 specification (including any accompanying claims, abstract and drawings), or to any novel one, or any novel combination, of the steps of any method or process so disclosed.

Sequences

Code:

"Tag1": +A+CG+CA+GC+C+A+A+AC+G+T (SEQ ID NO: 33)

"Tag2": +T+GTG+A+AGC+G+T+GC+G+T (SEQ ID NO: 34)

"Tag3" +T+TGC+T+TG+C+T+GGC+G+T (SEQ ID NO: 35)

Tag constant: +T+C+GC+T+G+T+TT+GCC+G+T (SEQ ID NO: 36)

- 5 The tags may also be the reverse complement such as:

"Tag1R" - +A+CG+TT+TG+G+C+T+GC+G+T (SEQ ID NO: 37)

"Tag2R" - +A+CGC+ACG+C+T+TC+A+C+A (SEQ ID NO: 38)

"Tag3R" - +A+CGC+C+AG+C+A+AGC+A+A (SEQ ID NO: 39)

10

An advantage of using a complement tag are that the tag is binding to a strand of DNA that is covalently attached to the bead, and therefore it is more stably attached to the polyacrylamide bead.

- 15 + is for a locked nucleic acid e.g. +A means locked nucleic acid A. The constant tag is for quality control as described and simply allows checking that the beads (i.e., a support) have consistent amounts of DNA copolymerised. Different sequences can be generated using the algorithm described herein.

NAME	SEQUENCE	SEQ ID NO
Linker 1	GTACTCATCATTAAATCATCAGGTGAGTGGTAGAGGAACCTCGCTGTTGCCGTTTCCAGAGACCAGTCT	1
Linker 2	GTACTCATCATTAAATCATCAGACTCACATTCAACAATAGAAAGAAAGTCATAGTAC	2
tag 1_first	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAGCCAAACGTCAGTCCTTTTCAATGTAACCTATGATAGGGAGTATTTGTTTCATAGAAATGACATATTGCATTCTCTCTCTCTCAAAGTACATAACATAACATTTTCGATGAGTACTAGACTAATTCAACAACATGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCTTATAATTAAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGACCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATACAATAGATACCTCTACCTTATGTCATTAAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTGAGGCCTGCATGCAAGCTTGGCGTA	3
tag 1_sec	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAGCCAAACGTCAGTCCTTTTCAATGTAACCTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTATCCATTCTCTCTCTCAAAGTACATAACATAACATTTTCGATGAGTACTAGACTAATTCAACAACAT	4

on d	TGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCTT ATAATTAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGACC CCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCTC TACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTGAGGCCT GCATGCAAGCTTGCGTA	
tag 1_t hir d	AGCAGATTGTAAGTACTGAGAGTGACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACCTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGATGAGTACTAGACTAATTCAACAAC ATTGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACC TTATAATTAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGA CCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACC TCTACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTGAGGC CTGCATGCAAGCTTGCGTA	5
tag 1_f ou rth	AGCAGATTGTAAGTACTGAGAGTGACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACCTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGCAGCCAAACGTTGTTCAACTAA GTAAAGACTGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACC TTATAATTAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGA CCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACC TCTACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTGAGGC CTGCATGCAAGCTTGCGTA	6
tag 1_f ifth	AGCAGATTGTAAGTACTGAGAGTGACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACCTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGCAGCCAAACGTTGTTCAACTAA GTAAAGACTACGCAGCCAAACGTTGAGTTGATGTACATACTATCAGAAAGGTCATTCTACTAGA CCTTATAATTAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAG GACCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGAT ACCTCTACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTGA GGCCTGCATGCAAGCTTGCGTA	7
tag 1_s ixt h	AGCAGATTGTAAGTACTGAGAGTGACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACCTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGCAGCCAAACGTTGTTCAACTAA GTAAAGACTACGCAGCCAAACGTTGAGTTGATGTACATACTATCAACGCAGCCAAACGTTGTCAA CAACATAATCATCTCATGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAG ACCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATAC CTCTACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTGAGG CCTGCATGCAAGCTTGCGTA	8
tag 1_s ev ent h	AGCAGATTGTAAGTACTGAGAGTGACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACCTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGCAGCCAAACGTTGTTCAACTAA GTAAAGACTACGCAGCCAAACGTTGAGTTGATGTACATACTATCAACGCAGCCAAACGTTGTCAA CAACATAATCATCTCATACGCAGCCAAACGTTCTATGAGTAATGACATCATTATGATAGGTAG GACCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGAT ACCTCTACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTGA GGCCTGCATGCAAGCTTGCGTA	9
tag 1_ eig hth	AGCAGATTGTAAGTACTGAGAGTGACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACCTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGCAGCCAAACGTTGTTCAACTAA GTAAAGACTACGCAGCCAAACGTTGAGTTGATGTACATACTATCAACGCAGCCAAACGTTGTCAA CAACATAATCATCTCATACGCAGCCAAACGTTCTATGAGTAATGACATCATTACGCAGCCAAAC GTGAGAGAGATTATGAAAATTAGGTGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGA	1 0

	TACCTCTACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCTG AGGCCTGCATGCAAGCTTGCGTA	
tag 1_ nin th	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGCAGCCAAACGTTGTTCAACTAA GTAAAGACTACGCAGCCAAACGTTGAGTTGATGTACATACTATCAACGCAGCCAAACGTTGTCAA CAACATAATCATCTCATACGCAGCCAAACGTCCTATGAGTAATGACATCATTACGCAGCCAAAC GTGAGAGAGATTATGAAAATTAGGACGCAGCCAAACGTTATATTCTCATAATGAGAGTACCATAG ATACCTCTACCTTATGTCATTAACCTTAGTACGCATACCCTCAATGTCCCTCCTAGAGACCAGTCT GAGGCCTGCATGCAAGCTTGCGTA	1 1
tag 1_t ent h	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATTCCGACACCTCCTCACAATCCACGCAG CCAAACGTCAGTCCTTTTCAATGTAACTAACGCAGCCAAACGTTGAAATGTTGAGAGAATGTAT CACGCAGCCAAACGTAATGAGAGTACATCAACATAGTACGCAGCCAAACGTTGTTCAACTAA GTAAAGACTACGCAGCCAAACGTTGAGTTGATGTACATACTATCAACGCAGCCAAACGTTGTCAA CAACATAATCATCTCATACGCAGCCAAACGTCCTATGAGTAATGACATCATTACGCAGCCAAAC GTGAGAGAGATTATGAAAATTAGGACGCAGCCAAACGTTATATTCTCATAATGAGAGTACCACGC AGCCAAACGTTGAGATGAGTAGTTGAGATATCGCATACCCTCAATGTCCCTCCTAGAGACCAGTC TGAGGCCTGCATGCAAGCTTGCGTA	1 2
tag 2_f irst	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACTATGATAGGGAGTATTTGTTGATAGAAATGACATATT GCATTCTCTCTCTCAAAGTACATAACATAACATTTGATGAGTACTAGACTAATTCAACAACAT TGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCTT ATAATTAAGTGGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGACC CCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCTC TACCTTATGTCATTAACCTTAGTACGTTACTCTCCACTACCCCTTAGAGACCAGTCTGAGGCCT GCATGCAAGCTTGCGTA	1 3
tag 2_s ec on d	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACTATGTAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CCATTCTCTCTCTCAAAGTACATAACATAACATTTGATGAGTACTAGACTAATTCAACAACAT TGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCTT ATAATTAAGTGGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGACC CCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCTC TACCTTATGTCATTAACCTTAGTACGTTACTCTCCACTACCCCTTAGAGACCAGTCTGAGGCCT GCATGCAAGCTTGCGTA	1 4
tag 2_t hir d	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACTATGTAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTCGATGAGTACTAGACTAATTCAACAAC ATTGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACC TTATAATTAAGTGGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGA CCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACC TCTACCTTATGTCATTAACCTTAGTACGTTACTCTCCACTACCCCTTAGAGACCAGTCTGAGGC CTGCATGCAAGCTTGCGTA	1 5
tag 2_f ou rth	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACTATGTAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTGTGAAGCGTGCGTTGTTCAACTAA GTAAAGACTGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACC TTATAATTAAGTGGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGA CCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACC TCTACCTTATGTCATTAACCTTAGTACGTTACTCTCCACTACCCCTTAGAGACCAGTCTGAGGC CTGCATGCAAGCTTGCGTA	1 6

tag 2_f ifth	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACCTATGTGAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTGTGAAGCGTGCGTGTGTTCAACTAA GTAAAGACTTGTGAAGCGTGCGTTGAGTTGATGTACATACTATCAGAAAGGTCATTCTACTAGA CCTTATAATTAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTA GACCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATAACAATAGAT ACCTCTACCTTATGTCATTAACTTAGTACGTTACTCTCCCACTCACCCCTTAGAGACCAGTCTGA GGCCTGCATGCAAGCTTGCGTA	1 7
tag 2_s ixt h	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACCTATGTGAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTGTGAAGCGTGCGTGTGTTCAACTAA GTAAAGACTTGTGAAGCGTGCGTTGAGTTGATGTACATACTATCATGTGAAGCGTGCGTGTCAA CAACATAATCATCTCATGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAG ACCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATAACAATAGATAC CTCTACCTTATGTCATTAACTTAGTACGTTACTCTCCCACTCACCCCTTAGAGACCAGTCTGAGG CCTGCATGCAAGCTTGCGTA	1 8
tag 2_s ev ent h	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACCTATGTGAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTGTGAAGCGTGCGTGTGTTCAACTAA GTAAAGACTTGTGAAGCGTGCGTTGAGTTGATGTACATACTATCATGTGAAGCGTGCGTGTCAA CAACATAATCATCTCATTGTGAAGCGTGCGTCCTATGAGTAATGACATCATTCTATGATAGGTA GACCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATAACAATAGAT ACCTCTACCTTATGTCATTAACTTAGTACGTTACTCTCCCACTCACCCCTTAGAGACCAGTCTGA GGCCTGCATGCAAGCTTGCGTA	1 9
tag 2_ eig hth	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACCTATGTGAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTGTGAAGCGTGCGTGTGTTCAACTAA GTAAAGACTTGTGAAGCGTGCGTTGAGTTGATGTACATACTATCATGTGAAGCGTGCGTGTCAA CAACATAATCATCTCATTGTGAAGCGTGCGTCCTATGAGTAATGACATCATTCTGTGAAGCGTGC GTGAGAGAGATTATGAAAATTAGGTGAGATCTCTCATACCTTTGTCTATCTATATATAACAATAGA TACCTCTACCTTATGTCATTAACTTAGTACGTTACTCTCCCACTCACCCCTTAGAGACCAGTCTGA GGCCTGCATGCAAGCTTGCGTA	2 0
tag 2_ nin th	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACCTATGTGAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTGTGAAGCGTGCGTGTGTTCAACTAA GTAAAGACTTGTGAAGCGTGCGTTGAGTTGATGTACATACTATCATGTGAAGCGTGCGTGTCAA CAACATAATCATCTCATTGTGAAGCGTGCGTCCTATGAGTAATGACATCATTCTGTGAAGCGTGC GTGAGAGAGATTATGAAAATTAGGTGTGAAGCGTGCGTATATTCTCATAATGAGAGTACCATAG ATACCTCTACCTTATGTCATTAACTTAGTACGTTACTCTCCCACTCACCCCTTAGAGACCAGTCTG AGGCCTGCATGCAAGCTTGCGTA	2 1
tag 2_t ent h	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCATCCTCAAATGACTCCCTACCCCTGTGAA GCGTGCGTCAGTCCTTTTCAATGTAACCTATGTGAAGCGTGCGTTGAAATGTTGAGAGAATGTAT CTGTGAAGCGTGCGTAATGAGAGTACATCAACATAGTTGTGAAGCGTGCGTGTGTTCAACTAA GTAAAGACTTGTGAAGCGTGCGTTGAGTTGATGTACATACTATCATGTGAAGCGTGCGTGTCAA CAACATAATCATCTCATTGTGAAGCGTGCGTCCTATGAGTAATGACATCATTCTGTGAAGCGTGC GTGAGAGAGATTATGAAAATTAGGTGTGAAGCGTGCGTATATTCTCATAATGAGAGTACCATAG AAGCGTGCGTTGAGATGAGTAGTTGAGATATCGTTACTCTCCCACTCACCCCTTAGAGACCAGTC TGAGGCCTGCATGCAAGCTTGCGTA	2 2
tag 3_f irst	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATGATAGGGAGTATTTGTTCATAGAAATGACATATT GCATTCTCTCTCTCAAAGTACATAACATAACATTCGATGAGTACTAGACTAATTCAACAACAT	2 3

	TGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCTT ATAATTTAAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGACC CCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCTC TACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAGCT TGGCGTA	
tag 3_s ec on d	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CCATTCTCTCTCTCAAAGTACATAACATAACATTTGATGAGTACTAGACTAATTCAACAACAT TGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCTT ATAATTTAAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGACC CCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCTC TACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAGCT TGGCGTA	2 4
tag 3_t hir d	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTCGATGAGTACTAGACTAATTCAACAACA TTGTACATGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCT TATAATTTAAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGAC CCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCT CTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAGC TTGGCGTA	2 5
tag 3_f ou rth	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTTTGCTTGCTGGCGTGTTGTTCAACTAAG TAAAGACTGTTGTCTATTACATTAGTCTTTGAGTTACTTTAGGTGAAAGGTCATTCTACTAGACCT TATAATTTAAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGAC CCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCT CTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAGC TTGGCGTA	2 6
tag 3_f ifth	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTTTGCTTGCTGGCGTGTTGTTCAACTAAG TAAAGACTTTGCTTGCTGGCGTTGAGTTGATGTACATACTATCAGAAAGGTCATTCTACTAGACC TTATAATTTAAAGTCGGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGA CCCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACC TCTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAG CTTGCGTA	2 7
tag 3_s ixt h	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTTTGCTTGCTGGCGTGTTGTTCAACTAAG TAAAGACTTTGCTTGCTGGCGTTGAGTTGATGTACATACTATCATTGCTTGCTGGCGTGTAACA ACATAATCATCTCATGAATGTCTTTGTTAGTTCTAAAATTGAGTAGTATCCATATGATAGGTAGAC CCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCT CTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAGC TTGGCGTA	2 8
tag 3_s ev ent h	AGCAGATTGACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTTTGCTTGCTGGCGTGTTGTTCAACTAAG TAAAGACTTTGCTTGCTGGCGTTGAGTTGATGTACATACTATCATTGCTTGCTGGCGTGTAACA ACATAATCATCTCATTTGCTTGCTGGCGTCTATGAGTAATGACATCATTATCATATGATAGGTAGAC CCCTATGAGTTAGTCTACCTATGAGATCTCTCATACCTTTGTCTATCTATATATACAATAGATACCT CTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAGC	2 9

	CTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAAGC TTGGCGTA	
tag 3_ eig hth	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTTTGCTTGCTGGCGTGTTGTTCAACTAAG TAAAGACTTTGCTTGCTGGCGTTGAGTTGATGTACATACTATCATTGCTTGCTGGCGTGTCACCA ACATAATCATCTCATTGCTTGCTGGCGTCCTATGAGTAATGACATCATTCTTGCTTGCTGGCGTG AGAGAGATTATGAAAATTAGGTGAGATCTCTCATACCTTTGCTATCTATATATACAATAGATAC CTCTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAA GCTTGCGTA	3 0
tag 3_ nin th	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTTTGCTTGCTGGCGTGTTGTTCAACTAAG TAAAGACTTTGCTTGCTGGCGTTGAGTTGATGTACATACTATCATTGCTTGCTGGCGTGTCACCA ACATAATCATCTCATTGCTTGCTGGCGTCCTATGAGTAATGACATCATTCTTGCTTGCTGGCGTG AGAGAGATTATGAAAATTAGGTGAGATCTCTCATACCTTTGCTATCTATATATACAATAGATAC CTCTACCTTATGTCATTAACCTTAGTACCACCTATGACCACTCACCAGTCTGAGGCCTGCATGCAA GCTTGCGTA	3 1
tag 3_ t ent h	AGCAGATTGTAAGTACTGAGAGTGCACCAAGACTGGTCTCACCTTGACCTCAAAAGTTGACCGTTGCTT GCTGGCGTCAGTCCTTTTCAATGTAACCTATTGCTTGCTGGCGTTGAAATGTTGAGAGAATGTAT CTTGCTTGCTGGCGTAATGAGAGTACATCAACATAGTTTGCTTGCTGGCGTGTTGTTCAACTAAG TAAAGACTTTGCTTGCTGGCGTTGAGTTGATGTACATACTATCATTGCTTGCTGGCGTGTCACCA ACATAATCATCTCATTGCTTGCTGGCGTCCTATGAGTAATGACATCATTCTTGCTTGCTGGCGTG AGAGAGATTATGAAAATTAGGTGAGATCTCTCATACCTTTGCTATCTATATATACAATAGATAC TGGCGTTGAGATGAGTAGTTGAGATATCCACCTATGACCACTCACCAGTCTGAGGCCTGCATGC AAGCTTGCGTA	3 2

References

- (1) Zilionis, R.; Nainys, J.; Veres, A.; Savova, V.; Zemmour, D.; Klein, A. M.; Mazutis, L.
5 Single-Cell Barcoding and Sequencing Using Droplet Microfluidics. *Nat Protoc* 2017, 12 (1), 44–73. <https://doi.org/10.1038/nprot.2016.154>.
- (2) Ward, E. N.; Hecker, L.; Christensen, C. N.; Lamb, J. R.; Lu, M.; Mascheroni, L.; Chung, C. W.; Wang, A.; Rowlands, C. J.; Schierle, G. S. K.; Kaminski, C. F. Machine Learning Assisted Interferometric Structured Illumination Microscopy for Dynamic Biological Imaging.
10 *Nat. Commun.* 2022, 13 (1), 7836. <https://doi.org/10.1038/s41467-022-35307-0>.
- (3) Walt, S. van der; Schönberger, J. L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J. D.; Yager, N.; Gouillart, E.; Yu, T.; contributors, scikit-image. Scikit-Image: Image Processing in Python. *PeerJ* 2014, 2, e453. <https://doi.org/10.7717/peerj.453>.
- (4) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.;
15 Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; Walt, S. J. van der; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.;

- Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; Mulbregt, P. van; Contributors, S. 1 0; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.;
- 5 Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; Cardoso, J. V. de M.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.;
- 10 Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* 2020, 17 (3), 261–272.
- 15 <https://doi.org/10.1038/s41592-019-0686-2>.
- (5) Gantz, M.; Neun, S.; Medcalf, E. J.; Vliet, L. D. van; Hollfelder, F. Ultrahigh-Throughput Enzyme Engineering and Discovery in In Vitro Compartments. *Chem. Rev.* 2023, 123 (9), 5571–5611. <https://doi.org/10.1021/acs.chemrev.2c00910>.
- (6) Salmen, F.; Jonghe, J. D.; Kaminski, T. S.; Alemany, A.; Parada, G. E.; Verity-Legg, J.;
- 20 Yanagida, A.; Kohler, T. N.; Battich, N.; Brekel, F. van den; Ellermann, A. L.; Arias, A. M.; Nichols, J.; Hemberg, M.; Hollfelder, F.; Oudenaarden, A. van. High-Throughput Total RNA Sequencing in Single Cells Using VASA-Seq. *Nat. Biotechnol.* 2022, 40 (12), 1780–1793. <https://doi.org/10.1038/s41587-022-01361-8>.

Claims

1. A nucleic acid encoded dual modality barcode (dual-barcode) for linking real time identification and subsequent analysis of one or more entities comprising:
 - a. one or more distinct tag binding regions each comprising one or more tag binding sequences for binding to a cognate identification tag;
wherein each tag binding region comprises a linker region for linking to an adjacent tag binding region or a support; and
wherein the tag binding sequences of each tag binding region are distinct for binding to distinct identification tags; and
 - b. at least one amplification region for amplification of the nucleic acid molecule.
2. The dual-barcode of claim 1, wherein the nucleic acid molecule comprises C tag binding regions, wherein C is the number of distinct tag binding regions.
3. The dual-barcode of claims 1 or 2, wherein each tag binding region comprises L tag binding sequences, wherein L is the number tag binding sequences of each distinct tag binding region and wherein each tag binding sequence in each individual tag binding region is for binding the same identification tag.
4. The dual-barcode of claim 3, wherein the number of distinct nucleic acid molecules (N) is equal to L^C .
5. The dual-barcode of any of claims 1 to 4, wherein each one of the tag binding regions is bound to one or more cognate identification tags and wherein each cognate identification tag comprises;
 - a tag sequence that specifically binds to its cognate tag binding sequence; and
 - an identifiable moiety.
6. The dual-barcode of claim 4 or 5, wherein the one or more cognate identification tags bound to each distinct tag binding region comprises distinct identifiable moieties.
7. The dual-barcode any of claims 4 to 6, wherein each tag binding region is bound to two or more of its cognate identification tag.

8. The dual-barcode any of claims 4 to 7, wherein the identifiable moiety comprises an optical moiety and/or a defined mass moiety.
9. The dual-barcode of any preceding claim, wherein each tag binding region comprises the same length.
10. The dual-barcode of any preceding claim, further comprising a unique molecule identifier sequence (UMI).
11. The dual-barcode of any preceding claim, further comprising a target region.
12. The dual-barcode of claim 11, wherein the target region:
 - a. is for binding to a cellular target;
 - b. comprises a sequence encoding a gene of interest; or
 - c. is for binding to a chemical moiety.
13. The dual-barcode of any preceding claim, wherein the nucleic acid molecule is attached to a support.
14. The dual-barcode of claim 13, wherein the support comprises a bead.
15. The dual-barcode of claim 13, wherein the support comprises a DNA-encoded library (DEL) member.
16. The dual-barcode of claim 13, wherein the support comprises a protein.
17. The dual-barcode of claim 13, wherein the support comprises a cell; optionally wherein the support is comprised within the cell and/or on the cell surface.
18. The dual-barcode of any of claim 13 to 17, wherein the support comprises a chemical agent.
19. A barcoded support for linking real time identification and subsequent analysis of one or more entities comprising:
 - at least one dual-barcode according to any of claims 1 to 4 and 9 to 11; and

a support comprising one or more attachment moieties for attachment of the dual-barcode to the support.

20. The barcoded support of claim 19, wherein each one of the tag binding regions is bound to one or more of its cognate identification tags according to any of claims 5 to 8.
21. The barcoded support of claims 19 or 20, wherein the support comprises a bead, a cell, a protein, a nucleic acid or a chemical agent.
22. The barcoded support any of claims 19 to 21, wherein the dual-barcode is attached to the bead via a linker region of the dual-barcode.
23. The barcoded support of claims 19 to 22, wherein the one or more attachment moieties comprises a nucleic acid.
24. The barcoded support according to claim 23, wherein the one or more attachment moieties are integral to the support.
25. The barcoded support according to any of claims 19 to 24, wherein each dual-barcode comprises identical tag binding regions.
26. The barcoded support according to any of claims 19 to 25, wherein each one of the at least one dual-barcodes comprises a distinct UMI.
27. A plurality of barcoded supports according to any of claims 19 to 26, wherein each support of the plurality comprises dual-barcodes comprising distinct target regions bound thereto to the each other support of the plurality.
28. A method of producing a plurality of barcoded supports according to claim 27; the method comprising:
 - a. providing plurality of first tag binding regions for binding to a cognate first identification tag and comprising at least one linking region for linking to a support and/or adjacent tag binding regions;
wherein one of the plurality of first tag binding regions comprises L tag binding sequences for binding to L cognate first identification tags; and

wherein each other one of the plurality of tag binding regions comprises $L+x$ tag binding sequences wherein x is an integer of 1 or more and x is not repeated for the plurality of tag binding regions;

thereby providing a plurality of first tag binding regions comprising a distinct number of tag binding sequences;

- b. attaching in separate vessels at least one of each of the plurality of first tag binding regions comprising a distinct number of tag binding sequences to a support;

thereby providing a plurality of supports each attached to at least one first binding region comprising a distinct number of tag binding sequences (Support-L1x);

29. The method of claim 28, further comprising:

- c. combining the plurality of Support-L1x and separating a plurality of portions of the plurality of Support-L1x into a plurality of separate vessels;
- d. adding to each separate vessel a plurality of a second tag binding regions for binding a second cognate identification tag;

wherein one of the plurality of the second cognate tag binding region comprises L second tag binding sequences for binding to L cognate second identification tags; and

wherein each other one of the plurality second tag binding regions comprises $L+x$ second tag binding sequences wherein x is an integer of 1 or more and x is not repeated for the plurality of second tag binding regions;

thereby providing a plurality of second tag binding regions each comprising a distinct number of second tag binding sequences ($L2x$);

- e. attaching each of the $L2x$ to the first tag binding region;

thereby providing a plurality of supports each attached to at least one first binding region comprising a distinct number of tag binding sequences and a second tag binding region comprising a distinct number of second tag binding sequences (Support-L1x-L2x).

30. A method of producing a plurality of barcoded supports according to claim 27; the method comprising:

- a. providing at least one first tag binding region (TBRa) for binding to a cognate first identification tag and at least one second tag binding region (TBRb) for binding to a cognate second identification tag;

wherein each of the first tag binding region comprises L tag binding sequences for binding to L cognate first identification tags and at least one linking region for linking to a support and/or adjacent tag binding regions; and

wherein the second tag binding region comprise Lb tag binding sequences for binding to Lb cognate second identification tags and at least one linking region for linking to adjacent tag binding regions;

- b. linking the second tag binding region (TBRa) to the first tag binding region (Support-TBRa(L) – TBRb(Lb))

thereby producing a dual-barcode;

- c. repeating steps (a) to (c) with a plurality of distinct tag binding regions to a produce a plurality of distinct dual-barcodes;
- d. attaching each distinct one of the plurality of distinct dual-barcodes to a different support or;

wherein step (a) comprises attaching the first binding region to a support thereby providing a plurality of at least one dual-barcodes each attached to a support.

31. The method of claim 30 further comprising, before step (d):

- f. linking one or more further tag binding regions (TBRy) to the second tag binding region wherein each of the further tag binding region comprises Ly tag binding sequences for binding to Ly cognate further identification tags and at least one linking region for linking to adjacent tag binding regions.

32. The method of claim 31, wherein step (f) is repeated for a plurality of further tag binding regions wherein each of the further tag binding region comprises L+x, Lb+x or Ly+x tag binding sequences wherein x is an integer of 1 or more and x is not repeated (Support-TBR1a(L) – TBRb (L) - TBRy(Lyx))

33. The method of claim 31 or 32, wherein at least one further tag binding region comprises a further first tag binding region or a further second tag binding region or wherein at least one further tag binding region comprises a further tag binding region with Lx tag binding sequences (Support-TBRa(L) – TBRb (Lb) – TBRa(Lx) – TBRb(Lbx) or Support-TBRa(L) – TBRb (Lb) – TBRy(Ly) – TBRy(Lyx)).

34. The method of claim 29 or 30 to 33, wherein steps d and e of claim 29 or step (c) of claim 30 or step (f) of claim 32 are repeated C times, wherein C is the number of distinct tag binding regions thereby producing N distinct barcoded supports, wherein $N = L^C$.

35. The method of claim 34, further comprising binding C distinct cognate identification tags to each tag binding sequence of each distinct tag binding region;
wherein the identification tags are according to any one of claims 5 to 8.
36. The method of any of claims 28 to 35, wherein the method further comprises, attaching a distinct UMI to each one of the plurality of dual-barcodes attached to each support.
37. The method of any of claims 28 to 36, wherein the method further comprises, attaching a distinct target region to at least one of the plurality of dual-barcodes, wherein each dual-barcode comprises at least one target region distinct from other ones of the plurality of dual-barcodes attached to other supports.
38. A method of linking real time identity of an entity to subsequent analysis the method comprising:
- a. providing a plurality of barcoded supports according to any of claims 20 to 26;
 - b. combining one of the plurality of barcoded supports with one entity;
 - c. detecting the identification tag of the barcoded support; and
 - d. optionally amplifying each dual-barcode of the plurality of barcoded supports and;
 - e. further optionally sequencing the amplified dual-barcodes.
39. The method of claim 38, wherein the identification tag comprises an optical moiety and detection comprises detecting one or more of optical properties.
40. The method of claim 38, wherein the identification tag comprises a defined mass moiety and detection comprises detecting at least a mass of each dual-barcode.
41. The method of any of claims 38 to 40, wherein the method comprises a method of RNA detection and the dual-barcode comprises a target region for binding to RNA.
42. The method of claim 39 or claim 41, wherein the optical moiety comprises a fluorophore and detection comprises:
- a. simultaneously illuminating two or more of the fluorophores of distinct identification tags;

- b. capturing fluorescence signals generated from the two or more fluorophores and separating each signal into distinct channels, each channel corresponding to each fluorophore; thereby allowing for simultaneous capture of multi-spectral fluorescence data from single entities.
43. The method of claim 42, further comprising extracting fluorescence measurements from the captured fluorescence signals, extracting comprising:
- a. simultaneously receiving signals including information corresponding to the fluorescence measurements;
 - b. processing the signals; the processing comprising:
 - i. Gaussian blurring the signals to reduce noise;
 - ii. thresholding the Gaussian blurred signals to produce binary masks which denote the presence of entities against a background;
 - iii. refining the binary masks and correcting for artefacts using hole-filling;
 - iv. optionally; wherein the refined binary masks comprises two or more entities in close proximity to each other, resolving the two or more entities using morphological erosion of the binary masks to isolate and delineate individual entities;
 - v. for each entity in the binary mask, edge detecting coordinates along each edge of each entity detected; and
 - vi. applying elliptical fitting to the coordinates.
44. The method of claim 43, further comprising for each entity detected:
- a. calculating the mean of each data point corresponding to each detected entity to estimate the measurement of each fluorophore in each entity.
45. The method of claims 43 or 44, further comprising:
- before calculating the mean, employing a fitting error metric for assessing accuracy of the elliptical fitting; and
 - excluding those detected entities exceeding a defined threshold of the fitting error metric thereby excluding detected entities that do not correspond to a single entity.
46. The method according to any one of claims, wherein the sequence of the dual-barcode of each of plurality of barcoded supports is known prior to step (a) or (b) and the method comprises pairing the known sequences to the detected identification tag.

47. Use of a dual-barcode according to any of claims 4 to 18, barcoded support according to any of claims 20 to 26 or a plurality thereof in a method of:
- a. single cell analyses;
 - b. genetic material analysis;
 - c. epigenetic analysis;
 - d. chemical library analysis;
 - e. peptide library analysis;
 - f. nucleic acid library analysis;
 - g. metabolic library analysis;
 - h. secretome analysis;
 - i. reaction condition analysis;
 - j. reaction analysis;
 - k. synthetic biology;
 - l. therapeutic agent analysis;
 - m. enzymology;
 - n. drug interaction analysis;
 - o. sample analysis;
 - p. pathogen analysis;
 - q. pathology;
 - r. cytology;
 - s. DNA encoded library analysis;
 - t. gene engineering;
 - u. cell engineering; or
 - v. in vitro transcription and/or in vitro transcription and translation.
48. A kit of parts comprising:
- a. at least one dual-barcode according to any one of claims 1 to 4 and 9 to 11;
and
 - b. at least one cognate identification tag according to any one of claims 5 to 8.
49. The kit according to claim 48, further comprising one or more further tag binding regions each for binding further cognate identification tags; and optionally one or more cognate further identification tags.
50. The kit according to claim 48 or 49, further comprising a support comprising one or more attachment moieties for attachment of the dual-barcode to the support.

51. The kit according to any of claims 48 to 50, further comprising one or more reagents for attaching the at least one dual-barcode to the support.
52. The kit according to any of claims 49 to 51, further comprising one or more reagents for linking the at least one dual-barcode to one or more further tag binding regions; and optionally for linking two or more further tag binding regions.
53. A method for extracting individual measurements from a plurality of simultaneously captured measurements of a plurality of entities comprising a dual-barcode; the method comprising:
- a. simultaneously receiving, from at least one sensor, signals including information corresponding to the plurality of measurements;
 - b. processing the signals; the processing comprising:
 - i. Gaussian blurring the signals to reduce noise;
 - ii. thresholding the Gaussian blurred signals to produce binary masks which denote the presence of one or more of the plurality of barcoded entity against a background;
 - iii. refining the binary masks and correcting for artefacts using hole-filling;
 - iv. optionally; wherein the refined binary masks comprises two or more entities in close proximity to each other, resolving the two or more entities using morphological erosion of the binary masks to isolate and delineate individual entities;
 - v. for each entity detected in the binary mask, edge detecting coordinates along each edge of each entity detected; and
 - vi. applying elliptical fitting to the detected coordinates; and
 - vii. for each entity detected, calculating the mean of each data point corresponding to each detected entity to estimate the measurement in each entity,thereby extracting measurements for all entities detected; and
 - c. outputting one or more signals indicative of the extracted measurements.
54. The method of claim 53, further comprising:
- before calculating the mean, employing a fitting error metric for assessing accuracy of the elliptical fitting; and

excluding those detected entities exceeding a defined threshold of the fitting error metric thereby excluding detected entities that do not correspond to a single entity.

55. The method of claim 53 or 54, wherein the data is analysed using machine learning.
56. The method of claims 53 to 55 or the method steps of claims 38 (C) are implemented using processing circuitry.
57. A method for designing a tag sequence for specifically binding to a tag binding sequence for a dual-barcode according to any one of claims 1 to 18, the method comprising:
- a. receiving data comprising information corresponding to a nucleic acid sequence, and processing the data by:
 - i. generating one or more tag sequences of a defined length picking randomly between the nucleotides A, T, C and G wherein each tag sequence comprises a GC content between 30-60%;
 - ii. assessing each tag sequence for (a) self hybridisation, (b) hairpin formation, and (c) secondary structure formation using delta G as an indication of a likelihood of (a) to (c) occurring;
 - iii. selecting the tag sequences with a likelihood of (a) to (c) occurring below a threshold;
 - iv. assessing each tag sequences for binding to a complementary sequence using delta G as an indication of binding affinity;
 - v. selecting the tag sequences with a binding affinity to a complementary sequence greater than a threshold;
 - vi. assessing each of the selected tag sequences for binding affinity to each other selected tag sequences using delta G as an indication of binding affinity;
 - vii. selecting the tag sequences that have a binding affinity to other selected tag sequences lower than a threshold and a binding affinity to a complementary sequence greater than a threshold; and
 - viii. assessing the tag sequences for binding to the complementary sequence of other tag sequences using delta G as an indication of binding affinity and selecting those tag sequences that specifically bind their complementary sequence and do not bind to the complementary sequence of other tag sequences.

58. The method of claim 57, further comprising assessing the binding affinity of each tag sequence selected in steps (c), (e), (g) or (f) to one or more further nucleic acid sequences using delta G as an indication of binding affinity and selecting tag sequences that have a binding affinity for the one or more further nucleic acid sequences below a threshold.
59. A library of genetically encoded barcodes comprising a plurality of dual-barcodes according to any one of claims 1 to 12, wherein each one of the plurality is a distinct dual-barcode.
60. A method of producing a library of dual-barcodes, the method comprising:
- a. designing one or more tag sequences using a method according to claims 57 or 59;
 - b. producing a plurality of dual-barcodes according to anyone of claims 1 to 4 and 9 to 12,
wherein each tag binding sequence comprises a complementary nucleic acid sequence to each of the designed tag sequences; and
wherein each of the dual-barcodes of the plurality are distinct.
61. The method of claim 60, wherein producing comprises in vitro synthesis of each distinct dual-barcode or recombinant production of each distinct dual-barcode.
62. The method of claim 61, wherein the target region of each distinct dual-barcode comprises a sequence of interest, wherein each sequence of interest is distinct for each distinct dual-barcode;
- thereby providing a plurality of distinct dual-barcodes each comprising a distinct sequence of interest and wherein the sequence of each target binding region of distinct dual-barcodes and its distinct sequence of interest are known.
63. The method of claim 62 wherein the sequence of interest comprises:
- a nucleic acid sequence encoding a gene of interest; or
 - a nucleic acid sequence encoding a protein of interest.

64. The method of claim 60 to 63, wherein the method comprises attaching one or more of each distinct dual-barcode to a different support.
65. The method of claim 60, wherein the method comprises attaching one or more of each distinct dual-barcodes to a different support; and
producing comprises a method according to any one of claims 28 to 37.