

CLAIMS

[Claim 1]. A method for identifying a subpopulation of cells comprising a biological feature, the method comprising:

- (a) providing a single-cell sequencing library,
 - wherein the sequencing library comprises a plurality of modified target nucleic acids,
 - wherein the modified target nucleic acids comprise at least one index sequence;
- (b) interrogating the sequencing library by targeted sequencing to identify the index sequences that are present on the same modified target nucleic acid as a biological feature,
 - wherein the index sequences associated with the biological feature are marker index sequences, optionally wherein more than one marker index sequence is identified;
- (c) altering the sequencing library to obtain a sub-library,
 - wherein the sub-library comprises increased representation of the modified target nucleic acids comprising the marker index sequences in comparison to other modified target nucleic acids present in the sequencing library that do not comprise a marker index sequence;
- (d) determining the nucleotide sequence of the modified target nucleic acids comprising a marker index sequence.

[Claim 2]. The method of claim 1, wherein the single-cell sequencing library comprises nucleic acids from multiple samples optionally, wherein the multiple samples comprise (i) samples of the same tissue obtained from different organisms, (ii) samples of different tissues from one organism, or (iii) samples of different tissues from different organisms.

[Claim 3]. The method of claim 1, wherein the single-cell combinatorial sequencing library comprises target nucleic acids representative of the whole genome of the cells or nuclei or a subset of the genome optionally, wherein the subset of the genome comprises target nucleic acids representative of transcriptome, accessible chromatin, DNA, conformational state, or proteins of the cells or nuclei.

[Claim 4]. The method of any one of claims 1-3, wherein the altering comprises enrichment of the modified target nucleic acids comprising the marker index sequences optionally, wherein the enriching comprises a hybridization-based method preferably, wherein the hybridization-based method comprises hybrid capture, amplification, or CRISPR (d)Cas9.

[Claim 5]. The method of claim 4, wherein the altering comprises depletion of the modified target nucleic acids that do not comprise the marker index sequences and/or the depletion comprises a hybridization-based method, preferably wherein the hybridization-based method comprises hybrid capture, amplification, or CRISPR (d)Cas9.

[Claim 6]. The method of claim 1, wherein the biological feature comprises

- a) a nucleotide sequence indicative of species type preferably, wherein the species type comprises the species of the cell,
- b) nucleotides of a 16s subunit, a 18s subunit, or an ITS non-transcriptional region
- c) a nucleotide sequence indicative of cell class preferably, wherein the cell class comprises expression pattern, epigenetic pattern, immune gene recombination, or a combination thereof and/or the epigenetic pattern comprises methylation mark, methylation pattern, accessible DNA, or a combination thereof
- d) a nucleotide sequence indicative of disease status or risk preferably, wherein disease status or risk comprises a variant DNA sequence, a variant expression pattern, or a variant epigenetic pattern that correlates with a disease;

wherein the variant DNA sequence comprises at least one single nucleotide polymorphism; the variant expression pattern comprises expression of a biomarker; and/or the variant epigenetic pattern comprises a methylation mark, methylation pattern.

[Claim 7]. The method of claim 1, wherein the modified target nucleic acids comprise a contiguous index of at least 2 compartment-specific index sequences, wherein there are no greater than 6 nucleotides between the 2 index sequences, wherein the contiguous index is present at each end of the modified target nucleic acids; the length of the contiguous index is at least 55 nucleotides; and/or

- (i) one copy of the contiguous index is present on the modified target nucleic acids; or
- (ii) two copies of the contiguous index are present on the modified target nucleic acids.

[Claim 8]. The method of claim 1, wherein the plurality of modified target nucleic acids of the sequencing library is representative of at least 100,000 different cells or nuclei.

[Claim 9]. The method of claim 1, wherein the providing the single-cell combinatorial sequencing library comprises:

processing a sample to produce a library, wherein the sample is a metagenomics sample obtained from an organism optionally, wherein the organism is a mammal and/or the metagenomics sample comprises a tissue suspected of comprising a commensal, a microbiome sample or pathogenic microbe preferably, wherein the microbe is prokaryotic or eukaryotic.

[Claim 10]. The method of claim 1, wherein the providing the single-cell combinatorial sequencing library comprises:

processing a sample, optionally cancer cells, to produce a library, wherein the sample is from an organism preferably, wherein the organism is a mammal and/or wherein a primary source of nucleic acids from the sample comprise

- (i) RNA preferably, wherein the RNA comprises mRNA;
- (ii) DNA preferably, wherein the DNA comprises whole cell genomic DNA;, wherein the whole cell genomic DNA comprises nucleosomes; or
- (iii) cell free DNA.

[Claim 11]. The method of claim 1, wherein the providing the single-cell combinatorial sequencing library comprises producing the library with a single-cell combinatorial indexing method selected from single-nuclei transcriptome sequencing, single-cell transcriptome sequencing, single-cell transcriptome and transposon-accessible chromatin sequencing, whole genome sequencing of single nuclei, single nuclei sequencing of transposon accessible chromatin, single-cell epitope sequencing, sci-HiC, and sci-MET

optionally, wherein the providing comprises providing two different single-cell combinatorial sequencing libraries from each cell or nucleus preferably, wherein the two different single-cell combinatorial sequencing libraries are selected from a single-cell combinatorial indexing method selected from single-nuclei transcriptome sequencing, single-cell transcriptome sequencing, single-cell transcriptome and transposon-accessible chromatin sequencing, whole genome sequencing of single nuclei, single nuclei sequencing of transposon accessible chromatin, sci-HiC, and sci-MET.

[Claim 12]. The method of claim 1, further comprising performing a sequencing procedure to determine the nucleotide sequences for the nucleic acids.

[Claim 13]. A method for preparing a sequencing library comprising nucleic acids from a plurality of single nuclei or cells, the method comprising:

- (a) providing a plurality of nuclei or cells, wherein the nuclei or cells comprise nucleosomes;
- (b) contacting the plurality of nuclei or cells with a transposome complex comprising a transposase and a universal sequence, wherein the contacting further comprises conditions suitable for incorporation of the universal sequence into DNA nucleic acids resulting in double stranded DNA nucleic acids comprising the universal sequence;
- (c) distributing the plurality of nuclei or cells, preferably by dilution, into a first plurality of compartments preferably, wherein the compartment comprises a well, microfluidic compartment, or a droplet

wherein each compartment comprises a subset of nuclei or cells preferably wherein compartments comprise from 50 to 100,000,000 nuclei or cells;

- (d) processing DNA molecules in each subset of nuclei or cells to generate indexed nuclei or cells,

wherein the processing comprises adding to DNA nucleic acids present in each subset of nuclei or cells a first compartment specific index sequence to result in indexed nucleic acids present in indexed nuclei or cells, preferably, wherein the adding of the compartment specific index sequence comprises a two-step process of adding a

nucleotide sequence comprising a universal sequence to the nucleic acids, and then adding the compartment specific index sequence to the nucleic acids;

wherein the processing comprises ligation, primer extension, hybridization, amplification, or a combination thereof; and

(e) combining the indexed nuclei or cells to generate pooled indexed nuclei or cells, optionally, further comprising obtaining the indexed nucleic acids from the pooled indexed nuclei or cells, thereby producing a sequencing library from the plurality of nuclei or cells.

[Claim 14]. The method of claim 13, wherein the providing comprises providing the plurality of nuclei or cells in a plurality of compartments, wherein each compartment comprises a subset of nuclei or cells, wherein the contacting comprises contacting each compartment with the transposome complex, and wherein the method further comprises combining the nuclei or cells after the contacting to generate pooled nuclei or cells, and/or wherein the adding of the compartment specific index sequence comprises a two-step process of adding a nucleotide sequence comprising a universal sequence to the nucleic acids, and then adding the compartment specific index sequence to the nucleic acids, optionally, further comprising obtaining the dual-indexed nucleic acids from the pooled dual-indexed nuclei or cells, thereby producing a sequencing library from the plurality of nuclei or cells.

[Claim 15]. The method of claim 13, wherein the providing comprises subjecting the nuclei to a chemical treatment to generate nucleosome-depleted nuclei while maintaining integrity of the isolated nuclei, and/or wherein the adding of the compartment specific index sequence comprises a two-step process of adding a nucleotide sequence comprising a universal sequence to the nucleic acids, and then adding the compartment specific index sequence to the nucleic acids, optionally, further comprising obtaining the triple-indexed nucleic acids from the pooled triple-indexed nuclei or cells, thereby producing a sequencing library from the plurality of nuclei or cells.

[Claim 16]. The method of claim 13, further comprising:

distributing the pooled indexed nuclei or cells comprising the indexed nuclei or cells, preferably by dilution, into a second plurality of compartments preferably, wherein the compartment comprises a well, microfluidic compartment, or a droplet

wherein each compartment comprises a subset of nuclei or cells preferably wherein compartments comprise from 50 to 100,000,000 nuclei or cells;

processing DNA molecules in each subset of nuclei or cells to generate dual-indexed nuclei or cells,

wherein the processing comprises adding to DNA nucleic acids present in each subset of nuclei or cells a second compartment specific index sequence to result in dual-indexed nucleic acids present in indexed nuclei or cells,

wherein the processing comprises ligation, primer extension, hybridization, amplification, or a combination thereof;

combining the dual-indexed nuclei or cells to generate pooled dual-indexed nuclei or cells;

optionally further comprising: distributing the pooled nuclei or cells comprising the dual-indexed nuclei or cells, preferably by dilution, into a third plurality of compartments preferably, wherein the compartment comprises a well, microfluidic compartment, or a droplet

wherein each compartment comprises a subset of nuclei or cells preferably wherein compartments comprise from 50 to 100,000,000 nuclei or cells;

processing DNA molecules in each subset of nuclei or cells to generate triple-indexed nuclei or cells,

wherein the processing comprises adding to DNA nucleic acids present in each subset of nuclei or cells a third compartment specific index sequence to result in triple-indexed nucleic acids present in indexed nuclei or cells,

wherein the processing comprises ligation, primer extension, hybridization, amplification, or a combination thereof; and/or

combining the triple-indexed nuclei or cells to generate pooled triple-indexed nuclei or cells.

[Claim 17]. The method of claim 13, wherein the contacting comprises contacting each subset with two transposome complexes, wherein one transposome complex comprises a first transposase comprising a first universal sequence and a second transposome complex comprises a second transposase comprising a second universal sequence, wherein the contacting further comprises conditions suitable for incorporation of the first universal sequence and the second universal sequence into DNA nucleic acids resulting in double stranded DNA nucleic acids comprising the first and second universal sequences.

[Claim 18]. The method of any one of claims 13-17, further comprising:
providing a surface comprising a plurality of amplification sites,
wherein the amplification sites comprise at least two populations of attached single stranded capture oligonucleotides having a free 3' end, and

contacting the surface comprising amplification sites with the nucleic acid fragments comprising one, two, or three index sequences under conditions suitable to produce a plurality of amplification sites that each comprise a clonal population of amplicons from an individual fragment comprising a plurality of indexes.

[Claim 19]. A method for preparing a nucleic acid library comprising:

(a) providing a plurality of samples, wherein each sample comprises a plurality of cells or nuclei, wherein the plurality of cells or nuclei of each sample are present in one or more separate compartments preferably, wherein the plurality of nuclei or cells are fixed;

(b) contacting the plurality of nuclei or cells with a transposome complex comprising a transposase and a universal sequence and with the proviso that the transposome complex does not comprise an index sequence, wherein the contacting further comprises conditions suitable for incorporation of the universal sequence into nucleic acids;

(c) adding a first index sequence to the nucleic acids of each separate compartment preferably, further comprising an amplification of indexed nucleic acids thereafter;

(d) combining the cells or nuclei of the separated compartments;

(e) distributing the cells or nuclei into a plurality of compartments; and

(f) adding a second index sequence to the nucleic acids of the plurality of compartments preferably, further comprising an amplification of indexed nucleic acids thereafter.

[Claim 20]. The method of claim 19, wherein the first index sequence, the second index sequence, or the combination thereof, are added by ligation, primer extension, hybridization, amplification, or a combination thereof, and/or wherein steps (d)-(e) are repeated to add a third or more index sequences to the cells or nuclei of the plurality of compartments.

[Claim 21]. The method of claim 20, further comprising step (g) combining the nucleic acids of the plurality of compartments and determining the sequence of the nucleic acids.

[Claim 22]. The method of claim 20, further comprising performing a sequencing procedure to determine the nucleotide sequences for the nucleic acids.

[Claim 23]. A method for sequencing a single cell or nucleus comprising:

- (a) uniquely indexing nucleic acids of each cell or nuclei in a sample, thereby generating an indexed library for each cell or nuclei;
- (b) using a biological feature to identify one or more indexed libraries of interest from step (a);
- (c) enriching the indexed libraries of interest of step (b) thereby generating an enriched library preferably, wherein the identifying the indexed library of interest comprises sequencing the indexes; and
- (d) sequencing the enriched library from step (c).

[Claim 24]. The method of claim 23, wherein the libraries are derived from DNA, RNA, or protein of the cells or nuclei, and/or wherein the biological feature is DNA, RNA, or protein or a combination thereof.

[Claim 25]. The method of any one of claims 23 or 24, wherein the uniquely indexing in step (a) comprises associating at least two different indexes to the nucleic acids of the cells or nuclei preferably, wherein the at least two different indexes are a contiguous index.

[Claim 26]. The method of any one of claims 23 or 24, wherein the enriched library is generated through positive enrichment preferably, wherein

- a) the positive enrichment comprises amplification;
- b) the positive enrichment comprises a capture agent;
- c) the positive enrichment comprises a solid support; or
- d) the enriched library is generated through negative enrichment.

[Claim 27]. A method for sequencing a single cell or nucleus comprising:

- (a) providing a sample, wherein the sample comprises a plurality of nuclei or cells;
- (b) associating a first index on each nucleus or cell in the sample;
- (c) dividing the sample into a plurality of compartments;
- (d) associating a second index on each nuclei or cell of the plurality of compartments;
- (e) pooling the plurality of compartments;
- (f) sequencing the pooled compartments;
- (g) identifying a combination of first and second indexes associated with a biological feature;
- (h) enriching the biological feature from the pooled compartments using the identified combination of first and second indexes from step (g).

[Claim 28]. A kit containing:

- (a) a plurality of transposome complexes, wherein each transposome complex comprises a transposase and a transposon sequence, wherein the transposon sequence is not indexed;
- (b) a first plurality of index oligonucleotides, wherein the first plurality of index oligonucleotides comprises oligonucleotides having at least two different sequences; and
- (c) a ligase enzyme for use with the index oligonucleotides.

[Claim 29]. The kit of claim 28, further comprising a second plurality of index oligonucleotides, wherein the second plurality of index oligonucleotides comprises oligonucleotide having different sequences from the first plurality of index oligonucleotides.

[Claim 30]. The kit of claim 28, further comprising a third plurality of index oligonucleotides, wherein the third plurality of index oligonucleotides comprises oligonucleotide having different sequences from the first plurality of index oligonucleotides and the second plurality of index oligonucleotides.