CLAIMS

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

(a) providing a single-cell sequencing library,

wherein the single-cell sequencing library comprises a plurality of modified target nucleic acids,

wherein the modified target nucleic acids are derived from DNA or protein of the cells or nuclei and comprise at least one index sequence;

(b) interrogating the single-cell sequencing library by targeted sequencing to identify index sequences associated with 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 single-cell 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 single-cell 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 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 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 any one of claims 1-3, wherein the altering comprises depleting of the modified target nucleic acids that do not comprise the marker index sequences, optionally wherein the depleting 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, optionally wherein the

contiguous index is present at each end of the modified target nucleic acids, optionally wherein
the length of the contiguous index is at least 55 nucleotides, and optionally wherein
(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 single-cell 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 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 sequencing library comprises:

processing a sample, optionally wherein the sample comprises 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) DNA preferably, wherein the DNA comprises whole cell genomic DNA, wherein the whole cell genomic DNA comprises nucleosomes; or
(ii) cell free DNA.

Claim 11. The method of claim 1, wherein the providing the single-cell sequencing library comprises producing the library with a single-cell combinatorial indexing method selected from 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 sequencing libraries from each cell or nucleus preferably, wherein the two different single-cell sequencing libraries are selected from a single-cell combinatorial indexing method selected from whole genome sequencing of single nuclei, single nuclei sequencing of transposon accessible chromatin, sci-HiC, and sci-MET.

Claim 12. A method for sequencing a single cell or nucleus comprising:

(a) uniquely indexing nucleic acids of each cell or nuclei in a sample, wherein the nucleic acids derived from DNA or protein of the cells or nuclei, 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 13. The method of claim 12, wherein the biological feature is DNA, RNA, or protein or a combination thereof.

Claim 14. The method of any one of claims 12 or 13, 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 15. The method of any one of claims 12 or 13, wherein the enriched library is generated through positive enrichment or through negative enrichment, optionally wherein the positive enrichment comprises amplification, a capture agent, or a solid support.

Claim 16. The method of claim 1, wherein the marker index sequence correlates with a first nucleotide region, and the biological feature correlates with a second nucleotide

region, and the step of interrogating comprises determining sequences at the first and second nucleotide regions.