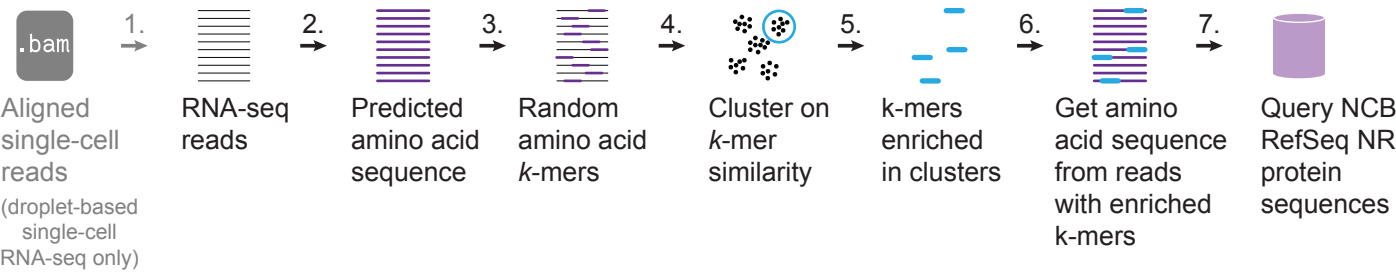
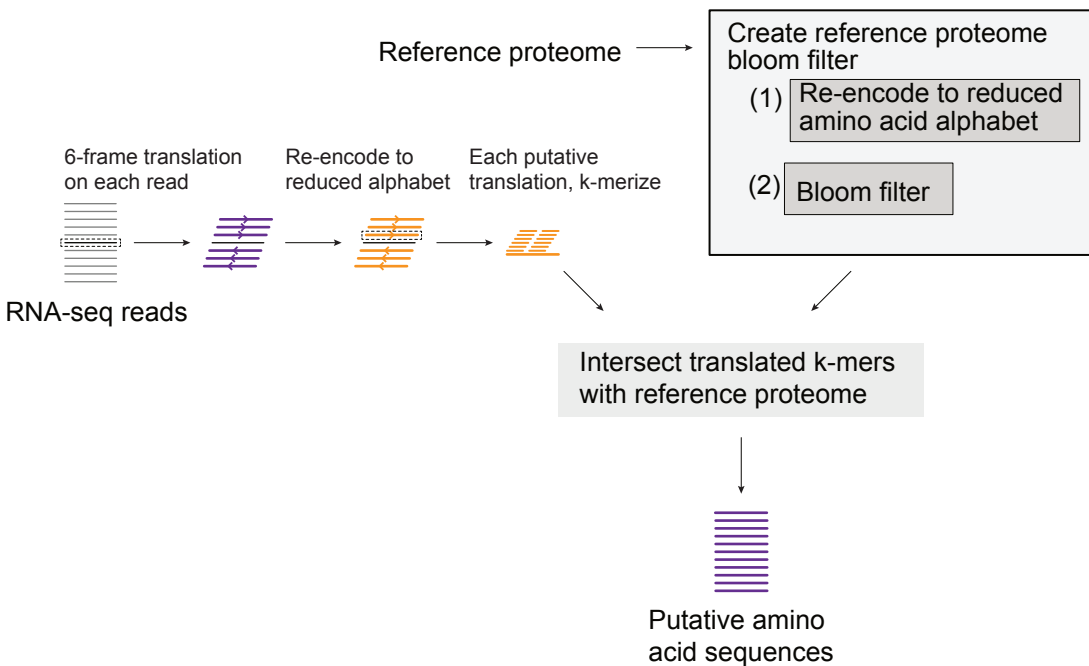


A



- A**
1. Extract cell barcodes with sufficient nUMI from single-cell bam via `bam2fasta`
 2. Predict amino acid sequence from RNA-seq reads via `khtools extract_coding`



- | | | |
|-----|-------------------------------|------------------------------|
| (b) | <code>sourmash sketch</code> | MinHash protein k-mers |
| (c) | <code>sourmash compare</code> | Compute cell-cell similarity |
| (d) | <code>sourmash knn</code> | Build nearest neighbor graph |

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads

Create reference proteome
bloom filter

(1) Re-encode to reduced
amino acid alphabet

(2) Bloom filter

RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads

RNA-seq reads



6-frame translation
+ reduced alphabet

Create reference proteome
bloom filter

(1) Re-encode to reduced
amino acid alphabet

(2) Bloom filter

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

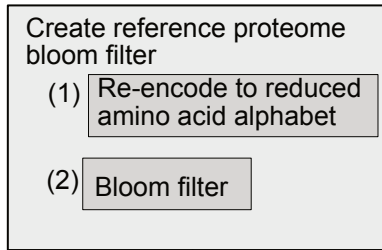
(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads



RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

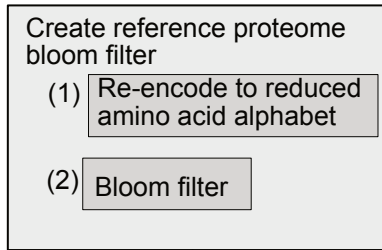
(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads



RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

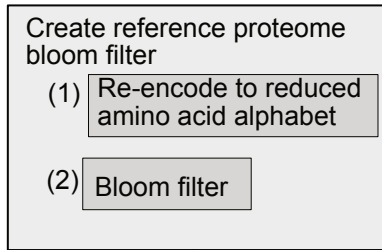
(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads



RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

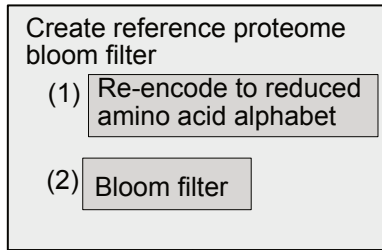
(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads



RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

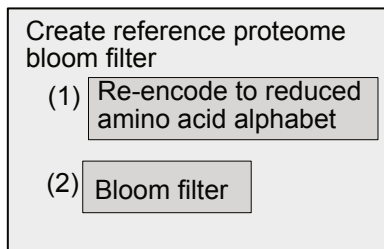
(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads



RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

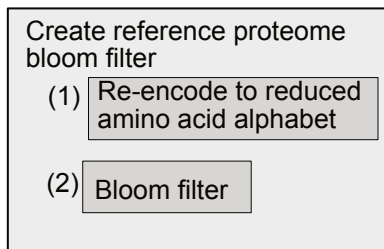
(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads



RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

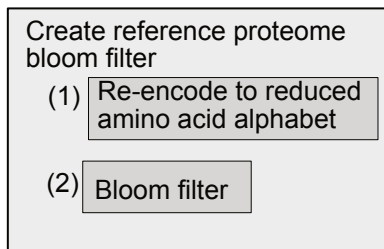
(d) `sourmash knn`

Build nearest neighbor graph

(a) `khtools extract_coding`

Reference proteome
(e.g. human)

Predict protein-coding
sequences from
RNA-seq reads



RNA-seq reads



6-frame translation
+ reduced alphabet

Intersect translated k-mers
with reference proteome

Putative protein-coding sequences

(2)

(b) `sourmash sketch`

MinHash protein k-mers

(c) `sourmash compare`

Compute cell-cell similarity

(d) `sourmash knn`

Build nearest neighbor graph