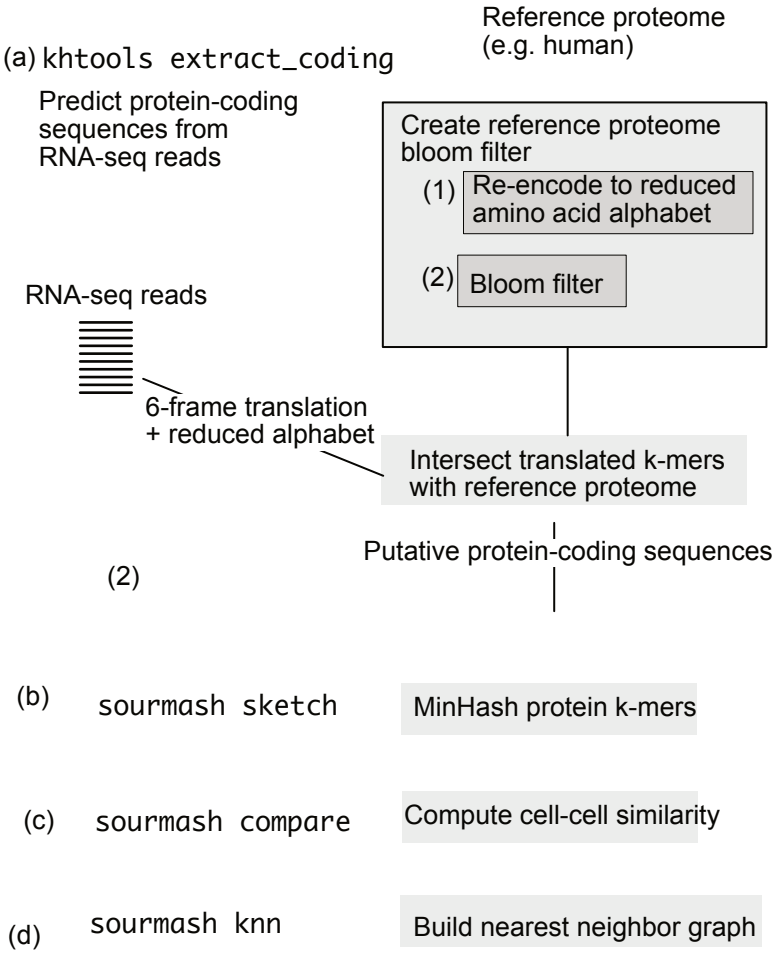
Reference proteome
(e.g. human)
(a) khtools extract_coding
Predict protein-coding sequences from RNA-seq reads
Create reference proteome bloom filter
(1) Re-encode to reduced amino acid alphabet
(2) Bloom filter
RNA-seq reads
6-frame translation + reduced alphabet
Intersect translated k-mers with reference 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