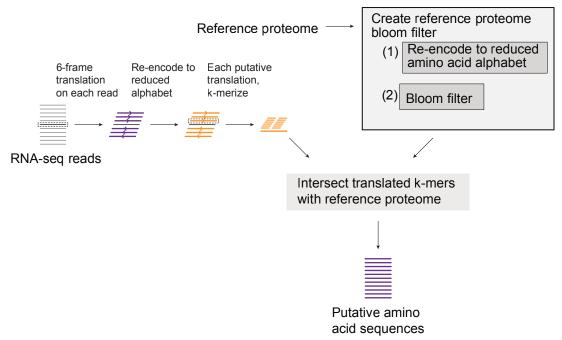
- 1. Extract cell barcodes with sufficient nUMI from single-cell bam via bam2fasta
 - 2. Predict amino acid sequence from RNA-seq reads via khtools extract coding



- 3. Randomly subsample amino acid k-mers via MinHash algorithm in sourmash sketch
- (b) sourmash sketch MinHash protein k-mers Compute cell-cell similarity sourmash compare (c) (d) sourmash index Build sequence bloom tree (SBT) (e) sourmash knn Build nearest neighbor graph (f) sourmash umap Build UMAP off of knn XP_008705901.1 PREDICTED: phosph XP_026370896.1 phospholipase A2 XP_011803952.1 PREDICTED: phosph NP_001177150.1 cytochrome c oxid XP_003419308.1 cytochrome c oxid В samtools fastq khtools extract coding diamond blastp (if input is bam, Adapter, Predicted **FASTQs** Putatitive homologs convert to fastq) poly-A, amino acid trimmed sequence UNIPROT/SwissProt NCBI RefSeq manually curated eukaryotic non-redundant protein protein sequences Stop codon in all translation frames Gorilla Gorilla Gorilla

