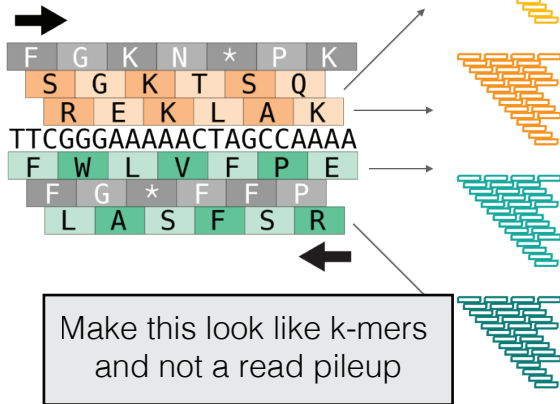


## A kmerslay extract-coding

### For each read, do six-frame translation

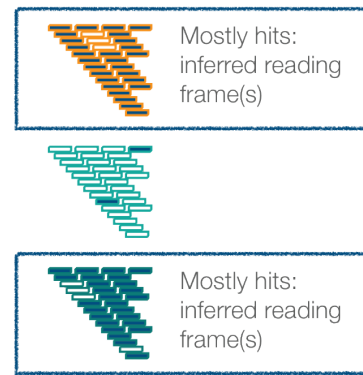
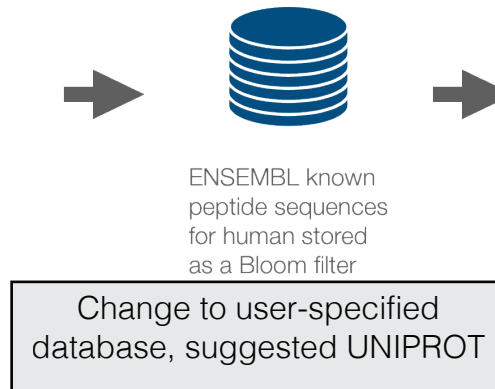
Discard reading frames with stop codons



### k-merize each "non-stop" reading frame

### Check against database of known human amino acid k-mers

### Infer reading frame from presence of known k-mers



## B

ATGGAATGTAACATGCATTATATGTGGACAACCCACCATGAA  
GAAGAAATGATGTTCTGTGATGTGTGTGACAGAGGTGCTGGATT

### Translation in forward direction:

frame +1  
LYVDNPTMKKK\*CSVMCVTEVAG X Stop codon

frame +2  
YMWTPP\*RRNDVL\*CV\*QRSLD X Stop codon

frame +3  
ICGQPHHEEMFCDVCDGRWI → 53% of 7-mers matched known peptides

✓ Accepted reading frame

### Translation in reverse direction:

frame -1  
NPATSVTHTEHHFFFMVGLSTY → 6% of 7-mers matched known peptides

frame -2  
IQRP LSHTSQNISSSWGCPHI → 0% of 7-mers matched known peptides

frame -3  
SSDLCHTHRTSFLHGGVVHI\* X Stop codon

## C

>A00111:133:H3VGJDSXX:1:2362:27805:31454 2:N:0:TTTGACAGGCTG+TCATTACATCAT  
CTCGTGGATGAGGCCAGATGCAAGAGAGCAACAGGAGGAGAGGAATCTCAGGAAAAATCTTGCTTGAATCCTTGCCAAGACAGATTCTC

>A00111:133:H3VGJDSXX:1:2362:27805:31454 2:N:0:TTTGACAGGCTG+TCATTACATCAT translation\_frame: -1 jaccard: 1.0

ENLVFGKGFQAKIFPEILLCLLLALFASGLIHR

>A00111:133:H3VGJDSXX:1:2362:27805:31454 2:N:0:TTTGACAGGCTG+TCATTACATCAT translation\_frame: -2 jaccard: 0.5185185185185185

RILSLAKDSKQRFLLRFSSASCWLSLHLASSTE

## UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

