Supplemental Figure 1 A oolong translate Check against database k-merize each "non-stop" Infer reading frame from For each read, do sixof known human amino reading frame presence of known k-mers frame translation acid k-mers Discard reading frames with stop codons Mostly hits: G K T S Q E K L A K inferred reading frame(s) TTCGGGAAAAACTAGCCAAAA ENSEMBL known peptide sequences for human stored as a Bloom filter Mostly hits: Make this look like k-mers inferred reading Change to user-specified and not a read pileup frame(s) database, suggested UNIPROT В ATGGAATGTAAAACATGCATTATATGTGGACAACCCCACCATGAA C GAAGAAATGATGTTCTGTGATGTGTGACAGAGGTCGCTGGATT



