Focus on variable region sequence

- Apply Kabat numbering
- Identify CDRs and framework
- Curate all full-length sequences
- Remove CDRs
- Curate all framework only sequences

All Human Databases:
- VH: 29,958 sequences
- VK: 5,042 sequences
- VL: 3,708 sequences

Full-length sequences:
- BLAST sequence database
- Identify the 20 sequences with highest % ID

Framework only sequences:
- BLAST sequence database
- Identify the 20 sequences with highest % ID

Obtain T20 score for each human sequence
Sort T20 scores from high to low
Remove bottom ~15% sequences (T20 score cutoff)

T20 Cutoff Human Databases:
- VH: 25,389 sequences (full-length; cutoff = 79)
- VH: 25,290 sequences (framework only; cutoff = 84)
- VK: 4,309 sequences (full-length; cutoff = 86)
- VK: 4,339 sequences (framework only; cutoff = 90)
- VL: 3,106 sequences (full-length; cutoff = 84)
- VL: 3,144 sequences (framework only; cutoff = 88)