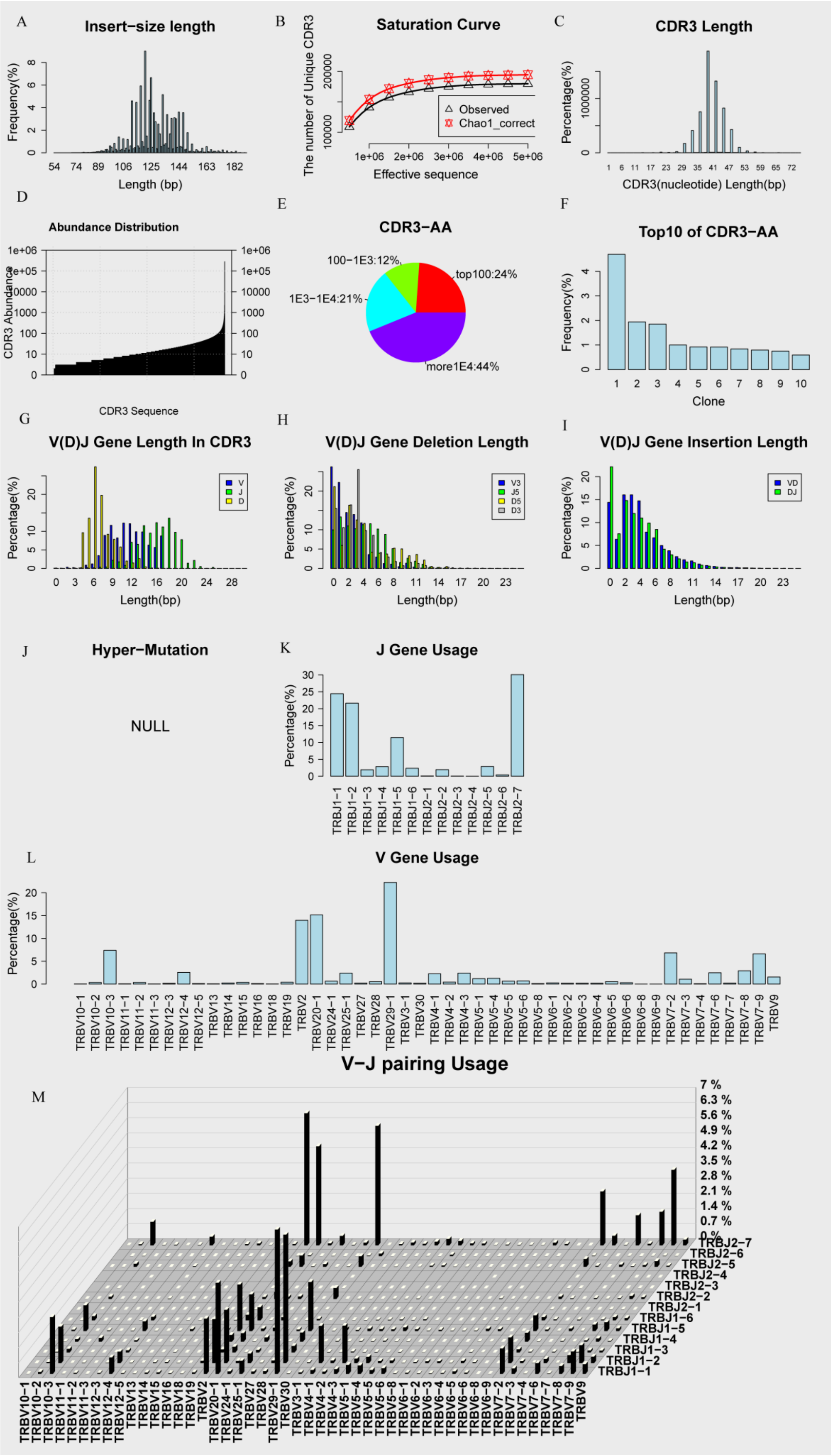
A
Insert−size length
Frequency(%)
0 2 4 6 8
54 74 89 106 125 144 163 182
Length (bp)
B
Saturation Curve
The number of Unique CDR3
100000 200000
1e+06 2e+06 3e+06 4e+06 5e+06
Effective sequence
Observed
Chao1_correct
C
CDR3 Length
Percentage(%)
0 1000000
1 6 11 17 23 29 35 41 47 53 59 65 72
CDR3(nucleotide) Length(bp)
D
Abundance Distribution
CDR3 Abundance
1e+06 1e+05 10000 1000 100 10 0
CDR3 Sequence
E
CDR3−AA
100−1E3:12%
top100:24%
1E3−1E4:21%
more1E4:44%
F
Top10 of CDR3−AA
Frequency(%)
0 1 2 3 4
1 2 3 4 5 6 7 8 9 10
Clone
G
V(D)J Gene Length In CDR3
Percentage(%)
0 5 10 20
0 3 6 9 12 16 20 24 28
Length(bp)
V
J
D
H
V(D)J Gene Deletion Length
Percentage(%)
0 5 10 20
0 2 4 6 8 11 14 17 20 23
Length(bp)
V3
J5
D5
D3
I
V(D)J Gene Insertion Length
Percentage(%)
0 5 10 15 20
0 2 4 6 8 11 14 17 20 23
Length(bp)
VD
DJ
J
Hyper−Mutation
NULL
K
J Gene Usage
Percentage(%)
0 5 10 15 20 25 30
TRBJ1−1 TRBJ1−2 TRBJ1−3 TRBJ1−4 TRBJ1−5 TRBJ1−6 TRBJ2−1 TRBJ2−2 TRBJ2−3 TRBJ2−4 TRBJ2−5 TRBJ2−6 TRBJ2−7
L
V Gene Usage
Percentage(%)
0 5 10 15 20
TRBV10−1 TRBV10−2 TRBV10−3 TRBV11−1 TRBV11−2 TRBV11−3 TRBV12−3 TRBV12−4 TRBV12−5 TRBV13 TRBV14 TRBV15 TRBV16 TRBV18 TRBV19 TRBV2 TRBV20−1 TRBV24−1 TRBV25−1 TRBV27 TRBV28 TRBV29−1 TRBV3−1 TRBV30 TRBV4−1 TRBV4−2 TRBV4−3 TRBV5−1 TRBV5−4 TRBV5−5 TRBV5−6 TRBV5−8 TRBV6−1 TRBV6−2 TRBV6−3 TRBV6−4 TRBV6−5 TRBV6−6 TRBV6−8 TRBV6−9 TRBV7−2 TRBV7−3 TRBV7−4 TRBV7−6 TRBV7−7 TRBV7−8 TRBV7−9 TRBV9
V−J pairing Usage
M
7 % 6.3 % 5.6 % 4.9 % 4.2 % 3.5 % 2.8 % 2.1 % 1.4 % 0.7 % 0 %
TRBJ2−7 TRBJ2−6 TRBJ2−5 TRBJ2−4 TRBJ2−3 TRBJ2−2 TRBJ2−1 TRBJ1−6 TRBJ1−5 TRBJ1−4 TRBJ1−3 TRBJ1−2 TRBJ1−1
TRBV10−1 TRBV10−2 TRBV10−3 TRBV11−1 TRBV11−2 TRBV11−3 TRBV12−3 TRBV12−4 TRBV12−5 TRBV13 TRBV14 TRBV15 TRBV16 TRBV18 TRBV19 TRBV2 TRBV20−1 TRBV24−1 TRBV25−1 TRBV27 TRBV28 TRBV29−1 TRBV30 TRBV3−1 TRBV4−1 TRBV4−2 TRBV4−3 TRBV5−1 TRBV5−4 TRBV5−5 TRBV5−6 TRBV5−8 TRBV6−1 TRBV6−2 TRBV6−3 TRBV6−4 TRBV6−5 TRBV6−6 TRBV6−8 TRBV6−9 TRBV7−2 TRBV7−3 TRBV7−4 TRBV7−6 TRBV7−7 TRBV7−8 TRBV7−9 TRBV9

Figure S4. H-B-01 sample output figure of IMonitor. (A) Sequence length distribution. (B) Saturation curve, rarefaction studies of sequences. Sub-sequences are randomly selected and observed unique CDR3 number and predicted CDR3 number (Chao1 corrected algorithm) are calculated. (C) CDR3 nucleotide length distribution. (D) CDR3 abundance distribution. (E) CDR3 amino acid frequencies sectional content. (F) Top ten frequency of CDR3 amino acid. (G) Length distribution of V/D/J gene in CDR3 region. (H) Deletion length distribution of V/D/J gene. (I) Insertion length distribution of between V and D gene, D and J gene. (J) Hyper-mutation, Only for Ig. (K), J gene usage. (L) V gene usage. (M) Three-dimensional graph of V-J pairing.

16 SI **Wei Zhang et al.**