Table 3.1: Analysis of raw sequence data before accounting Primer ID for downstream analysis.

|  |  |  |
| --- | --- | --- |
|  | Run1 | Run2 |
| Total raw reads | 125865 | 40544 |
| Total read discarded for no primer | 2962 | 751 |
| Total reads discarded for no MID | 7557 | 1109 |
| Total reads discarded for ambiguous base in Primer ID | 257 | 37 |
| Total reads after demultiplexing | 115089 | 38647 |

Table 3.2: Amplicon specific forward and reverse sequence read length of interest

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *ENV* | | GAG54 | | GAG472 | | NEF23 | |
| Forward | Reverse | Forward | Reverse | Forward | Reverse | Forward | Reverse |
| 377 | 160 | 315 | 162 | 318 | 160 | 364 | 177 |

Table 3.3: Number of reads discarded for not being full length

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | *ENV* | | GAG54 | | GAG472 | | NEF23 | |
| F | R | F | R | F | R | F | R |
| Run 1 | Patient A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Patient B | 0 | 10 | 0 | 2 | 0 | 7 | 0 | 13 |
| Patient C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Patient D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Run 1 | Patient E | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 2 |
| Patient F | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 7 |
| Patient G | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 0 |
| Patient H | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |

Table 3.4 Number of reads retained for being full length

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | *ENV* | | GAG54 | | GAG472 | | NEF23 | |
| F | R | F | R | F | R | F | R |
| Run 1 | Patient A | 7239 | 13422 | 946 | 3036 | 1197 | 2988 | 1350 | 1940 |
| Patient B | 3780 | 6159 | 1490 | 2493 | 2326 | 4272 | 2119 | 2846 |
| Patient C | 5158 | 7556 | 2687 | 4675 | 2054 | 2865 | 2112 | 2999 |
| Patient D | 4038 | 5709 | 1861 | 3380 | 2544 | 3782 | 2354 | 3589 |
| Run 2 | Patient E | 862 | 1915 | 457 | 741 | 524 | 1997 | 22 | 2024 |
| Patient F | 1118 | 2374 | 658 | 243 | 312 | 1023 | 1357 | 1320 |
| Patient G | 1445 | 4020 | 671 | 726 | 363 | 1756 | 1123 | 1998 |
| Patient H | 2066 | 3603 | 83 | 712 | 211 | 1394 | 8 | 1498 |

Table 3.5 Total number of unique Primer ID tags in each dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *ENV* | GAG54 | GAG472 | NEF23 |
| Run 1 | Patient A | 1064 | 1051 | 232 | 282 |
| Patient B | 4193 | 3031 | 4097 | 470 |
| Patient C | 750 | 422 | 228 | 479 |
| Patient D | 1030 | 1710 | 145 | 374 |
| Run 2 | Patient E | 253 | 590 | 110 | 546 |
| Patient F | 515 | 424 | 115 | 126 |
| Patient G | 2105 | 1048 | 660 | 930 |
| Patient H | 607 | 279 | 229 | 157 |

Table 3.6 Breakdown of the percentage of Primer IDs tags for each unique dataset with two or less and three or more representative sequences.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | *ENV* | | GAG54 | | GAG472 | | NEF23 | |
| <3 | ≥3 | <3 | ≥3 | <3 | ≥3 | <3 | ≥3 |
| Run 1 | Patient A | 77.35 | 22.65 | 59.47 | 40.53 | 76.72 | 23.28 | 80.50 | 19.50 |
| Patient B | 68.78 | 31.22 | 94.09 | 5.91 | 87.06 | 12.94 | 73.62 | 26.38 |
| Patient C | 76.67 | 23.33 | 69.91 | 30.09 | 67.11 | 32.89 | 63.67 | 36.33 |
| Patient D | 64.66 | 35.34 | 57.08 | 42.92 | 83.45 | 16.55 | 69.79 | 30.21 |
| Run 2 | Patient E | 80.24 | 19.76 | 74.58 | 25.42 | 82.73 | 17.27 | 50.18 | 49.82 |
| Patient F | 60.39 | 39.61 | 70.75 | 29.25 | 73.04 | 26.96 | 65.87 | 34.13 |
| Patient G | 67.27 | 32.73 | 93.99 | 6.01 | 62.27 | 37.73 | 63.98 | 36.02 |
| Patient H | 55.85 | 44.15 | 59.14 | 40.86 | 56.33 | 43.67 | 83.12 | 16.88 |

Table 3.7: Average number of representative sequence per Primer ID tag for each of the unique datasets. The average values were round to the nearest integer value.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | *ENV* | GAG54 | GAG472 | NEF23 | Average |
| Run1 | Patient A | 76.9 | 4.60 | 65.82 | 24.35 | 43 |
| Patient B | 7.50 | 3.54 | 54.01 | 5.30 | 18 |
| Patient C | 64.33 | 3.86 | 56.46 | 229.0 | 89 |
| Patient D | 54.87 | 35.93 | 27.04 | 47.77 | 42 |
| Run2 | Patient E | 50.98 | 4.3 | 127.47 | 6.21 | 48 |
| Patient F | 15.35 | 4.29 | 39.93 | 59.93 | 30 |
| Patient G | 5.34 | 3.36 | 6.43 | 7.071 | 6 |
| Patient H | 19.71 | 5.10 | 14.54 | 34.27 | 19 |
|  | Average | 37 | 9 | 49 | 52 |  |

Table 3.8: The number of duplicated Primer IDs in Run1 and Run2.

|  |  |  |
| --- | --- | --- |
|  | Sample | Number of Duplicated Primer IDs |
| Run1 | Patient A | 77 |
| Patient B | 1103 |
| Patient C | 44 |
| Patient D | 75 |
| Run2 | Patient E | 32 |
| Patient F | 15 |
| Patient G | 242 |
| Patient H | 42 |