Table 4. 1: Generic names of antiretroviral drugs, their drug class and abbreviations.

|  |  |  |
| --- | --- | --- |
| **Drug Class** | **Generic Drug Name** | **Abbreviation** |
| Nucleoside Reverse Integrase Inhibitors (NRTIs) | 1. Abacavir | ABC |
| 1. didanosine | DDI |
| 1. Emtricitabine | FTC |
| 1. Lamivudine | 3TC |
| 1. Stavudine | d4T |
| 1. Tenofovir | TDF |
| 1. Zidovudine | AZT |
| Non-Nucleoside Reverse Integrase Inhibitors  (NNRTIs) | 1. Efavirenz | EFV |
| 1. Etravirine | ETR |
| 1. Nevirapine | NVP |
| 1. Rilpivirine | RPV |
| Protease Inhibitors (PIs) | 1. Atazanavir/r | ATV/r |
| 1. Darunavir/r | DRV/r |
| 1. Fosamprenavir/r | FPV/r |
| 1. Indinavir/r | IDV/r |
| 1. Lopinavir/r | LPV/r |
| 1. Nelfinavir | NFV |
| 1. Saquinavir/r | SQV/r |
| 1. Tipranavir/r | TPV/r |
| Integrase Inhibitors (IN) | 1. Raltegravir | RAL |
| 1. Elvitegravir | EVG |
| 1. Dolutegravir | DTG |

Table 4. 2: The standard list of HIV drug resistant mutation codon positions with respect to the genes

|  |  |
| --- | --- |
| HIV gene | Standard drug resistant mutation codon positions in the gene |
| PR | 10,11,13,20,23,24,30,32,33,34,35,36,43,46,47,48,50,53,54,55,58,60,62,  63,64,71,73,74,76,77,82,83,84,85,88,89,90,93 |
| RT | 40,41,44,62,65,66,67,68,69,70,71,74,75,77,90,98,100,101,103,106,108,  115,116,118,138,151,179,181,184,188,190,210,215,219,221,225,227,  230,236,238,318,333,348 |
| IN | 51,66,68,72,74,92,95,97,114,121,125,128,138,140,143,145,146,147,148,  151,153,154,155,157,163,201,203,206,230,263 |

Table 4. : Number of sequences mixed to generate simulated sequence data with different percentage of resistant sequences.

|  |  |  |
| --- | --- | --- |
| Percentage of resistant sequences | Number of resistant sequences | Number of sensitive sequences |
| 0.1 | 10 | 9990 |
| 1 | 100 | 9900 |
| 5 | 500 | 9500 |
| 10 | 1000 | 9000 |
| 15 | 1500 | 8500 |
| 20 | 2000 | 8000 |
| 50 | 5000 | 5000 |

Table 4. : Drug resistant mutations grouped by drug class to which the mutation is highly resistant. The drug resistant mutations are present in five resistant sequences (indicated by ‘\_R’ in sequence names) and none are present in five susceptible sequences (indicated by ‘\_S’ in sequence names) as reported from web Sierra.

|  |  |  |  |
| --- | --- | --- | --- |
| Sequences names | NRTIs | NNRTIs | PIs |
| 56252\_R | K65R, K70R, V75I, F77L, Y115F, F116Y, Q151M | K103N | A71V, L90M |
| 56252\_S | None | None | None |
| 21354\_R | K65R, D67N, K70G | K103N | |  | | --- | | L10I, G48V, I50V, I54V, A71V, V82A | |
| 21354\_S | None | None | None |
| 63377\_R | K65R, D67N, K70G | K103N | L10I, G48V, I50V, I54V, A71V, V82T |
| 63377\_S | None | None | None |
| 4134\_R | K65R, D67N, K70G | K103N, Y181C | L10I, G48V, I54V, V82A |
| 4134\_S | None | None | None |
| 2368\_R | M41L, A62V, K65R, V75I | |  | | --- | | K103N, Y181C | | A71V, G73S, L90M |
| 2368\_S | None | None | None |

Table 4. : Resistance Calls of five resistant sequences (indicated by ‘\_R’ in sequence names) and five susceptible sequences (indicated by ‘\_S’ in sequence names) for simulated data using web Sierra

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Drugs | 56252\_R | 56252\_S | 21354\_R | 21354\_S | 63377\_R | 63377\_S | 4134\_R | 4134\_S | 2368\_R | 2368\_S |
| 3TC | R | S | R | S | R | S | R | S | R | S |
| ABC | S | S | S | S | S | S | I | S | I | S |
| ATV/r | R | S | R | S | R | S | R | S | R | S |
| AZT | S | S | S | S | S | S | I | S | I | S |
| D4T | R | S | R | S | R | S | R | S | R | S |
| DDI | R | S | R | S | R | S | R | S | R | S |
| DRV/r | R | S | I | S | I | S | I | S | I | S |
| EFV | R | S | I | S | I | S | I | S | I | S |
| ETR | R | S | R | S | R | S | R | S | R | S |
| FPV/r | R | S | R | S | R | S | R | S | R | S |
| FTC | R | S | S | S | S | S | S | S | S | S |
| IDV/r | I | S | R | S | R | S | R | S | I | S |
| LPV/r | S | S | I | S | I | S | S | S | S | S |
| NFV | I | S | R | S | R | S | I | S | I | S |
| NVP | I | S | R | S | R | S | R | S | I | S |
| RPV | S | S | R | S | R | S | I | S | I | S |
| SQV/r | R | S | R | S | R | S | R | S | R | S |
| TDF | I | S | R | S | R | S | R | S | R | S |
| TPV/r | S | S | S | S | I | S | I | S | S | S |
| **R: Resistant, I: Intermediate, S: Susceptible** | | | | | | | | | | |

Table 4. 6 The start and end nucleotide positions of full-length amplicons, codon positions of full-length amplicons and codon positions of optimal full-length amplicons in the simulated datasets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Simulated Amplicon name | Full length amplicon Nucleotide position | | Full length amplicon Codon positions | | Optimal full length amplicon codon positions | |
| Start | End | Start | End | Start | End |
| PR | 169 | 469 | 57 | 157 | 66 | 149 |
| RT1 | 466 | 812 | 156 | 271 | 195 | 270 |
| RT2 | 672 | 1017 | 224 | 339 | 224 | 336 |

Table 4. : The mean observed prevalence of all DRMs from all three simulated amplicons datasets - PR, RT1 and RT2 in all samples. The mean across samples calculated from the mean observed prevalence is shown at the last column.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Expected Prevalence (%) | 56252 | 21354 | 63377 | 4134 | 2368 | **Mean across samples** |
| 0.1 | 0.11 | 0.13 | 0.13 | 0.13 | 0.14 | **0.128** |
| 1 | 0.99 | 1.04 | 1.05 | 1.00 | 1.07 | **1.03** |
| 5 | 5.03 | 5.00 | 5.10 | 5.00 | 5.09 | **5.044** |
| 10 | 9.94 | 9.95 | 10.18 | 9.86 | 10.16 | **10.018** |
| 15 | 14.86 | 14.94 | 15.24 | 14.88 | 15.37 | **15.058** |
| 20 | 19.86 | 19.75 | 20.28 | 19.83 | 20.10 | **19.964** |
| 50 | 49.50 | 49.46 | 50.16 | 49.36 | 50.13 | **49.722** |

Table 4. 8: Resistance calls to PI, NRTI and NNRTI drugs in PR, RT1 and RT2 amplicons simulated with the expected DRM prevalence of 0.1%, 1%, 5% and 10% in all samples. The green color denoted that drugs are sensitive to the samples.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Drug class | Sample/Drug | 56252 | 21354 | 63377 | 4134 | 2368 |
| PI | ATV/r | S | S | S | S | S |
| DRV/r | S | S | S | S | S |
| FPV/r | S | S | S | S | S |
| IDV/r | S | S | S | S | S |
| LPV/r | S | S | S | S | S |
| NFV | S | S | S | S | S |
| SQV/r | S | S | S | S | S |
| TPV/r | S | S | S | S | S |
| NRTI | ABC | S | S | S | S | S |
| DDI | S | S | S | S | S |
| FTC | S | S | S | S | S |
| 3TC | S | S | S | S | S |
| D4T | S | S | S | S | S |
| TDF | S | S | S | S | S |
| AZT | S | S | S | S | S |
| NNRTI | EFV | S | S | S | S | S |
| ETR | S | S | S | S | S |
| NVP | S | S | S | S | S |
| RPV | S | S | S | S | S |

Table 4. 9: Resistance calls to PI, NRTI and NNRTI drugs in RT1 amplicon simulated datasets with the expected DRM prevalence of 15% from all samples.

R – Resistant (colored Red), I – Intermediate Resistant (colored Orange), S - Susceptible (colored Green)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Drug class | Sample/Drug | 56252 | 21354 | 63377 | 4134 | 2368 |
| PI | ATV/r | S | I | S | S | I |
| DRV/r | S | S | S | S | S |
| FPV/r | S | S | S | S | I |
| IDV/r | S | I | S | S | I |
| LPV/r | S | I | S | S | S |
| NFV | S | I | S | S | R |
| SQV/r | S | S | S | S | I |
| TPV/r | S | S | S | S | S |
| NRTI | ABC | R | R | R | R | R |
| DDI | R | R | R | R | R |
| FTC | I | I | I | I | I |
| 3TC | I | I | I | I | I |
| D4T | R | R | R | R | R |
| TDF | R | R | R | R | R |
| AZT | I | I | S | I | I |
| NNRTI | EFV | R | R | R | R | R |
| ETR | S | S | S | S | S |
| NVP | R | R | R | R | R |
| RPV | S | S | S | S | S |

Table 4. 10: Resistance calls to PI, NRTI and NNRTI drugs in PR, RT1 and RT2 amplicons simulated datasets with the expected DRM prevalence of 20% and 50% from all samples.

R – Resistant (colored Red), I – Intermediate Resistant (colored Orange), S - Susceptible (colored Green)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Drug class | Sample/Drug | 56252 | 21354 | 63377 | 4134 | 2368 |
| PI | ATV/r | I | R | R | R | I |
| DRV/r | S | I | I | S | S |
| FPV/r | I | R | R | I | I |
| IDV/r | I | R | R | R | I |
| LPV/r | S | R | R | I | I |
| NFV | R | R | R | R | R |
| SQV/r | I | R | R | R | R |
| TPV/r | S | I | I | I | S |
| NRTI | ABC | R | R | R | R | R |
| DDI | R | R | R | R | R |
| FTC | I | I | I | I | I |
| 3TC | I | I | I | I | I |
| D4T | R | R | R | R | R |
| TDF | R | R | R | R | R |
| AZT | I | I | I | I | I |
| NNRTI | EFV | R | R | R | R | R |
| ETR | S | S | S | I | I |
| NVP | R | R | R | R | R |
| RPV | S | S | S | I | I |

Table 4. 11: The selected drug resistant mutations in the simulated datasets within or adjacent to the homopolymer region in HXB2 reference sequence, the wild type sequence and mutated sequence of the drug resistant mutations.

|  |  |  |  |
| --- | --- | --- | --- |
| Drug Class | Drug Resistant Mutation (DRM) | Wild type codon sequence in Homopolymer Region of DRM (bold and underlined) | Mutated codon sequence in Homopolymer Region of DRM (bold and italic) |
| PI | G48V | **GGG**GG | ***GUG***GG |
| PI | I54V | TTTT**AUC**AAA | TTTT***GUC***AAA |
| NRTI | K65R | AAAG**AAA**AAAG | AAAG***AGA***AAAG |
| NNRTI | K103N | AAAAAAG**AAA**AAA | AAAAAAG***AGA***AAA |

Table 4. : The percentage of simulated sequence reads with insertions and deletions in the simulated datasets from Sample 2368.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Drug Resistant Mutation | Number of sequences analyzed | Derived from Sensitive (S) or Resistant (R) “seed” sequence | Number of Insertions (+)/deletions (-) | % of sequences with indels |
| K103N | 9990 | S | +4 | 0.1 |
| 9990 | S | +3 | 0.6 |
| 9990 | S | +2 | 2.47 |
| 9990 | S | +1 | 24.24 |
| 9990 | S | 0 | 70.07 |
| 9990 | S | -1 | 3.06 |
| 9990 | S | -2 | 0.08 |
| **5000** | **R** | **+2** | **0.12** |
| **5000** | **R** | **+1** | **6.64** |
| **5000** | **R** | **0** | **90.94** |
| **5000** | **R** | **-1** | **2.3** |
| K65R | 9990 | S | +2 | 0.35 |
| 9990 | S | +1 | 14.02 |
| 9990 | S | 0 | 83.01 |
| 9990 | S | -1 | 2.09 |
| 9990 | S | -2 | 0.01 |
| 9990 | S | -3 | 0.03 |
| **5000** | **R** | **+1** | **2.34** |
| **5000** | **R** | **0** | **95.54** |
| **5000** | **R** | **-1** | **1.68** |
| **5000** | **R** | **-2** | **0.06** |
| **5000** | **R** | **-3** | **0.02** |
| G48V | 9990 | S | +2 | 0.02 |
| 9990 | S | +1 | 5.21 |
| 9990 | S | 0 | 92.5 |
| 9990 | S | -1 | 1.95 |
| 9990 | S | -3 | 0.03 |
| **5000** | **R** | **+2** | **0.06** |
| **5000** | **R** | **+1** | **5.28** |
| **5000** | **R** | **0** | **92.58** |
| **5000** | **R** | **-1** | **1.74** |
| **5000** | **R** | **-2** | **0.02** |
| I54V | 9990 | S | +1 | 2.21 |
| 9990 | S | 0 | 95.76 |
| 9990 | S | -1 | 1.57 |
| 9990 | S | -2 | 0.01 |
| **5000** | **R** | **+2** | **0.02** |
| **5000** | **R** | **+1** | **2.16** |
| **5000** | **R** | **0** | **95.96** |
| **5000** | **R** | **-1** | **1.46** |

Table 4. : Total number of sequences discarded as non-optimal full-length in the simulated datasets for every sample at all expected prevalences for each DRMs.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Expected  Prevalence (%) | L90M | A71V | K65R | K70R | V75I | F77L | Y115F | F116Y | Q151M |
| 56252 | 0.1 | 0 | 13 | 0 | 0 | 15 | 24 | 13 | 0 | 0 |
| 1 | 1 | 0 | 2 | 0 | 104 | 103 | 103 | 1 | 2 |
| 5 | 15 | 15 | 15 | 3 | 2 | -9 | 6 | 9 | 3 |
| 10 | 35 | 34 | 23 | 4 | 0 | -8 | 6 | 14 | 8 |
| 15 | 39 | 39 | 36 | 10 | 10 | 2 | 6 | 13 | 18 |
| 20 | 48 | 51 | 51 | 10 | 8 | 3 | 18 | 25 | 21 |
| 50 | 111 | 155 | 119 | 26 | 29 | 32 | 45 | 61 | 31 |
| Sample | Expected  Prevalence (%) | G48V | I50V | I54V | V82A | K65R | D67N | K70G | K103N |
| 21354 | 0.1 | 0 | 0 | 11 | 17 | 2 | 17 | 0 | 0 |
| 1 | 2 | 2 | 4 | 105 | 25 | 15 | 2 | 3 |
| 5 | 15 | 7 | 17 | 0 | 143 | 133 | 27 | 28 |
| 10 | 28 | 24 | 37 | 17 | 266 | 260 | 58 | 68 |
| 15 | 41 | 32 | 58 | 26 | 404 | 392 | 76 | 94 |
| 20 | 58 | 44 | 92 | 35 | 592 | 577 | 102 | 124 |
| 50 | 128 | 108 | 208 | 95 | 1406 | 1379 | 269 | 313 |
| Sample | Expected  Prevalence (%) | G48V | I50V | I54V | V82T | L10I | A71V | K65R | D67N | K70G | K103N |
| 63377 | 0.1 | 0 | 0 | 14 | 0 | 0 | 14 | 4 | 16 | 1 | 1 |
| 1 | 2 | 2 | 101 | 3 | 3 | 102 | 34 | 23 | 5 | 6 |
| 5 | 6 | 10 | 18 | 10 | 10 | 3 | 120 | 109 | 27 | 33 |
| 10 | 28 | 24 | 34 | 20 | 19 | 19 | 270 | 261 | 50 | 61 |
| 15 | 39 | 32 | 65 | 36 | 31 | 23 | 413 | 393 | 85 | 103 |
| 20 | 54 | 35 | 84 | 38 | 39 | 31 | 529 | 516 | 96 | 120 |
| 50 | 153 | 121 | 229 | 117 | 117 | 106 | 1322 | 1306 | 252 | 310 |
| Sample | Expected  Prevalence (%) | G48V | I54V | V82A | L10I | K65R | D67N | K70G | K103N | Y181C |
| 4134 | 0.1 | 0 | 14 | 0 | 0 | 1 | 14 | 0 | 0 | 19 |
| 1 | 7 | 4 | 6 | 6 | 24 | 18 | 3 | 3 | 105 |
| 5 | 9 | 17 | 10 | 9 | 103 | 97 | 30 | 38 | 24 |
| 10 | 30 | 38 | 23 | 22 | 244 | 238 | 56 | 68 | 54 |
| 15 | 63 | 68 | 46 | 44 | 313 | 298 | 73 | 87 | 79 |
| 20 | 75 | 100 | 51 | 54 | 429 | 413 | 91 | 11 | 95 |
| 50 | 161 | 232 | 118 | 123 | 1138 | 1117 | 273 | 322 | 298 |
| Sample | Expected  Prevalence (%) | L90M | A71V | G73S | M41L | A62V | K65R | V75I | K103N | Y181C |
| 2368 | 0.1 | 0 | 16 | 1 | 11 | 2 | 2 | 15 | 1 | 22 |
| 1 | 2 | 104 | 3 | 18 | 20 | 22 | 102 | 7 | 108 |
| 5 | 10 | 3 | 13 | 115 | 117 | 124 | 24 | 57 | 17 |
| 10 | 16 | 7 | 17 | 230 | 232 | 231 | 57 | 127 | 53 |
| 15 | 23 | 10 | 28 | 311 | 309 | 319 | 81 | 162 | 82 |
| 20 | 47 | 33 | 48 | 472 | 472 | 484 | 138 | 245 | 127 |
| 50 | 122 | 87 | 106 | 1112 | 1112 | 1152 | 275 | 560 | 267 |

Table 4. : Total number of sequence reads discarded as non full-length after reference mapping in the amplicons of the samples at all expected prevalences.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Amplicon | Prevalence (%) | 56252 | 21354 | 63377 | 4134 | 2368 |
| PR | 0.1 | 172 | 192 | 178 | 186 | 183 |
| 1 | 172 | 193 | 180 | 192 | 182 |
| 5 | 175 | 188 | 170 | 188 | 181 |
| 10 | 180 | 187 | 170 | 195 | 177 |
| 15 | 179 | 190 | 185 | 195 | 160 |
| 20 | 174 | 192 | 155 | 196 | 183 |
| 50 | 150 | 170 | 172 | 192 | 163 |
| RT1 | 0.1 | 2378 | 2838 | 2802 | 2309 | 2854 |
| 1 | 2356 | 2830 | 2804 | 2312 | 2846 |
| 5 | 2261 | 2837 | 2775 | 2295 | 2826 |
| 10 | 2136 | 2800 | 2799 | 2306 | 2800 |
| 15 | 2014 | 2809 | 2772 | 2255 | 2742 |
| 20 | 1925 | 2834 | 2739 | 2254 | 2739 |
| 50 | 1192 | 2761 | 2707 | 2275 | 2525 |
| RT2 | 0.1 | 0 | 501 | 478 | 479 | 516 |
| 1 | 0 | 494 | 479 | 475 | 514 |
| 5 | 0 | 501 | 485 | 482 | 515 |
| 10 | 0 | 503 | 484 | 480 | 524 |
| 15 | 0 | 501 | 489 | 476 | 525 |
| 20 | 0 | 483 | 480 | 463 | 539 |
| 50 | 0 | 496 | 475 | 504 | 503 |