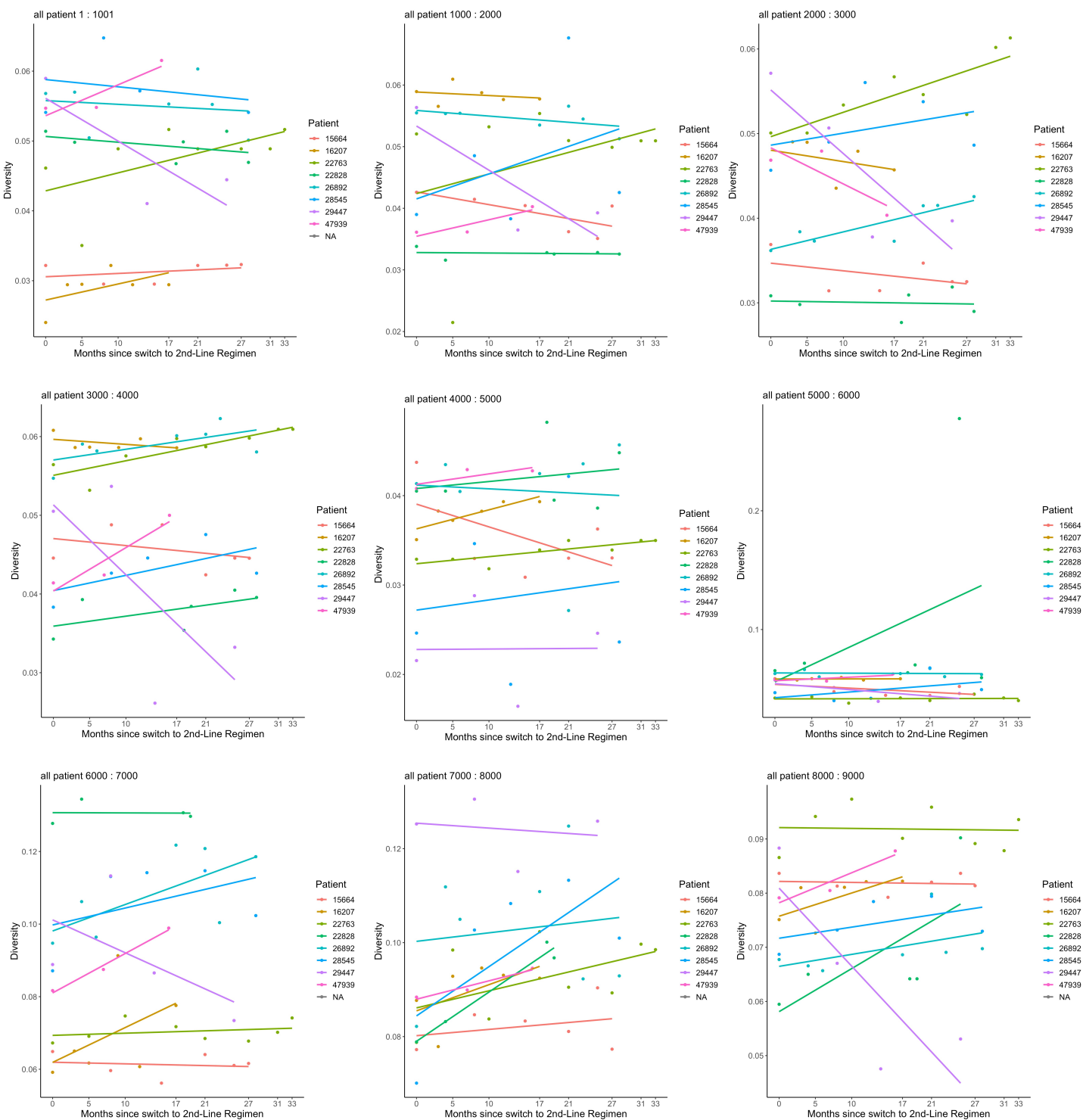
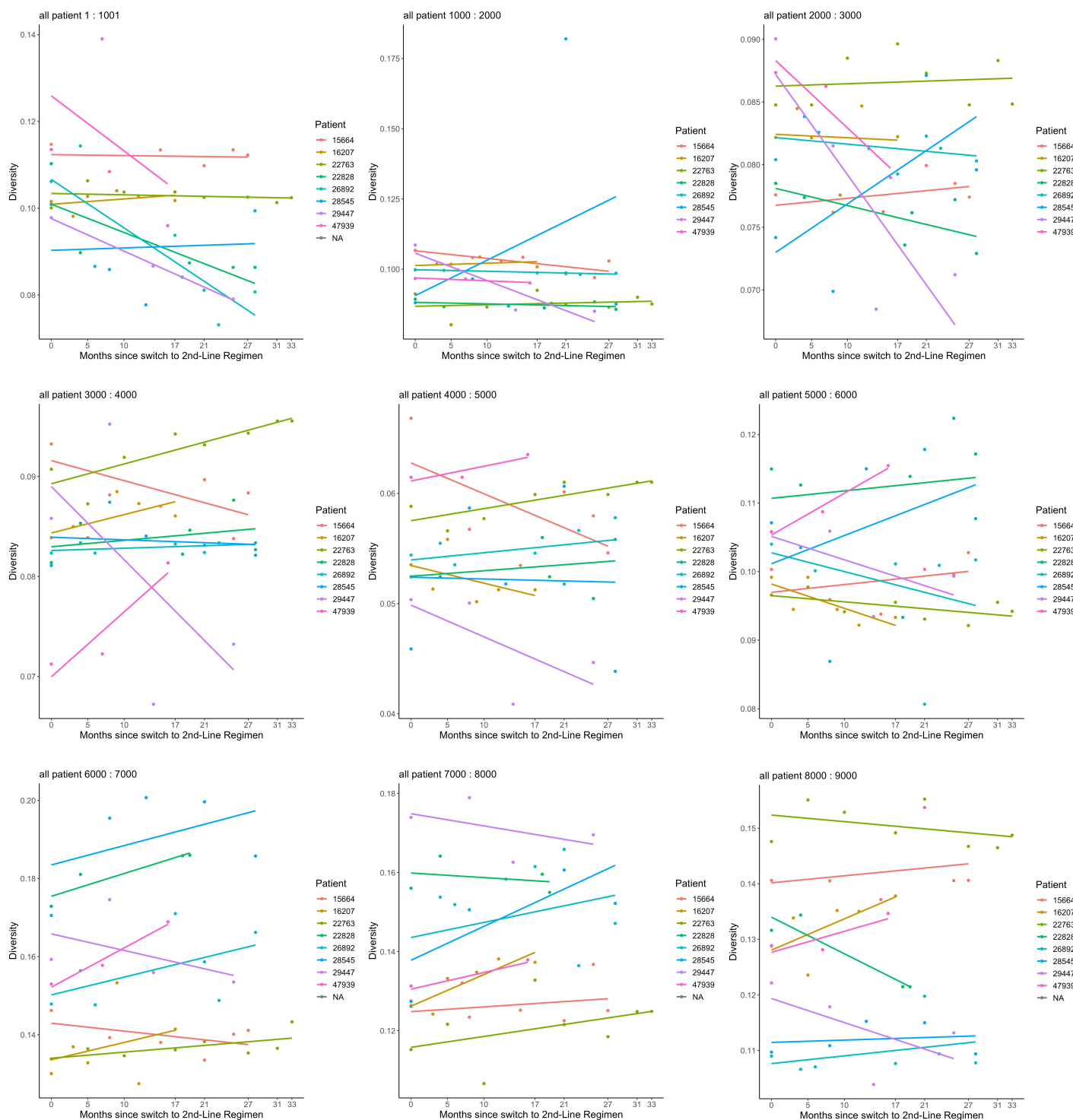


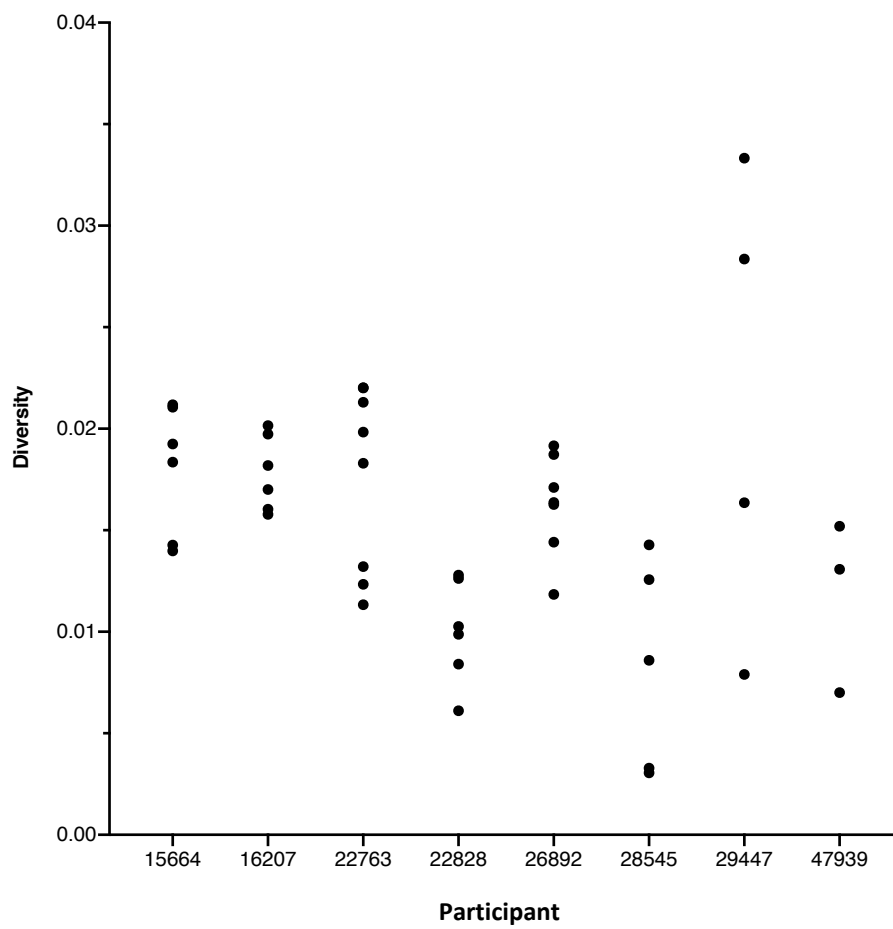
Supplementary Figure 1. Measure of divergence from the SNPs relative to the infecting (baseline) strain at successive timepoints in eight individuals under non-suppressive ART. These data were for SNPs detected by Illumina NGS at a threshold of 2% abundance. Sites in this analysis had coverage of at least 10 reads.



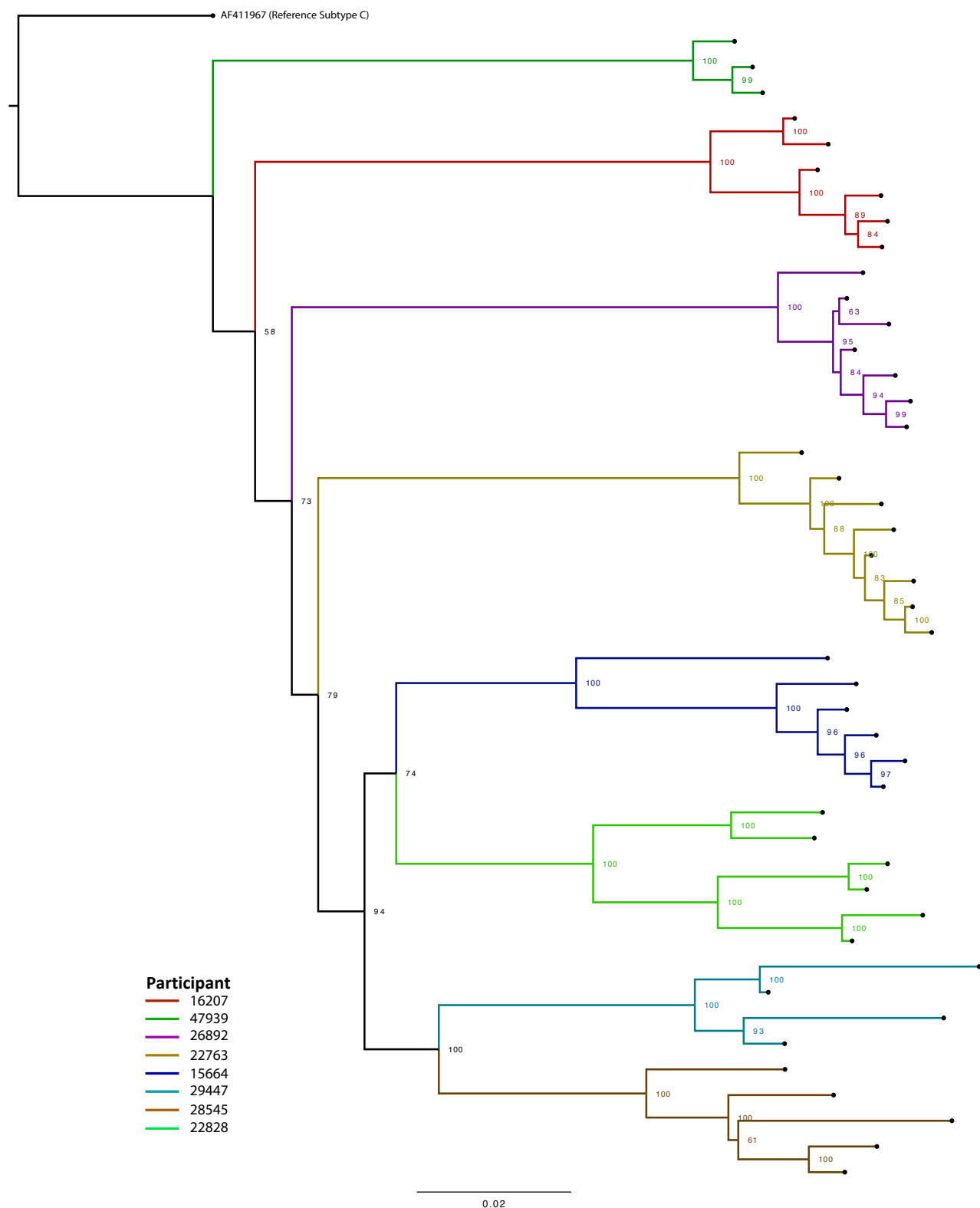
Supplementary Figure 2. Measure of divergence from the SNPs relative to the ancestral Subtype C strain at successive timepoints in eight individuals under non-suppressive ART. These data were for SNPs detected by Illumina NGS at a threshold of 2% abundance. Sites in this analysis had coverage of at least 10 reads.



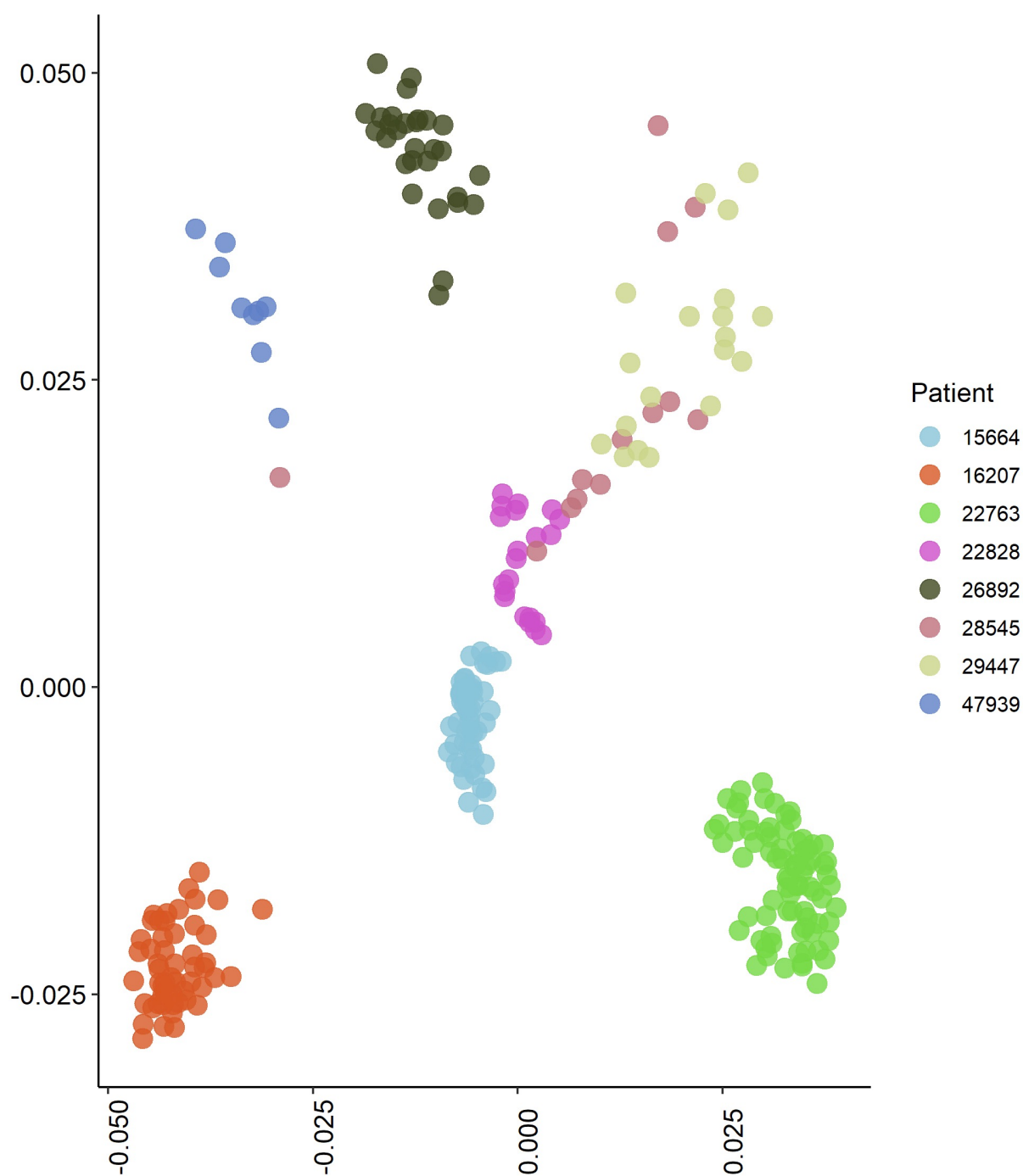
Supplementary Figure 3. Measure of divergence from the SNPs relative to the ancestral subtype M strain at successive timepoints in eight individuals under non-suppressive ART. These data were for SNPs detected by Illumina NGS at a threshold of 2% abundance. Sites in this analysis had coverage of at least 10 reads.



Supplementary Figure 4. Whole-genome nucleotide diversity of longitudinal timepoints from each patient. Diversity was calculated using all information from BAM files by fitting observed variant frequencies to two distributions (a β -distribution and 4D Dirichlet plus Δ function). Each dot in the scatter represents a different timepoint and highlights differences in whole-genome diversity between successive timepoints.

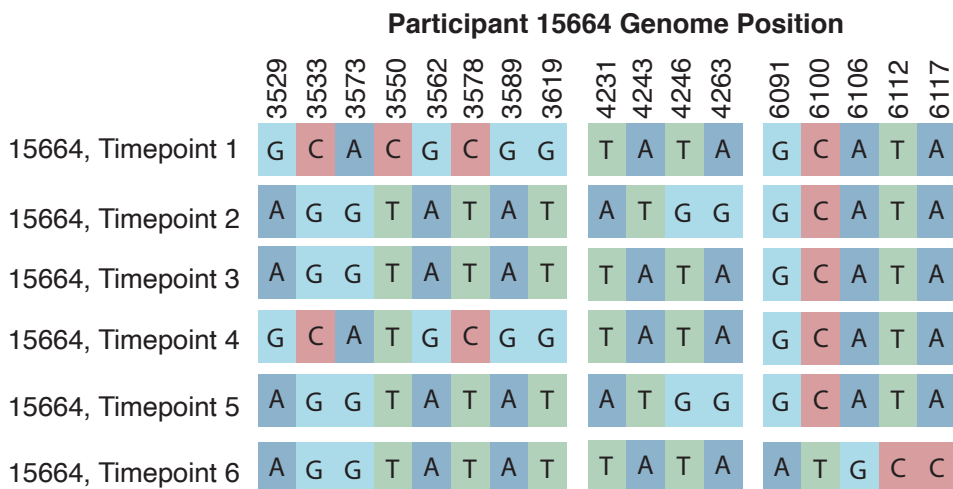


Supplementary Figure 5. Maximum likelihood phylogeny of consensus sequences from all timepoints from all patients. Phylogenies were rooted on a South-African origin subtype C reference genome, AF411967. Trees were inferred with a GTR model with 1000 bootstrap replicates. Bootstrap values are indicated at all nodes. The phylogeny is largely consistent with the haplotype tree shown in Figure 1C suggesting that haplotypes were accurate representations of sequences.

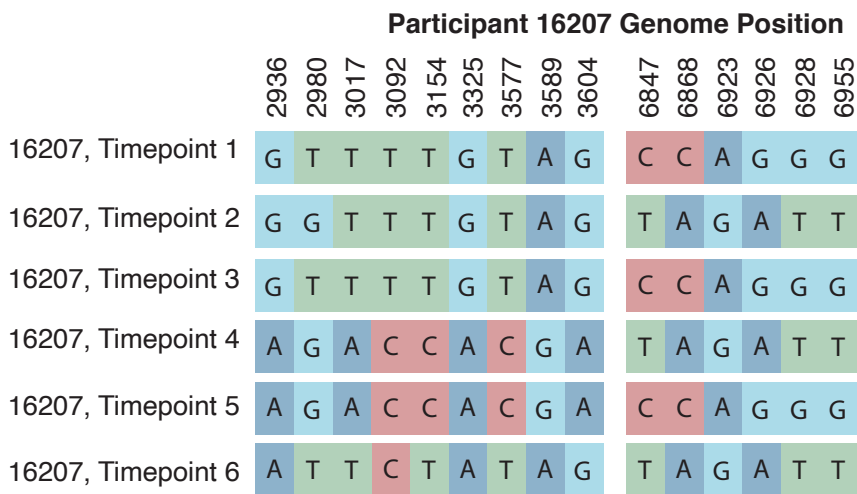


Supplementary Figure 6. MDS scatterplot of reconstructed haplotypes. Plots were produced by obtaining a multiple sequence alignment, calculating average pairwise distances between all pairs and then multi-dimensional scaling under a TN93 substitution matrix. Each axis represents the component scores of the most variable axis and the second-most variable axis. Haplotypes show an increased measure of diversity compared to consensus-level variants. This is due to increased resolution of potential viral quasispecies.

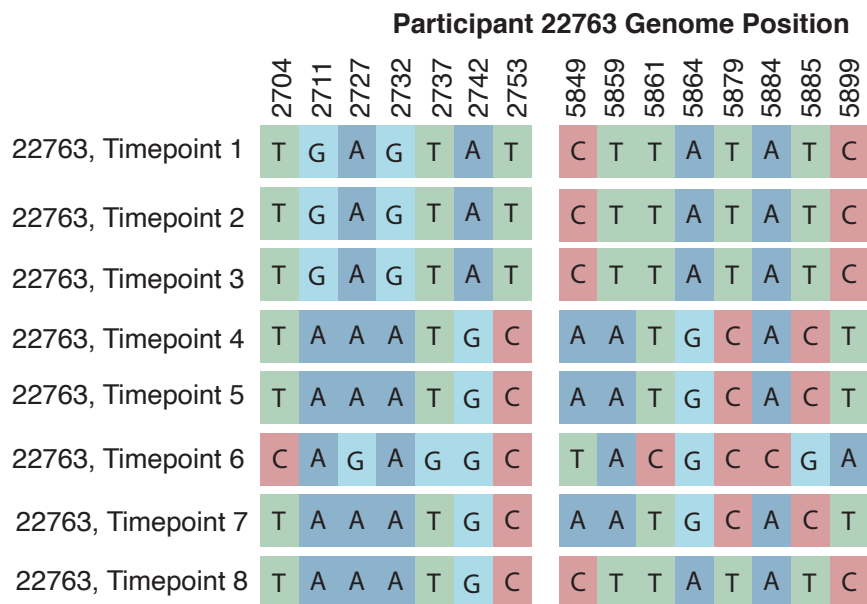
A)



B)



C)



Supplementary Figure 7. Patterns of SNPs at perceived recombination breakpoint locations. In all participants where there was recombination detected, there were distinct patterns or haplotypes observable across multiple sites. Distinct patterns were observable between recombination breakpoints, lending support to the theory that recombination between different genomes was occurring, to give rise to numerous haplotypes. In numerical terms where each number represents an individual pattern (i.e. 0, 1 or 2), Participant 15664 haplotype pattern 1, (A, Left), assumes a 011011 distribution, pattern 2 (A, middle) as 0100010 and pattern 3 (A, right) as 000001. Participant 16207, pattern one (B, Left) is 010223 and pattern 2 (B, right) is 010101. Participant 22763, pattern one (C, left) is 00011211 and pattern 2 (C, right) is 00011210.