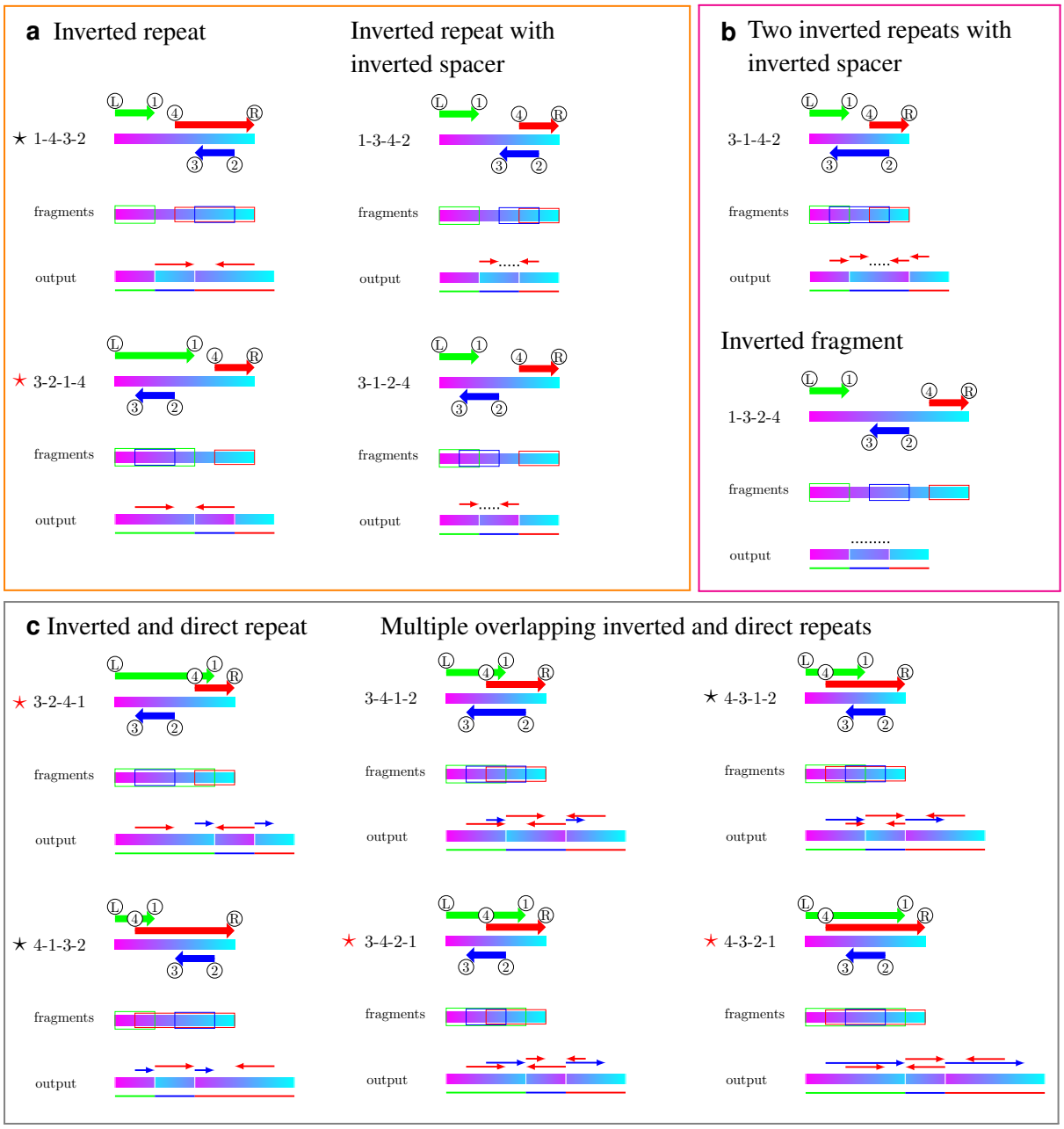
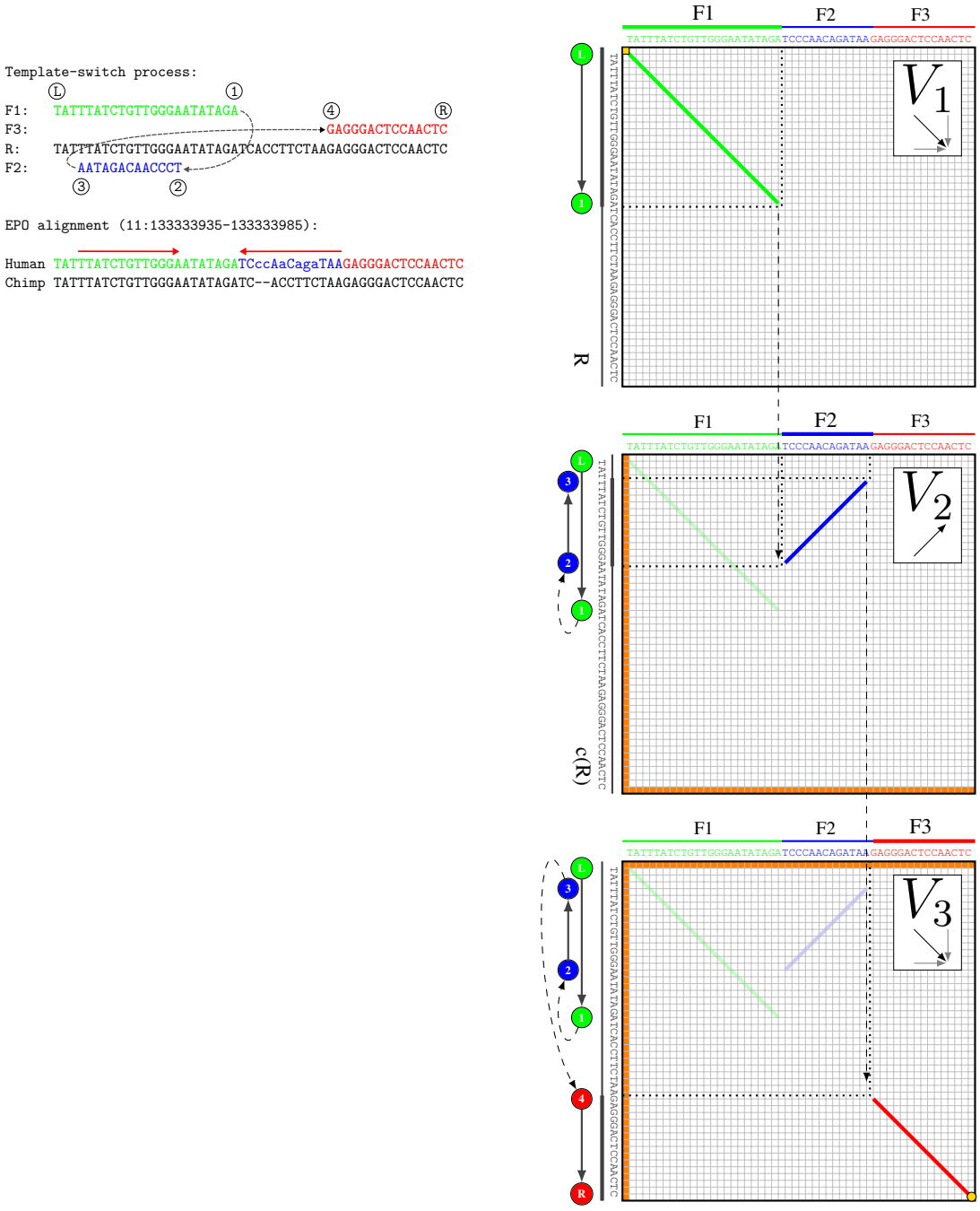
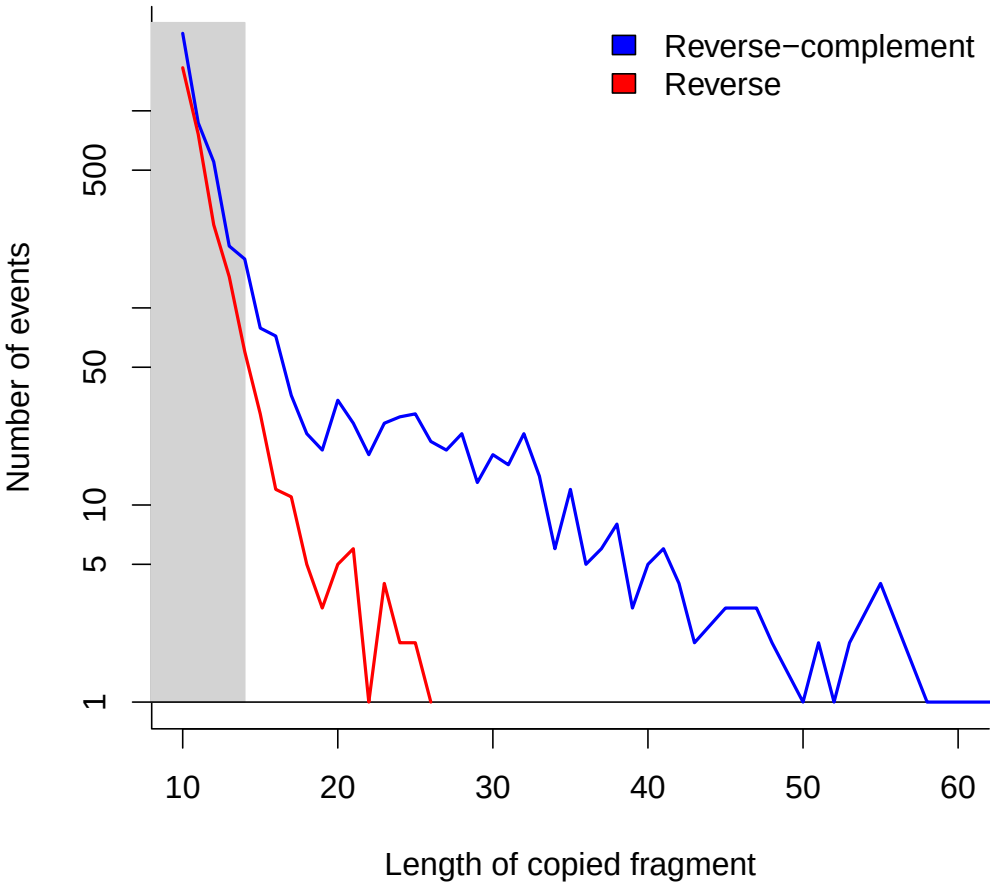


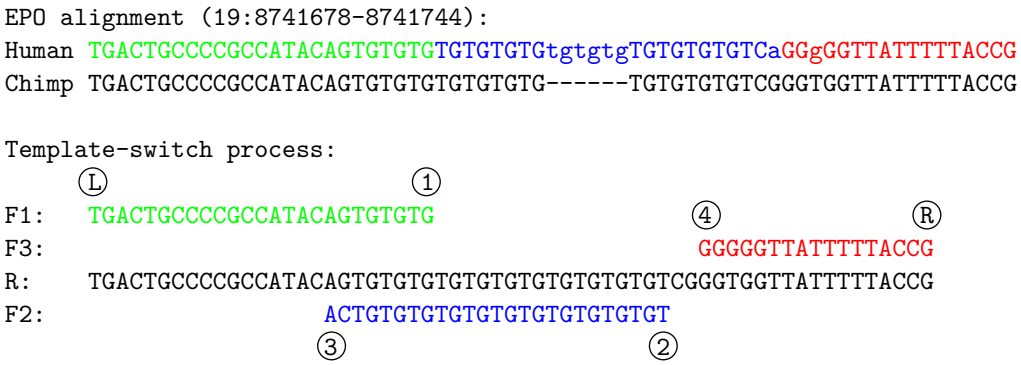
Supplemental Fig. S1: All possible arrangements of switch points and the repeats they create. For each of the 12 possible cases (e.g. "1-4-3-2"), the green, blue and red arrows at top indicate the three replication fragments (labelled Ⓛ→①, ③←②, ④→Ⓡ, respectively), defined by the four points ①–④, and show the process of copying the sequence from the template indicated by the magenta-cyan bar. Below that, the fragments are projected onto the template with matching colors (middle) and then concatenated to create the complete replication output (bottom). Newly created inverted repeats are shown with red arrows, new direct repeats are shown with blue arrows, and inverted sequence fragments are indicated with dotted lines. Some event types are 'mirror cases': "1-4-3-2" becomes "3-2-1-4" if replication progresses from right to left, and similarly for the pairs "1-3-4-2" and "3-1-2-4", "3-2-4-1" and "4-1-3-2", and "3-4-2-1" and "4-3-1-2". **a**, **b**, The first six event types are observed in the human-chimp and human-human comparisons. Of these, the first four (**a**; accounting for slightly >50% of observed cases) are consistent with the model previously proposed for bacteria whereas the last two (**b**; nearly 50% of observed cases) are novel and only achievable under our new four-point model. **c**, In the cases not observed in the data, ④ precedes ① and the second template switch would require opening of the newly synthesized DNA double helix. Event types that could happen via intra-strand switches are indicated with red stars, their mirror cases with black stars. All other events can only happen inter-strand.



Supplemental Fig. S2: Example of four-point alignment. The optimal template switch path is found using a dynamic programming algorithm consisting of three matrices, V_1, V_2 and V_3. Two edges of matrices V_2 and V_3 are initialised to $-\infty$ (shown in orange) and the top-left corner of V_1 is set to 0 (gold square). The matrices are then filled by choosing, from a set of possible moves, the move that maximises the score. V_1 and V_3 are filled from top-left corner and the moves can be either base matches (diagonal moves) or indels (horizontal and vertical moves); V_2 is filled from bottom-left corner and only base matches are allowed. The score for a base match is positive if the two bases (in the end of the corresponding row and column) are identical, otherwise negative; the score for an indel is always negative. (Note that in V_2 the reference sequence R is complemented—denoted c(R)—but the scoring is otherwise similar.) In addition to moves within a matrix, jumps from V_1 to V_2 and from V_2 to V_3 are possible: unlike moves within a matrix, these jumps can be from any cell in the previous column of the preceding matrix. The score for the optimal solution is found in the bottom-right corner of V_3 (gold circle). The optimal template switch path is found by back-tracking the optimal dynamic programming path, including the jumps between the matrices. See also Supplemental Algorithm S1.

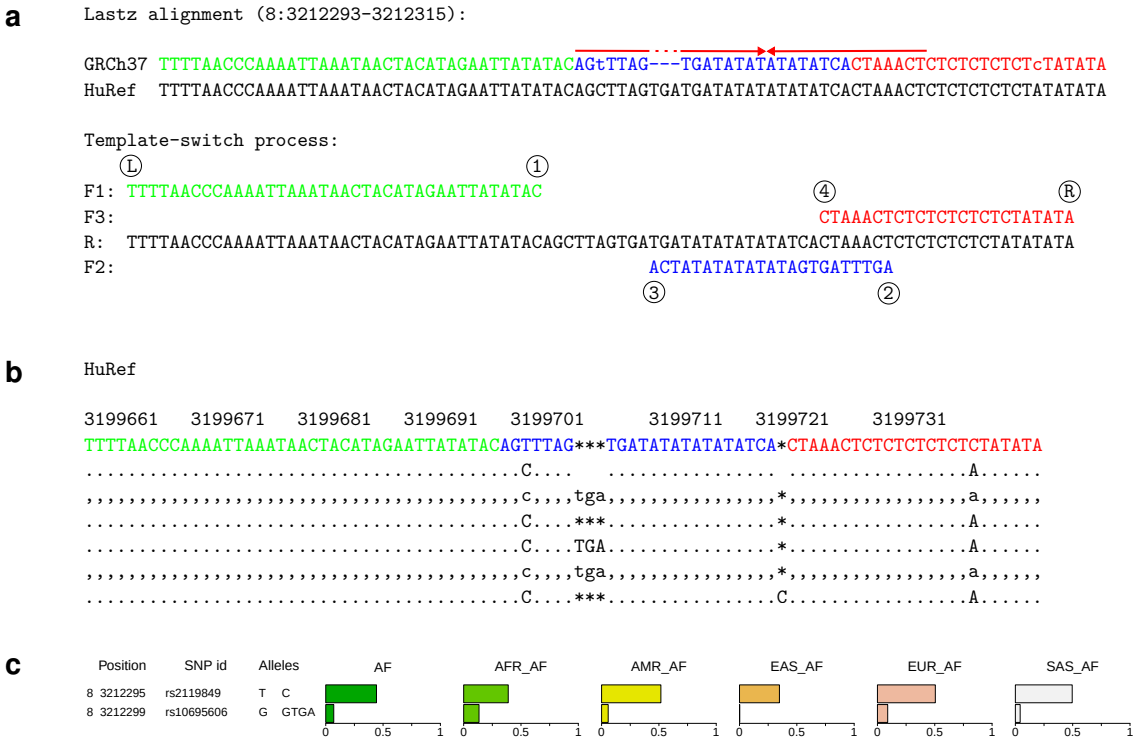


Supplemental Fig. S3: Length distribution of copied fragments. The length of the ②→③ fragment of true candidate events (fragment ②→③ copied in reverse-complement; blue) follows a logarithmic distribution with a heavy tail. The longest ②→③ fragment is 547 bases, close to the upper limit of the detection method used. In the control distribution (apparent reverse repeats, see Methods; red), short events occur at a relatively high frequency but events of 15 bases or longer are rare. To improve the signal to noise ratio, we discarded true candidate events shorter than 14 bases (shaded area) and focused on the remaining 794 cases. Note the logarithmic scale on the *y*-axis.



Supplemental Fig. S4: An example of an apparent reverse (non-complemented) copy event. As a control to assist in assessing the occurrence of false positives, we repeated the analysis without complementing the ②→③ fragment. There are few events longer than 15 bases (Supplemental Fig. S3) and, as with this example, reverse hits longer than 12–13 bases typically overlap with low-complexity sequence that was not detected by the sequence masking. Notice the two apparent substitutions, located 11 and 14 bases 3′ to the gapped region in the top alignment, which suggest this may not be a simple slippage event.

Original data:
http://grch37.ensembl.org/Homo%5Fsapiens/Location/Compara%5FAlignments?align=548&r=19:8741678-8741744

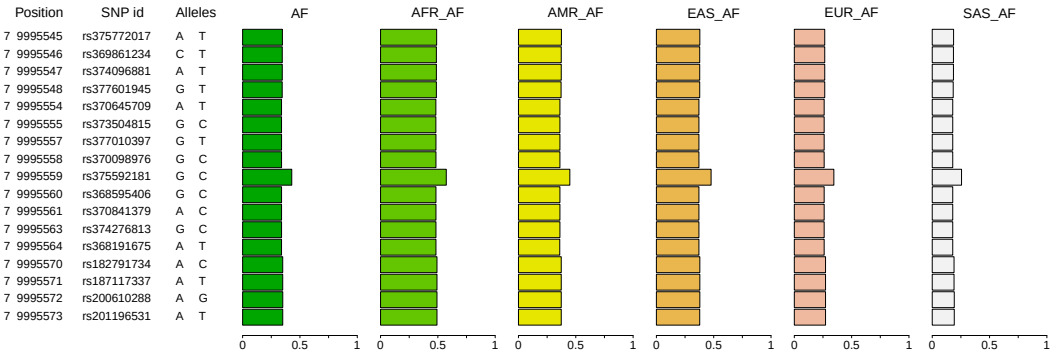


Supplemental Fig. S5: Independent mutation events. **a**, A mutation pattern that can be explained as a template switch event. **b**, Sequence reads show that HuRef is homozygous for the first variant and heterozygous for the second, indicating independent evolutionary histories. **c**, The 1KG variation data confirm that the two events are independent.

a

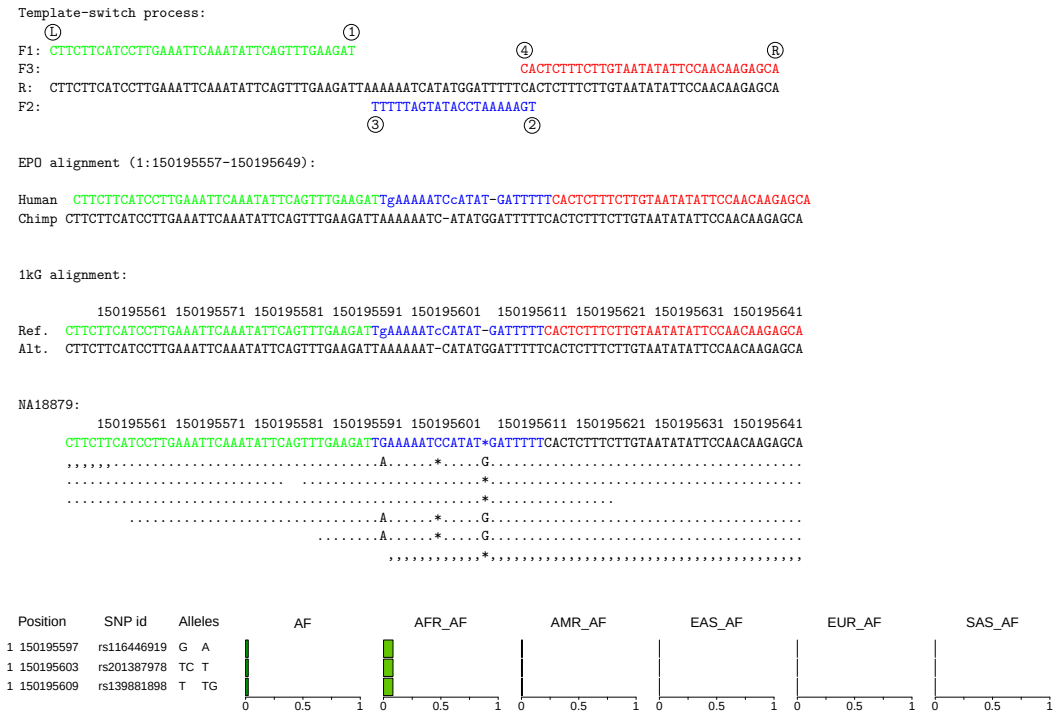
```
Lastz alignment (7:9995542-9995576):

GRCh37 TTAATCACACTAGTCAAATTCTACGTACCCAGTTTTGATAAacagCTTTGagTggggaAgaCAAAGaaaaTTAACATTTATTGAGTACATGAACTCTGATAGATATTAAT
HuRef  TTAATCACACTAGTCAAATTCTACGTACCCAGTTTTGATAATTTTCTTTGTCTTCCCCACTCAAAGCTGTTTAACATTTATTGAGTACATGAACTCTGATAGATATTAAT

Template-switch process:
   Ⓛ                                    ①
F1: TTAATCACACTAGTCAAATTCTACGTACCCAGTTTTGA                                    ④                                    Ⓡ
F3:                                                                          ACATTTATTGAGTACATGAACTCTGATAGATATTAAT
R:  TTAATCACACTAGTCAAATTCTACGTACCCAGTTTTGATAATTTTCTTTGTCTTCCCCACTCAAAGCTGTTTAACATTTATTGAGTACATGAACTCTGATAGATATTAAT
F2:                                      ATTAAAAGAAACAGAAGGGGTGAGTTTCGACAAAT
                                         ③                                   ②
```

b

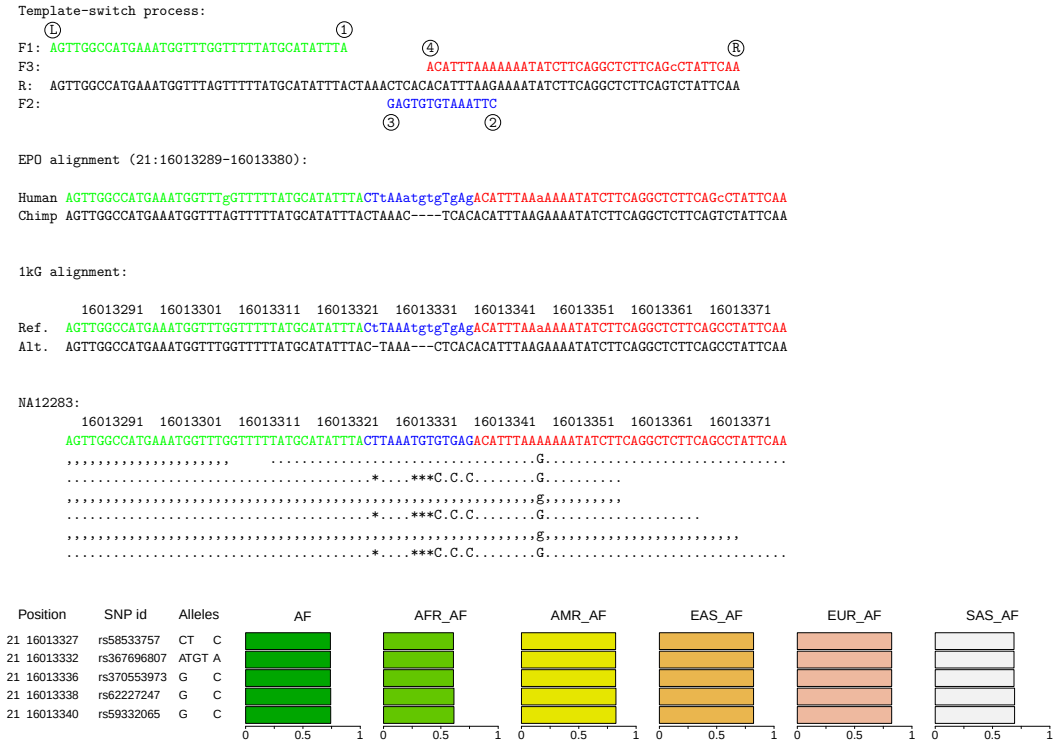


Supplemental Fig. S6: Complex mutation explained with 17 SNPs in in 1KG data. **a**, A mutation pattern explained by a template switch event. **b**, 1KG variation data report the event as 17 independent SNPs.

a



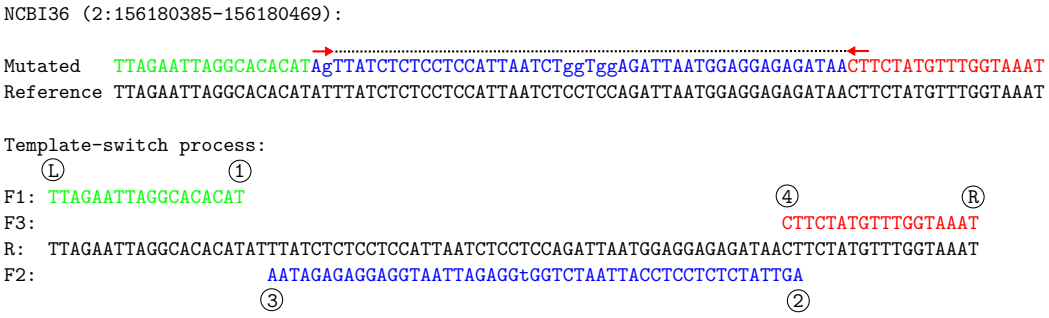
b



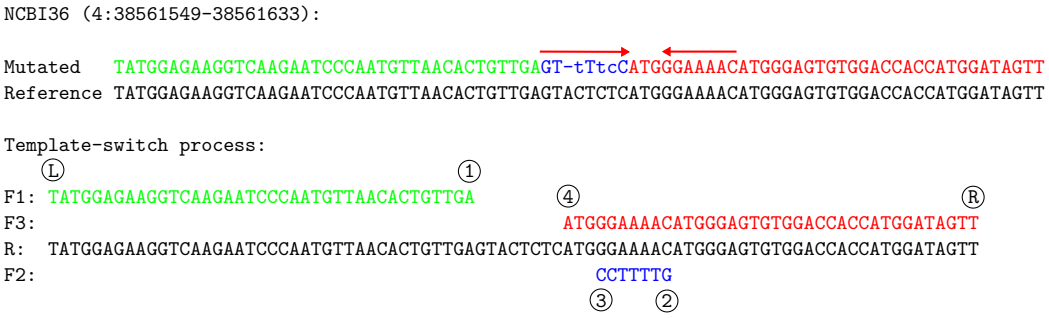
Supplemental Fig. S7: Complex mutations found in EPO and 1KG data. **a**, A mutation pattern explained by a template switch event that is present at uniformly low frequencies in 1KG population data. **b**, A mutation pattern that is present at uniformly high frequencies in 1KG population data.

a: http://grch37.ensembl.org/Homo%5Fsapiens/Location/Compara%5FAlignments?align=548&r=1:150195557-150195649
b: http://grch37.ensembl.org/Homo%5Fsapiens/Location/Compara%5FAlignments?align=548&r=21:16013289-16013380

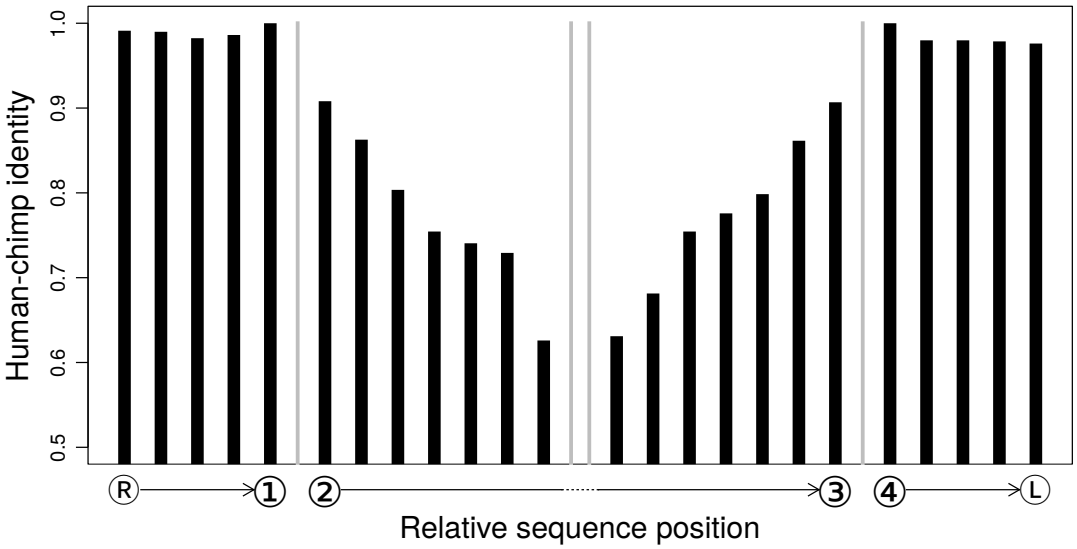
a



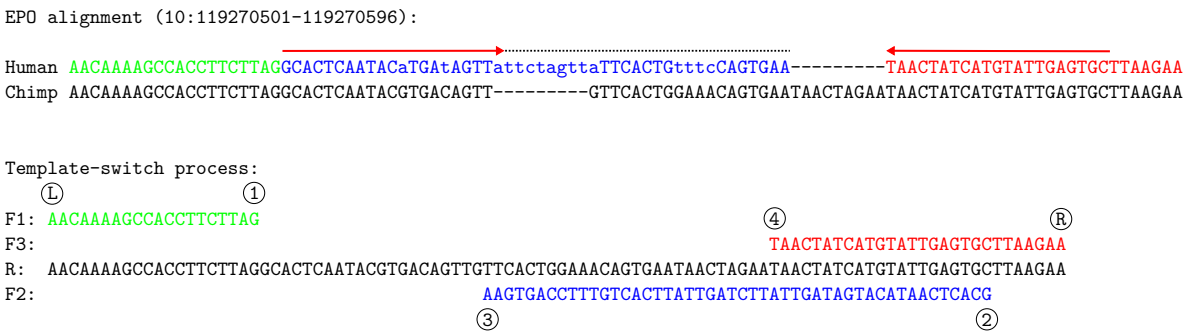
b



Supplemental Fig. S8: Complex *de novo* mutations found in Icelandic trios. a, A *de novo* mutation found in a human parent-offspring trio (Besenbacher et al., 2016) that can be explained by a "1-3-4-2" template switch event and one additional mismatch, indicated by a lower-case letter in F2. **b**, A *de novo* mutation that can be fully explained by a "1-4-3-2" template switch event.



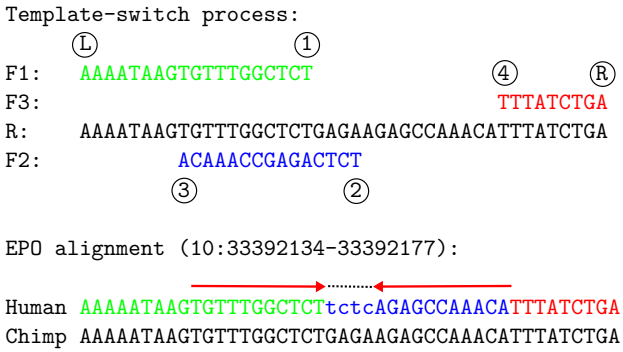
Supplemental Fig. S9: Sequence identity around template switch sites. While the algorithm ensures that positions ① and ④ are always identical between the two sequences, upstream and downstream sites may differ and the events we identified have multiple differences in the intervening fragment. Dissimilarities in the inserted fragment ②→③ are not evenly distributed and the sequence identity decreases with distance from the switch site. The higher sequence similarity at the start and end of the inserted fragment is consistent with very short regions of sequence identity observed at switch sites of major genomic rearrangements. The plot is based on the 794 high-confidence template-switch events detected in the human-chimp comparisons.



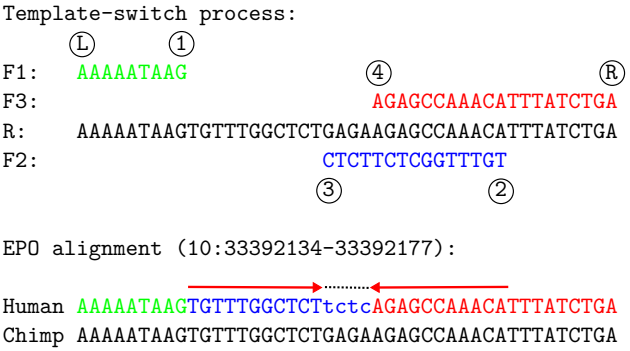
Supplemental Fig. S10: Example of indels appearing mutagenic. An example of a template switch event that does not change the sequence length but causes indels in the linear alignment. Indels (dashes) are associated with base mismatches (lower case bases), potentially giving the impression that indels are mutagenic.

Original data:
http:/ /grch37.ensembl.org/Homo%5Fsapiens/Location/Compara%5FAlignments?align=548&r=10:119270501-119270596

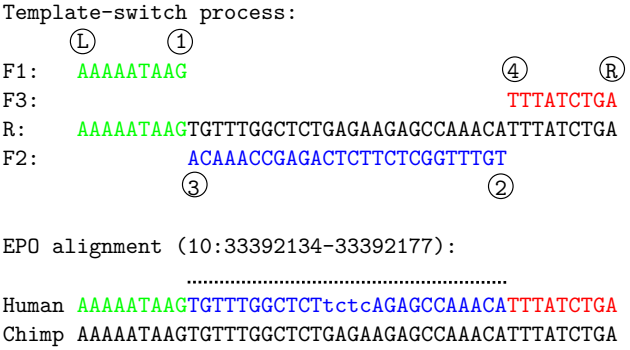
a



b



c



Supplemental Fig. S11: Uncertainty when the event overlaps an existing inverted repeat. An inversion of the spacer of an existing inverted repeat does not allow distinguishing the event type as there are multiple equally good placements for the switch points. **a–c**, Three possible locations for the switch points ①–④ give identical outputs matching the chimp reference sequence for this region on chromosome 10. The approach used maximizes the length of fragment ②→③, and reports this particular event as "1-3-2-4" (**c**).

Original data:
http://grch37.ensembl.org/Homo%5Fsapiens/Location/Compara%5FAlignments?align=548&r=10:33392134-33392177