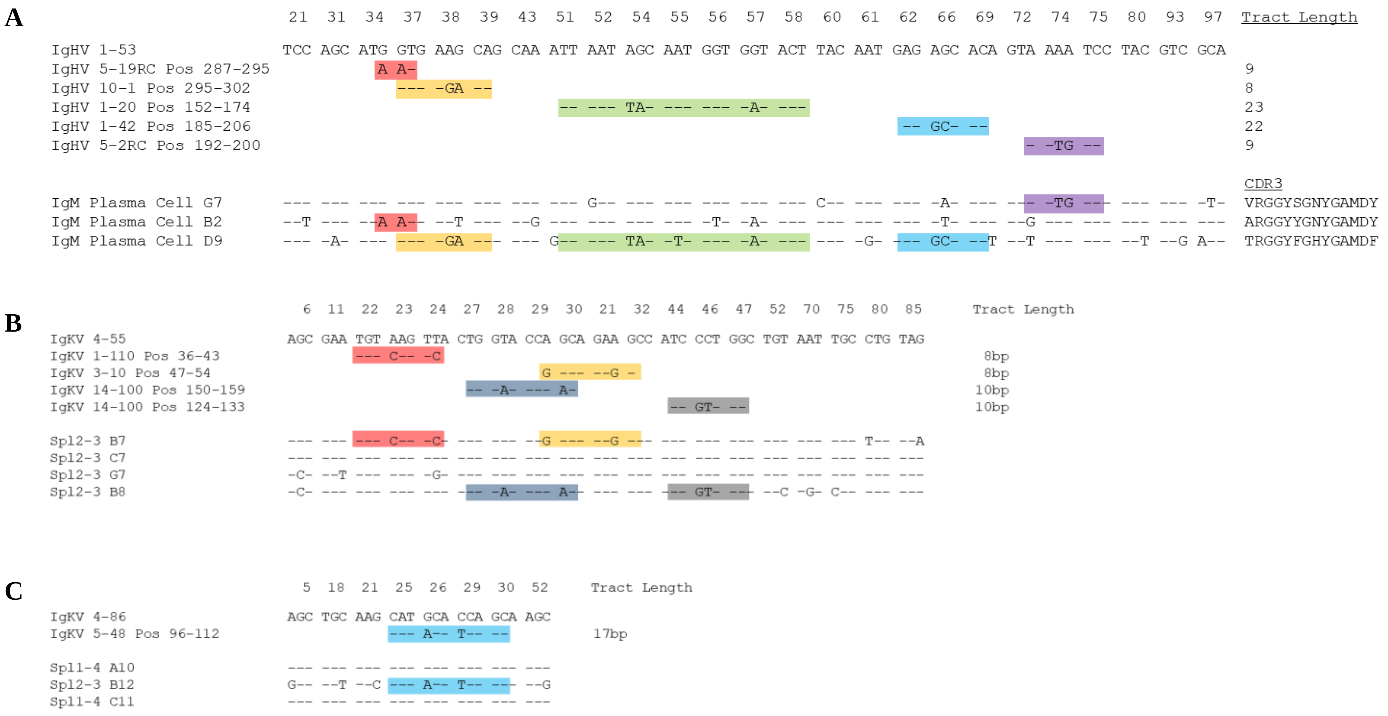
Figure S1



A

```
                              21  31  34  37  38  39  43  51  52  54  55  56  57  58  60  61  62  66  69  72  74  75  80  93  97  Tract Length

IgHV 1-53                    TCC AGC ATG GTG AAG CAG CAA ATT AAT AGC AAT GGT GGT ACT TAC AAT GAG AGC ACA GTA AAA TCC TAC GTC GCA
IgHV 5-19RC Pos 287-295                A A-                                                                                       9
IgHV 10-1 Pos 295-302                      --- -GA --                                                                             8
IgHV 1-20 Pos 152-174                                           -- --- TA- --- --- -A- ---                                        23
IgHV 1-42 Pos 185-206                                                                               -- GC- --                     22
IgHV 5-2RC Pos 192-200                                                                                           - -TG --         9

                                                                                                                                  CDR3
IgM Plasma Cell G7           --- --- --- --- --- --- --- --- G-- --- --- --- --- --- C-- --- --- -A- --- --- -TG --- --- --- -T-  VRGGYSGNYGAMDY
IgM Plasma Cell B2           --T --- --A A-- --T --- --G --- --- --- --- -T- -A- --- --- --- --- -T- --- --G --- --- --- --- ---  ARGGYYGNYGAMDY
IgM Plasma Cell D9           --- -A- --- --- -GA --- --- G-- --- TA- -T- --- -A- --- --- -G- --- GC- --T --T --- --- --T --G A--  TRGGYFGHYGAMDF
```

B

```
                               6  11  22  23  24  27  28  29  30  21  32  44  46  47  52  70  75  80  85        Tract Length

IgKV 4-55                    AGC GAA TGT AAG TTA CTG GTA CCA GCA GAA GCC ATC CCT GGC TGT AAT TGC CTG TAG
IgKV 1-110 Pos 36-43                 --- C-- -C                                                                  8bp
IgKV 3-10 Pos 47-54                                        G --- --G -                                           8bp
IgKV 14-100 Pos 150-159                               -- -A- --- A-                                             10bp
IgKV 14-100 Pos 124-133                                                   -- GT- --                             10bp

Spl2-3 B7                    --- --- --- C-- -C- --- --- --G --- --G --- --- --- --- --- --- --- T-- --A
Spl2-3 C7                    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Spl2-3 G7                    -C- --T --- --- -G- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Spl2-3 B8                    -C- --- --- --- --- --- -A- --- A-- --- --- --- GT- --- --C -G- C-- --- ---
```

C

```
                               5  18  21  25  26  29  30  52      Tract Length

IgKV 4-86                    AGC TGC AAG CAT GCA CCA GCA AGC
IgKV 5-48 Pos 96-112                     --- A-- T-- --            17bp

Spl1-4 A10                   --- --- --- --- --- --- --- ---
Spl2-3 B12                   G-- --T --C --- A-- T-- --- --G
Spl1-4 C11                   --- --- --- --- --- --- --- ---
```

Fig. S1: Templated tracts of mutations are observed in sanger sequenced IgM plasma cells. (**A**) Nucleotide alignment of somatically-mutated IgM plasma cell sequences to germline IgHV 1-53. (**B-C**) Nucleotide alignment of somatically-mutated IgKV sequences from IgM plasma cells to their respective germline sequence. Data is presented as in Figure 1C.

Figure S2

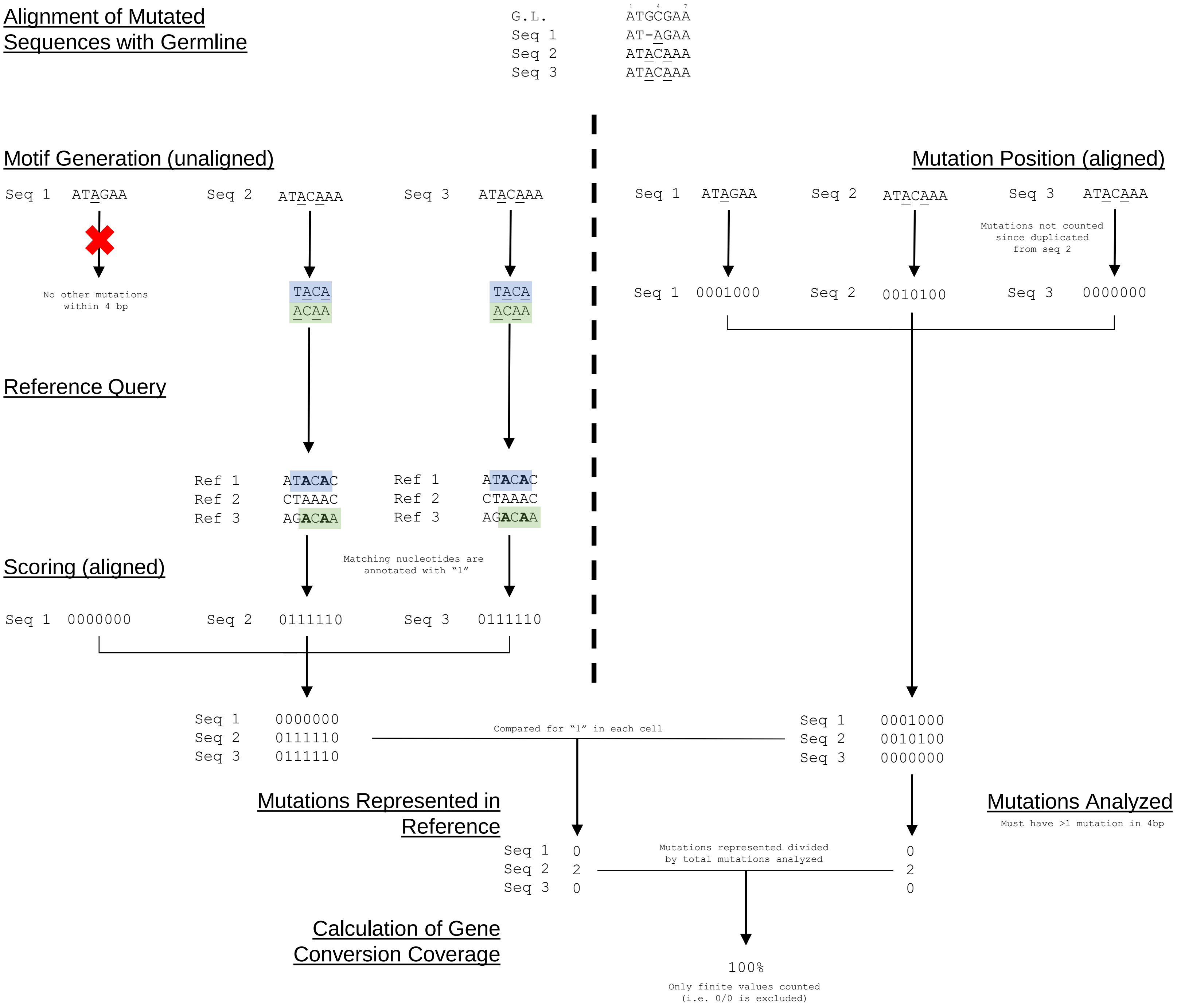

Alignment of Mutated Sequences with Germline
G.L. ATGCGAA
1 4 7
Seq 1 AT-AGAA
Seq 2 ATACAAA
Seq 3 ATACAAA
Motif Generation (unaligned)
Seq 1 ATAGAA
Seq 2 ATACAAA
Seq 3 ATACAAA
No other mutations within 4 bp
TACA
ACAA
TACA
ACAA
Reference Query
Ref 1 ATACAC
Ref 2 CTAAAC
Ref 3 AGACAA
Ref 1 ATACAC
Ref 2 CTAAAC
Ref 3 AGACAA
Matching nucleotides are annotated with "1"
Scoring (aligned)
Seq 1 0000000
Seq 2 0111110
Seq 3 0111110
Mutation Position (aligned)
Seq 1 ATAGAA
Seq 2 ATACAAA
Seq 3 ATACAAA
Mutations not counted since duplicated from seq 2
Seq 1 0001000
Seq 2 0010100
Seq 3 0000000
Seq 1 0000000
Seq 2 0111110
Seq 3 0111110
Compared for "1" in each cell
Seq 1 0001000
Seq 2 0010100
Seq 3 0000000
Mutations Represented in Reference
Mutations Analyzed
Must have >1 mutation in 4bp
Seq 1 0
Seq 2 2
Seq 3 0
Mutations represented divided by total mutations analyzed
0
2
0
Calculation of Gene Conversion Coverage
100%
Only finite values counted (i.e. 0/0 is excluded)

Fig. S2: Schematic of sequence analysis by PolyMotifFinder. Depicted are three somatically-mutated sequences aligned to a reference gene. Initially, mutation positions are defined by comparing the identity of nucleotides within the sequence to the reference sequence. Once positions of mutations are found, PolyMotifFinder creates a numeric array whose length is equal to the length of sequences analyzed and height is equal to the number of sequences analyzed. This array is filled with either "0" to denote a position in which the sequence has the germline nucleotide at a given position, or "1" should the sequence differ from germline at that position. Importantly, if a pair of mutations matches the position and identity of another pair of mutations already marked in the array, these mutations will be marked with "0", such that identical pairs of mutations are excluded from analysis. Next, PolyMotifFinder will generate motifs of k-mer length. Here k=4, whereas in our analyses k=8. These motifs must contain two or more mutations over their length. The generated motifs are then compared to a reference set of sequences for matches. If a match is found, another numeric array is annotated with "1" over the length of the motif that matched a reference, otherwise the array is annotated with "0". Each row of the mutation position array is compared to the respective row of the motif matched array. Corresponding cells are compared for matches in which both cells contain "1", denoting a mutation that was part of a motif that matched the reference sequence. These mutation matches are tallied and divided by the number of mutations within that sequence to generate a gene conversion (GC) coverage value.

Figure S3

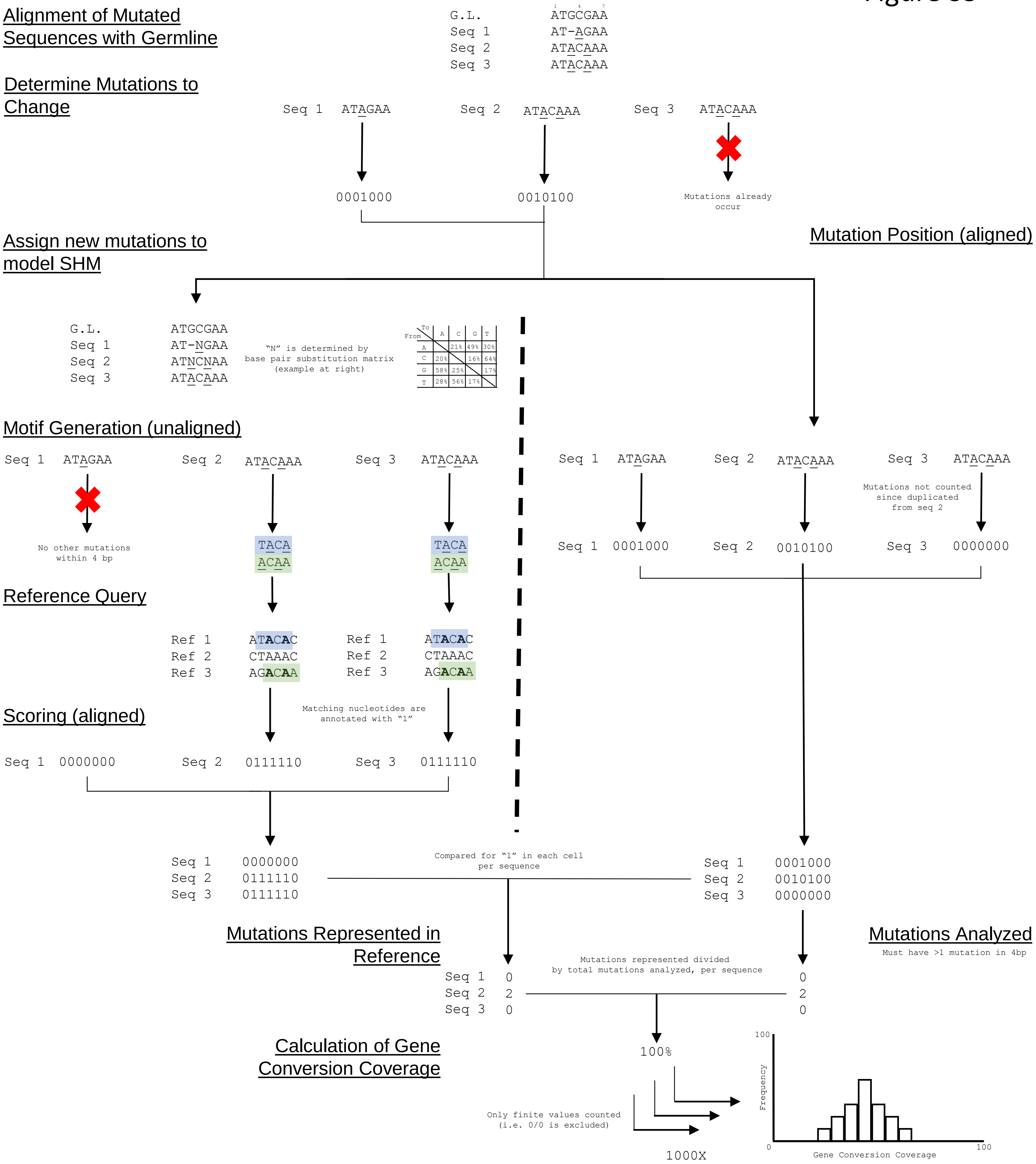


Fig. S3: Schematic of sequence analysis via RandomCheck. Depicted are three somatically-mutated sequences aligned to a reference gene. As in PolyMotifFinder, the positions of mutations are generated in a numeric array with duplicate mutation pairs being removed from the final array. Next, RandomCheck simulates the effect of canonical somatic hypermutation by changing mutations to any of the three other non-germline nucleotides with the probability of any given change being determined by the base pair substitution profiles reported in Maul et al. (2016) and Longo et al. (2009), for murine and human sequences, respectively. This is followed by the generation of motifs, and matching to references, as done by PolyMotifFinder. The motif matched array is then compared to the mutation position array by row for cells that both contain "1" indicating a mutation that also matched a reference sequence. The GC coverage is then determined for this sequence. This process is then iterated 100 to 1000 times to generate a background population for each sequence based on the activity of canonical somatic hypermutation. The results from PolyMotifFinder for that respective sequence is then compared to its population to generate a Z-score. Application of Stouffer's method to the set of Z-scores for each sequence set generates Stouffer's Z value.

Figure S4

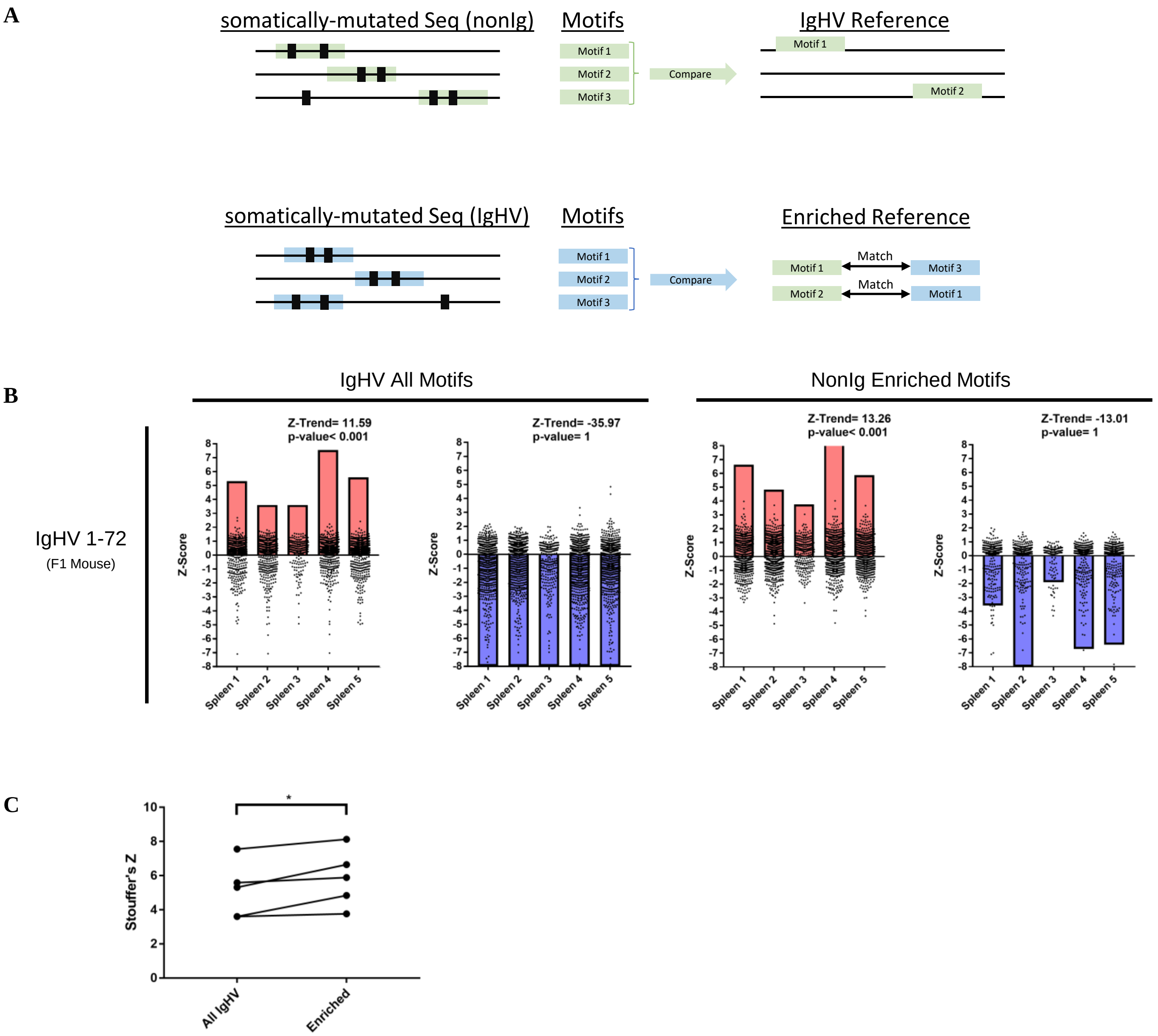


Fig. S4: IgHV genes from F1 germinal centers are enriched for motifs occurring in the IgHV repertoire and the subset of the repertoire that somatically-mutated GPT and β-globin match to. (**A**) Schematic depicting the strategy for enriching motifs used in (**B**). somatically-mutated GPT and β-globin were matched to the murine IgHV repertoire via PolyMotifFinder. Motifs that matched the repertoire were then used as a reference for somatically-mutated IgHV 1-72 sequences isolated from the day 12 germinal center in CB6F1/J mice. (**B**) somatically-mutated IgHV 1-72 sequences from CB6F1/J mice were compared via PolyMotifFinder/RandomCheck to either the IgHV repertoire or the enriched set of motifs defined in (**A**). Data is presented as in Figure 5B-D. (**C**) Scatter plot depicts the effect of enrichment on Stouffer's Z score shown in (**B**). *p<0.05, Paired t-test.