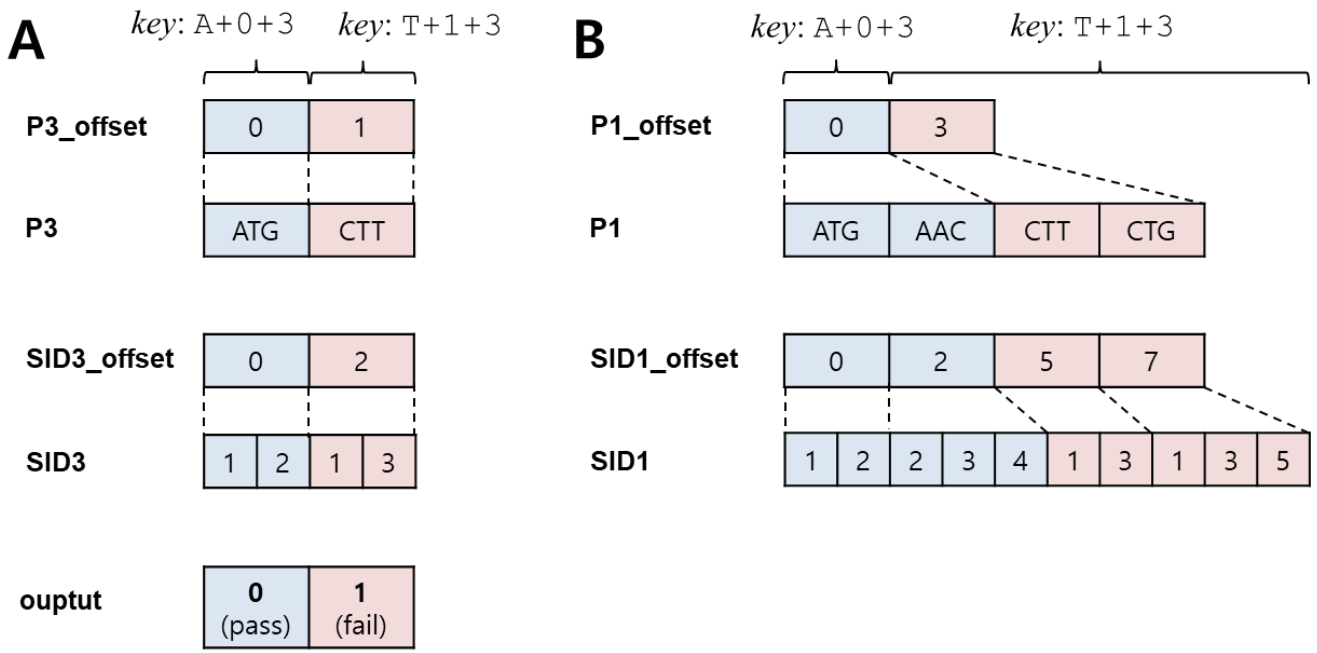


**Figure S1. Arrays for GPU computation in MRPrimerW2.** There are four arrays, P3, P3_offset, SID3, SID3_offset, by grouping the output of Step 3 by using seed+offset+|P| as a grouping key (A). Likewise, there are four arrays, P1, P1_offset, SID1, SID1_offset, by grouping the output of Step 1 by using the same grouping key (B). There are two groups of A+0+3 (blue color) and T+1+3 (red color). The first group means the seed "A" occurs at the offset 0 in the candidate primer "ATG" in the output of Step 3, which again occurs in the sequence IDs {1, 2} (A). It also means the seed "A" occurs at the offset 0 in the subsequence "AAC" in the output of Step 1, which again occurs in the sequence IDs {2, 3, 4} (B). Since "ATG" in P3 is quite different from "AAC" in P1 (i.e., two mismatches), "ATG" passes homology test, i.e., the corresponding element value in the output array becomes 0. However, in the second group, "CTT" in P3 is similar with "CTG" in P1 (i.e., single mismatch), and at the same time, the sequence IDs {1,3} where "CTT" occurs are different from the sequence IDs {1,3,5} where "CTG" occurs. That means "CTT" in P3 may amplify the off-target sequence {5}. Thus, the corresponding element value of "CTT" in the output array becomes 1 (fail).

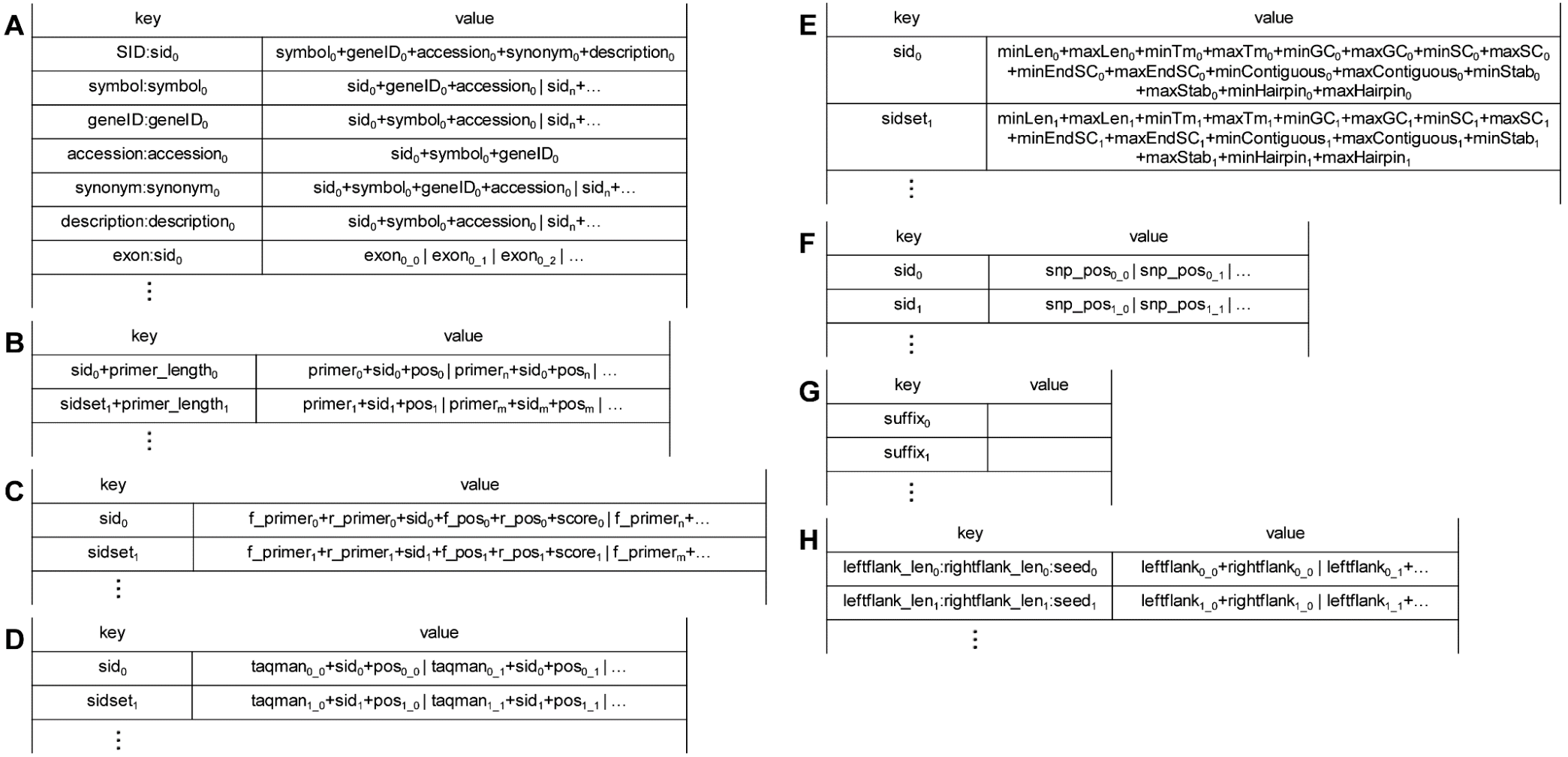
A
key
value
SID:sid0
symbol0+geneID0+accession0+synonym0+description0
symbol:symbol0
sid0+geneID0+accession0 | sidn+…
geneID:geneID0
sid0+symbol0+accession0 | sidn+…
accession:accession0
sid0+symbol0+geneID0
synonym:synonym0
sid0+symbol0+geneID0+accession0 | sidn+…
description:description0
sid0+symbol0+accession0 | sidn+…
exon:sid0
exon0_0 | exon0_1 | exon0_2 | …
B
key
value
sid0+primer_length0
primer0+sid0+pos0 | primern+sid0+posn | …
sidset1+primer_length1
primer1+sid1+pos1 | primerm+sidm+posm | …
C
key
value
sid0
f_primer0+r_primer0+sid0+f_pos0+r_pos0+score0 | f_primern+…
sidset1
f_primer1+r_primer1+sid1+f_pos1+r_pos1+score1 | f_primerm+…
D
key
value
sid0
taqman0_0+sid0+pos0_0 | taqman0_1+sid0+pos0_1 | …
sidset1
taqman1_0+sid1+pos1_0 | taqman1_1+sid1+pos1_1 | …
E
key
value
sid0
minLen0+maxLen0+minTm0+maxTm0+minGC0+maxGC0+minSC0+maxSC0 +minEndSC0+maxEndSC0+minContiguous0+maxContiguous0+minStab0 +maxStab0+minHairpin0+maxHairpin0
sidset1
minLen1+maxLen1+minTm1+maxTm1+minGC1+maxGC1+minSC1+maxSC1 +minEndSC1+maxEndSC1+minContiguous1+maxContiguous1+minStab1 +maxStab1+minHairpin1+maxHairpin1
F
key
value
sid0
snp_pos0_0 | snp_pos0_1 | …
sid1
snp_pos1_0 | snp_pos1_1 | …
G
key
value
suffix0
suffix1
H
key
value
leftflank_len0:rightflank_len0:seed0
leftflank0_0+rightflank0_0 | leftflank0_1+…
leftflank_len1:rightflank_len1:seed1
leftflank1_0+rightflank1_0 | leftflank1_1+…

**Figure S2. Structures of indices used in MRPrimerW2.** There are two kinds of DBs, DB-1 and DB-2. DB-1 is for the search mode in DB sequences, and DB-2 is for the search mode in FASTA sequences, which contains six indices (A-F) and three indices (G, H), respectively. The structure of indices is composed of a key and value pair. In details, indices of the search mode in DB sequences consist of annotation index (A), valid primer index (B), top-1 primer pair index (C), probe index (D), metadata index (E) and SNP locus index (for human only) (F). (A) stores the whole annotation of genes including NCBI gene symbol, NCBI gene ID, GenBank accession number, GenBank alias and keywords, and the exon positions. The key is formatted as "searchtype:query." For example, when the user input is GenBank accession number, NM_001101.3, the format is accession:NM_001101.3. The value of the annotation index is an arbitrary unique integer for sequence (sid) that points to the key in the other indices (B-F). The key portion of indices of (B-E) can have sidset (a set of sid) referring to the multi-targets. (B) The valid primer index contains the primer sequence and position in the sequence of the sid. The key is formatted sid+primer_length. Example key 1+18 contains candidate primers from sid 1 having length of 18. The value is a list of primer data with primer, sid, and pos (position) concatenating + tag. (C) The top-1 primer pair index ontains pre-computed top-1 primer pairs for each target(s). The key is formatted as sid and the value is a list of primer data formatted f_primer (forward primer), r_primer (reverse primer), sid, f_pos (forward position of f_primer), r_pos (reverse position of r_primer) in the sequence sid, and score. (D) The key of the probe index is formatted as sid, and the value is the list of taqman+sid+pos. (E) The metadata index contains minimum and maximum value of the primer length, melting temperature, GC content, self-complementarity, 3′ end self-complementarity, contiguous residue, stability and number of hairpin for sid. The index is used for checking the existence of valid primers for the query within the user-defined constraint. (F) The SNP locus index stores the whole SNP locus of each gene which has the heterozygosity greater than 0. For the search mode in FASTA sequences, (G) is an index for the 5′ cross-hybridization filtering and the key contains the suffixes of all possible subsequences by removing a prefix of length four. The two indices of (H) are for single mismatch and two mismatches for the general cross-hybridization filtering. The key is formatted leftflank_len (length of left flank), rightflank_len (length of right flank), and seed concatenating : tag, and the value is formatted leftflank (left flank) and rightflank (right flank) concatenating + tag. The example of general cross-hybridization filtering is illustrated in Supplementary Figure S2.

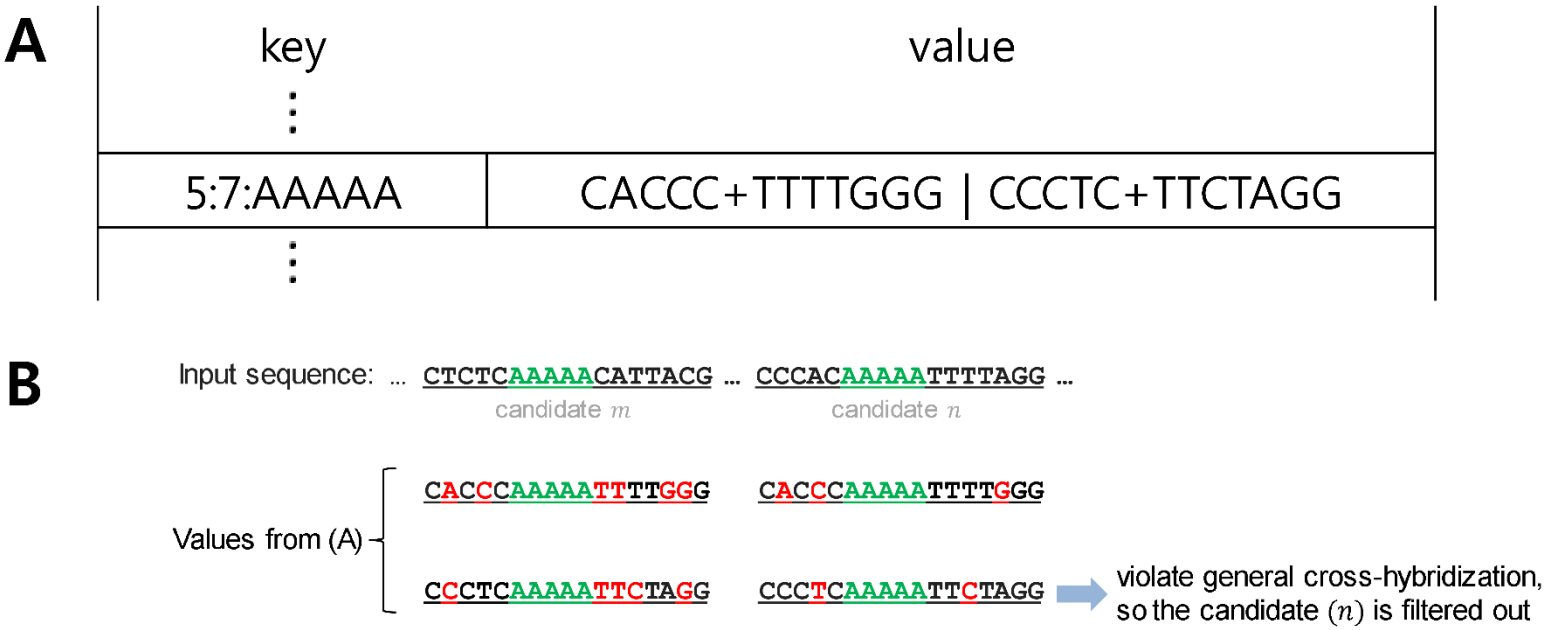


**Figure S3. An example of the general cross-hybridization filtering step in FASTA sequences search mode.** Example index for general cross-hybridization filtering (mismatch = 2) (A) and example of the general cross-hybridization filtering (mismatch = 2) (B). (A) There are a 17-mer candidate primers having common seed 'AAAAA' with 5 and 7 of length of left and right flanks. (B) From the input sequence, MRPrimerW2 generates 17-mer candidate primers and split them into 5bp seed to performing general cross-hybridization filtering. Then, for each candidate $m$ and $n$, MRPrimerW2 retrieves the key of '5:7:AAAAA' and its values in (A) since the candidate m and n have common seed 'AAAAA' in green colored with 5 and 7 of length of left and right flanks. As a results, the values from (A) are not homologue with the candidate $m$ (i.e., both have more than two mismatches), so the candidate m passes general cross-hybridization filtering step. On the contrary, candidate $n$ has two mismatches compared with the one of the values, so it should be filtered out.