

APPENDIX B: CROSS-METHOD SPECIFICITY BENCHMARK

This appendix accompanies Appendix A (scoring and substitution matrices). All numbers are computed by `bench/benchmark.py`; no raw sample data is reproduced.

B.1 Datasets and conditions

Table 1: Benchmark conditions, references and sample sizes.

ID	Loci	Reference	Mode	n (pos/neg)	Purpose
C1	TRB	VDJdb2026 (exact-filtered)	classify	–	Discriminate two co-presented HLA-A*02 epitopes (YFV LLW vs BST-2 LLL) in a sorted sample
C2	TRA, TRB	VDJdb2026 (exact-filtered)	classify	–	Detect CMV NLV-specific TCRs vs tetramer-negative controls in a sorted sample
C3	TRA, TRB, paired	VDJdb2026 (exact-filtered)	classify	–	Rank functionally-validated (padj1e-5) vs non-validated TCRvdb clonotypes, GLC + YLQ
C4	TRA, TRB, paired	VDJdb2025 short-list ($i=2$ refs), self	classify	–	Per-epitope one-vs-rest within the VDJdb2025 reference-replicated shortlist
C5	TRA, TRB, paired	VDJdb2025 short-list (temporal)	classify	–	Per-epitope one-vs-rest on clonotypes NEW in VDJdb2026 (temporal hold-out)
C6	TRA (s5), TRB (s4)	VDJdb2025	FP-only	–	False-positive rate: OLGA-generated TCRs (no true specificity), any hit is a FP
C1	TRA, TRB, paired	VDJdb2026 short-list, self	cluster	–	Clustering quality on the VDJdb2026 shortlist (retention / purity)

B.2 Methods

B.3 Results

Table 2: Compared methods: algorithm and classifier/clustering mode.

Method	Mode	Chains	Algorithm
vdjmatch	classify + cluster	A / B / paired	First-hit control-calibrated E-value: nearest reference within an adaptive edit ball, epitope enrichment as a Poisson tail against an OLGA/repertoire background; neighbour-vote clustering.
terdist3	classify + cluster	A / B / paired	V-gene-aware weighted CDR1/2/2.5/3 Hamming distance (TCRdist); 1-NN / k-NN label transfer for classification, distance-threshold single-linkage for clustering. Run in parallel (parasail).
termatch	classify	B only	BLOSUM62 k-mer similarity kernel between trimmed CDR3 β and a labelled reference set; score 0–1, nearest-reference epitope transfer.
MixTCRpred	classify	paired (A+B)	Per-pMHC transformer over CDR1/2/3 of both chains; one pretrained model per epitope (146 pMHCs; covers NLV/LLW/GLC/YLQ/GIL, not LLL). Paired input required.
NetTCR-2.2	classify	B (+paired)	Convolutional network on CDR3 (and optionally CDR1/2) + peptide; pan-specific binding probability for arbitrary (TCR, peptide) pairs.
GLIPH2	cluster	B	Groups CDR3 β by shared global and local (motif) similarity with V-gene/length/HLA enrichment; outputs specificity clusters (no per-pair score).
imw-DETECT	classify	as provided	ImmuneWatch DETECT (commercial); pre-computed per-sample binding scores supplied as .ods.
ERGO-II	classify	B (+paired)	Autoencoder/LSTM TCR embedding + peptide; supervised TCR–peptide binding probability.