

Computational Evaluation of Retriever v1 Beta Using Public Canine and Human Datasets

Prepared by Shashank Padala · N1 Therapeutics

Research report v0.2 · Revised October 19, 2026

Prepared for scientific review. Not journal peer reviewed.

Research report v0.2, revised October 19, 2026, reporting the computational evaluation of Retriever v1 beta completed October 18. Prepared for scientific review; not submitted to or accepted by a journal. No independent reviewer reproduction is recorded. The revision restructures existing evidence without adding new experiments.

Abstract

Objective: Evaluate supported components of Retriever v1 beta, a computational candidate-selection research workflow, using public canine and human datasets. Sequence construction, tumour RNA support, peptide presentation, binding association and immune-recognition ranking were treated as distinct endpoints.

Methods: Published canine mammary data supported ten source-labelled sequence/gene-RNA cases and one selected matched raw DNA/RNA case. Independent software reconstructed emitted peptide windows. A fixed MHCflurry ensemble ranked measured canine ligands against shuffled peptide backgrounds. Human TESLA panels and IMPROVE supported endpoint-specific binding and ranking comparisons. Component protocol timing, exclusions and training exposure were retained.

Results: All 14,109 emitted windows in the ten-case analysis agreed with independent reconstruction within the supported scope. The selected raw case produced 70 DNA PASS variants and 605 checked windows. Seven variants met the declared mutant-RNA threshold; one produced candidate windows, yielding 38 correlated windows. Observed-ligand recovery in three canine donors was 69.0%, 72.6% and 60.0% against approximately 50% random expectation. In TESLA S7, top-ten lists recovered three positive labels using the fixed expression policy versus one using binding-only. In IMPROVE, corresponding totals were 39 versus 42; configured original Rosie recovered 34. Human binding comparisons included 496 measured-positive rows across 11 HLA alleles.

Conclusion: The study supports scoped sequence-construction correctness, repeatable component analyses and explicit missing-evidence handling. It does not demonstrate general ranking superiority. Missing patient DLA typing, model training exposure and limited biological labels constrain interpretation; canine immunogenicity requires new experiments.

1. Introduction

Retriever v1 beta is N1 Therapeutics' bioinformatics pipeline for identifying and prioritizing candidate cancer-vaccine targets from tumour and matched normal sequencing data. It connects mutations to altered peptide sequences, incorporates tumour RNA evidence and supported DLA typing, and ranks candidate targets for laboratory evaluation and subsequent mRNA vaccine design.

Computational neoantigen selection combines several operations: identify variants, construct altered peptide sequences, evaluate molecular support, predict peptide-MHC compatibility and prioritize candidates for experiments. Correctness at one operation does not establish performance at the next. In canine datasets, DLA coverage and matched experimental measurements can limit evaluation.

Retriever extends the Project Rosie research prototype with explicit input provenance, exclusions and separate evidence fields. The objective of this study was to evaluate its supported components against available references, rather than estimate a single end-to-end clinical accuracy.

The study combined a canine mammary sequencing resource [1], canine DLA ligand observations [2], human TESLA assay panels [3] and human IMPROVE features and recognition labels [4]. Human results evaluate methods under human HLA settings; they do not establish canine immune recognition.

2.1 Study design and analysis status

Pipeline evaluated: Retriever v1 beta. The exact implementation, model and reference fingerprints are recorded in the evidence supplement [6–8]. Phase 1 is the in silico evaluation stage reported here. The proposed Phase 2 study would use this release as the starting pipeline for candidate selection and biological testing, subject to cohort-specific input and model qualification.

This is a computational evaluation of published resources. The ten-case construction and selected raw-case checks used recorded component protocols. TESLA S4 was a development resource. The S7 ranking policy, code and inputs were locked before row-level S7 outcomes and ranking results were inspected; the publication's aggregate sample and positive counts were already known. S7 was not blinded to the source publication or proven independent of all predictor development.

Canine MHCflurry, pVACbind, measured-binding and IMPROVE comparisons were retrospective or exploratory, with fixed settings recorded for their execution. IMPROVE was selected after earlier TESLA results and aggregate IMPROVE outcomes had been inspected. The final processed snapshot does not retrospectively make these analyses prospective.

Resource selection followed endpoint and input availability. The mammary resource provided canine molecular inputs with overlapping modalities; the DLA study provided measured displayed peptides; TESLA provided human recognition labels and some

measured binding. The public TESLA peptide tables did not provide the per-candidate DNA variant frequency needed by the original Rosie scorer. IMPROVE was therefore selected to support that comparator, after earlier outcomes had been inspected, and remains exploratory. These resources do not form one linked cohort or one independent validation population.

Table 1. Evaluation populations and endpoints

Resource	Population used	Endpoint
Canine mammary [1]	10 exact-source-label malignant cases	Emitted-sequence agreement and linked gene RNA.
Selected mammary raw case [1]	CMT-002: matched normal/tumour DNA; matched tumour RNA	Workflow qualification, peptide checks and mutant-fragment support.
Canine DLA ligands [2]	Three fully covered canine donors; one donor excluded	Observed-ligand recovery against artificial shuffles.
TESLA S4 [3]	608 source rows; 404 common-feature ranking rows; 496 measured-positive binding rows	Development ranking and exploratory within-allele binding association.
TESLA S7 [3]	310 rows, three patients, four positive labels	Separate fixed-policy recognition ranking.
IMPROVE [4]	17,520 source rows; 17,412 common-feature rows; 70 patients	Exploratory comparison including original Rosie scorer.

2.2 Data acquisition, curation and cohort selection

Original processed files were retained with checksums and transformed through recorded curation steps. Clinical, specimen, sequencing and expression labels were preserved. Variant and allele representations were normalized within declared reference contexts. Missing values remained missing; assay labels were explicitly parsed. Canonical peptide checks and exact peptide-allele joins preceded model comparisons. Duplicate keys and exclusions were accounted for.

The canine inventory contained 197 cases. Sequential selection retained 140 with a source malignant diagnosis, 132 with matched DNA metadata, 107 with a tumour RNA column and ten with exact paired published-VCF labels. This source-label linkage does not establish genetic identity. Legacy 16K/17K aliases were not inferred from label similarity.

For the selected raw case, normal DNA SRR7780922 and tumour DNA SRR7780923 were acquired with publisher byte counts and MD5 verification, SHA256 acquisition records and complete paired-read checks. Tumour RNA SRR7779580 shared BioSample SAMN09948202 with the tumour DNA. A similarly labelled RNA submission was not pooled because its identity was unresolved.

Canine presentation observations were grouped by biological donor and host context, with duplicate captures retained as provenance rather than additional donors. Human rows retained source cohort and patient identities. ICDC was a discovery route; no separate

ICDC cohort is claimed in this evaluation.

2.3 DNA processing and peptide construction

The reference was CanFam3.1 with Ensembl 91 annotation. The selected raw DNA workflow used BWA 0.7.17, samtools 1.13 and GATK 4.6.2.0 for alignment, duplicate handling, calling and filtering. Qualification intervals were annotation-derived coding regions plus 100 bases, merged and clipped to reference lengths. These intervals are not the original capture BED; the workflow is not an exact reproduction of the source study's raw analysis.

Peptide construction used normalized alleles, reference coding sequences and the evaluated adapter scope. Reference-allele agreement and sequence eligibility were checked. Unsupported contexts, including mitochondrial translation and partial or noncanonical coding sequences, were excluded explicitly. BCFtools 1.13 csq consequences and reconstruction from reference proteins supplied an independent software check of emitted windows. Shared annotation/reference inputs limit the independence of this check.

Construction acceptance required zero recorded reference mismatches, agreement for all emitted windows and exact repeated candidate/exclusion outputs. This evaluates emitted sequences; it does not establish sensitivity across all mutations, transcripts or neoantigen sources. Raw calling occurred once; fresh downstream construction reused completed calls.

2.4 Gene expression and mutant RNA evidence

Published gene FPKM measurements were linked by exact source labels. A positive gene measurement was reported separately from evidence for the mutated allele. Gene expression did not supply DLA genotype or establish peptide display.

Selected tumour RNA was aligned with HISAT2 2.2.1 using the same genome and annotation-derived splice sites. Paired default unstranded settings were used because strandedness was not established. Alignments were coordinate-sorted, duplicate-marked and indexed with the pinned samtools environment.

All 70 normalized DNA PASS variants were evaluated. Informative fragments required mapped primary, non-QC-failed, nonduplicate alignments, mapping quality at least 20 and base quality at least 20. MAPQ 255, missing NH tags and NH other than one were excluded. Exact reference/alternate context included three flanking reference bases on each side; splice skips across the required window were excluded. Overlapping mates counted once; conflicting informative mates were ambiguous.

The declared descriptive threshold was at least three alternate fragments and ten usable fragments. Reference, alternate, other, ambiguous and coverage outcomes were retained. No local haplotype realignment was performed; nearby variants and indel alignment can reduce exact support. Counts were repeated from the same BAM. The threshold is operational and has no demonstrated clinical sensitivity.

2.5 Canine presentation component

MHCflurry 2.0.6 with the upstream 2.0.0 selected pan-affinity ensemble was used without model fitting [5]. Exact supported DLA identities were required; no HLA surrogate was substituted. Complete reported-genotype coverage was required for donor-level ranking. Three human HCT116 monoallelic DLA contexts were evaluated separately and are not counted as canine donors.

Canonical 8–14mer observations were deduplicated within context. Up to one composition-preserving shuffled peptide per observed sequence was generated with seed 20261016 and at most 100 attempts, excluding study-observed sequences and prior context decoys. Failed generation was counted. For multiallelic donors, the minimum predicted affinity across exact reported alleles was used; this does not establish experimental allele restriction.

At K equal to the number of eligible observed peptides, recovery was the observed count in the top K divided by all eligible observed peptides. Random expected recovery was K divided by the combined observed/background count. Exact peptide and peptide-allele exposure in bundled training data was audited. Shuffles are artificial backgrounds, not measured negative ligands.

2.6 Human recognition ranking and comparators

The primary fixed-policy comparison used binding-only and a within-patient percentile policy with 80% binding and 20% expression weight. Missing or invalid required features generated named exclusions; zero expression was valid. TESLA used the recorded source affinity and tumour-abundance columns. Measured binding and recognition outcomes did not enter ranking. Deterministic ties and patient-level ranking budgets were retained.

Top-ten positive-label counts, precision at ten, recall and actual selection counts were recorded. The S7 primary directional criterion was a positive mean paired patient-level precision-at-ten difference. A positive direction alone does not establish statistical superiority. TESLA S4 and S7 were reported separately. A configured pVACbind 7.1.5/MHCflurry 2.0.6 peptide-input workflow supplied an exploratory exact-peptide comparator; this was not a full genomic pVACseq reproduction.

For IMPROVE, canonical 8–11mer source peptides and explicit HLA identities were evaluated using newly executed MHCflurry affinities. The unchanged original Rosie scorer used those affinities, source DNA variant allele frequency and source NetMHCpan binding percentile through its documented percentile fallback. Original 500 nM and 0.01 DNA-VAF filters, 50/30/20 weights and top-20 output were preserved. Top-ten outcomes were derived from the native output.

The IMPROVE common-feature population required finite expression, DNA VAF and source percentile. This aligns input availability but does not eliminate native policy eligibility differences. The original scorer's historical predictor ensemble was not reproduced. Assay selection, shared source cohorts and training overlap prevent pooling the resources into one general accuracy estimate.

2.7 Measured binding, repeatability and statistical scope

Source NetMHCpan and executed MHCflurry predictions were compared with positive measured TESLA S4 competition-IC50 values on identical peptide/HLA rows. Per-allele Spearman correlations used average ranks for ties; different alleles were not pooled. Missing/nonfinite measurements and seven zero values with unresolved censoring semantics were excluded. This is rank association, not numerical affinity calibration or a stability assay.

Candidate/exclusion outputs and downstream metrics were repeated under their recorded rules. Model repeats required identical ranking/recovery counts and score agreement within the declared absolute 0.01 nM / relative 1e-6 tolerances. A separate-environment replay reconstructed IMPROVE curation, scoring and arithmetic using cached source files and recorded predictions. It did not repeat inference or raw alignment.

This report presents descriptive results and the recorded S7 directional endpoint. No new inferential tests or confidence intervals were computed for this editorial revision. Correlated peptide windows, repeated source observations and shared training exposure must not be treated as independent biological replicates. No independent person has reproduced the work.

3.1 Cohort curation and sequence construction

The ten selected mammary cases produced 14,109 peptide windows with zero recorded reference-allele mismatches. All emitted windows agreed with independent BCFtools-based reconstruction within the declared scope. Repeated construction and linked gene-RNA outputs matched exactly.

Positive gene RNA measurements were present for 12,477 windows; 1,632 had zero measurements. These are gene-level measurements, not evidence that each mutated allele is transcribed. Patient-specific DLA typing was unavailable, so no patient-specific nominations were issued.

Table 2. Sequence and expression outcomes

Population	Output	Interpretation
Ten source-labelled canine cases	14,109 checked windows; zero recorded reference mismatches	Agreement for emitted supported sequences; excluded contexts are outside this count.
Ten-case gene RNA linkage	12,477 positive / 1,632 zero gene measurements	Gene expression only; windows are correlated within mutations/transcripts.
Selected raw case	70 DNA PASS variants; 605 checked windows	Workflow execution and scoped construction; no biological mutation truth set.
Raw-case gene RNA linkage	567 windows with positive gene measurement; 38 with zero; none missing	Distinct from mutant-fragment support.

3.2 Raw DNA case qualification

All four selected DNA FASTQs passed publisher digests and complete paired-read integrity checks. A damaged transfer was replaced before alignment. Alignment, duplicate marking, BAM integrity checks, calling and filtering completed. Seventy normalized alternate-bearing tumour records passed the caller filters.

The supported adapter emitted 605 windows with zero reference mismatches; all 605 were reconstructed from independent BCFtools consequences. Three further protein changes were outside declared scope: two lacked an ATG-starting coding sequence and one was mitochondrial. Downstream repetition reproduced candidate and exclusion outputs exactly.

An exploratory comparison found 13 shared calls with legacy alias 16K002-tumor, which contained 18 calls within comparison scope. Because alias identity and the original capture/callable regions are unresolved, these counts are not estimates of mutation precision or sensitivity. Optical-duplicate assessment is limited by archived read names. Host-clock changes prevented a reliable whole-run runtime or cost estimate.

3.3 Mutant RNA support

The tumour RNA input contained 37,348,256 read pairs. Alignment accounting included all 74,696,512 primary read records, with 97.16% primary mapping. Sorting was interrupted and recovered from completed mate preparation; the fixed counting rules were preserved. The alignment itself was not repeated.

Seven of 70 DNA variants met the declared mutant-RNA threshold. Only one of those seven generated candidates within the supported sequence scope, accounting for 38 windows. Repeated counting from the same BAM produced identical observations and candidate evidence.

Table 3. Mutant-RNA evidence in the selected case

Evidence category	DNA variants (n = 70)	Candidate windows (n = 605)
Meets ≥ 3 alternate / ≥ 10 usable fragments	7	38
Insufficient usable coverage	57	377
Adequate depth but below alternate threshold	6	190

Only one RNA-supported variant contributes the 38 windows. These windows are not independent neoantigens. Low coverage or below-threshold support does not establish absence.

3.4 Canine ligand-ranking results

Three canine donors had complete reported-genotype model coverage. Lily was excluded because DLA-88*501:02 was unsupported. Recovery exceeded the artificial-background random expectation in each included donor, but exposure of observed peptides in model

training prevents interpretation as untouched validation.

Table 4. Canine donor ligand recovery and training exposure

Donor	Observed peptides	Recovered in top K	Recovery	Exact training peptides / pairs
Lola	377	260	68.97%	191 / 102
163828A	1,335	969	72.58%	432 / 0
Bogey	7,520	4,513	60.01%	1,733 / 0

K equals observed count. Background counts are 377, 1,334 and 7,517 respectively; random recovery is approximately 50%. Training peptide matches are across any allele; zero exact pairs does not establish independence.

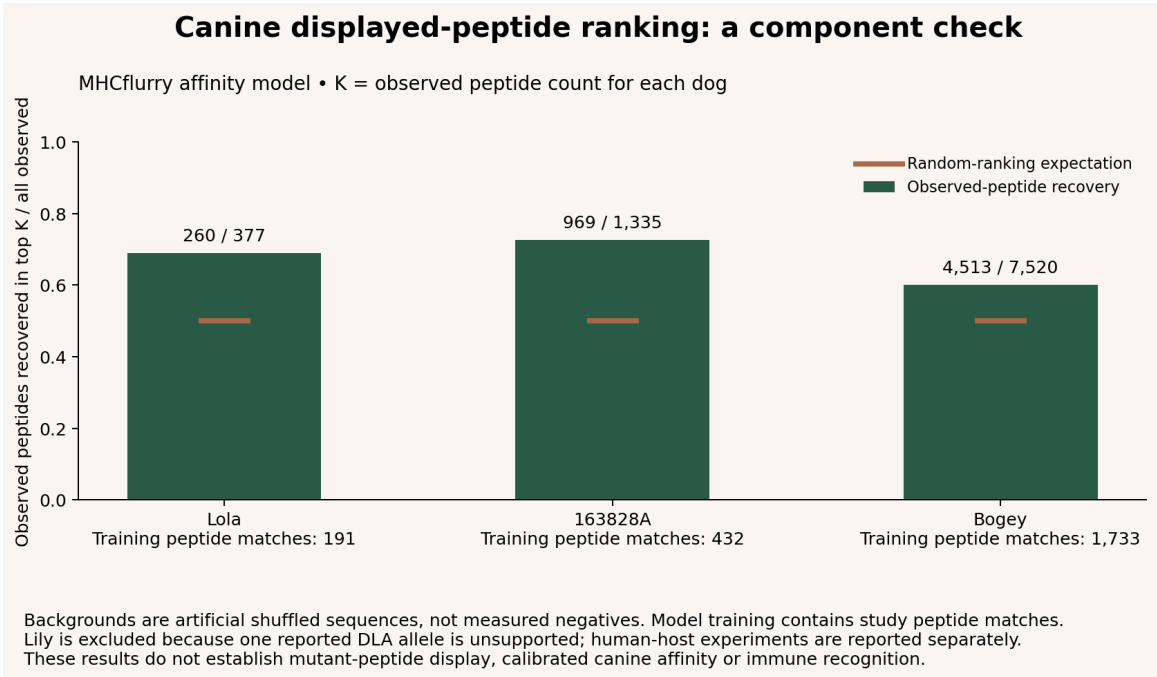


Figure 1. Observed-ligand recovery in three canine donors against composition-preserving shuffles. The chart reports training exposure and excluded donor coverage. This endpoint does not measure mutant-target recognition or calibrated canine affinity.

3.5 Human recognition-ranking results

The separate S7 panel contained three patients and four positive labels. The fixed expression policy recovered three positive labels in patient top-ten lists versus one with binding-only; configured pVACbind also recovered three. The positive paired precision direction met the recorded S7 criterion, but the small number of patients and positives limits inference.

The larger IMPROVE common-feature comparison contained 17,412 rows from 70 patients. Binding-only recovered 42 positive labels in top-ten lists, the expression policy 39 and configured Rosie 34. Each policy made 700 selections. Added expression improved seven patients, tied 52 and worsened 11. Removing any single patient retained the negative aggregate difference; this was a post hoc descriptive sensitivity check.

Table 5. Human panel top-ten positive-label totals

Panel	Common rows / patients	Binding-only	80/20 expression policy	Comparator
TESLA S4 development	404 / 6	8	12	pVACbind: 9
TESLA S7 separate fixed-policy evaluation	310 / 3	1	3	pVACbind: 3
IMPROVE exploratory	17,412 / 70	42	39	Configured original Rosie: 34

Counts are recovered positive labels, not a common cross-study accuracy measure. Native comparator eligibility differs. TESLA and IMPROVE are not pooled.

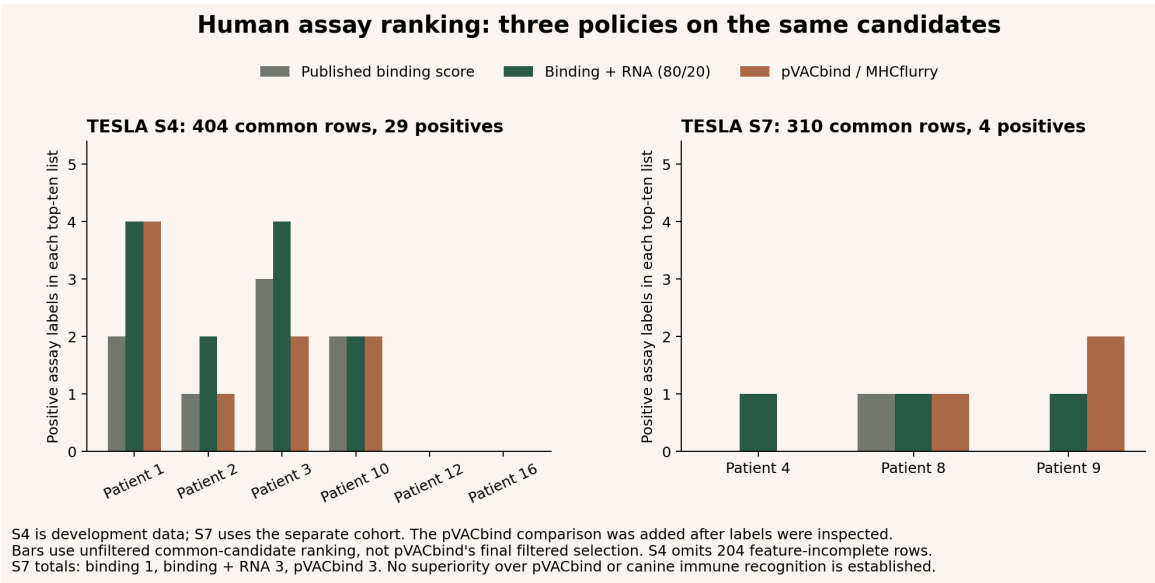


Figure 2. Separate TESLA development and S7 panel results. S4 includes 404 common-feature rows; S7 includes 310 rows and four positive labels. Comparator configuration and native eligibility remain distinct.

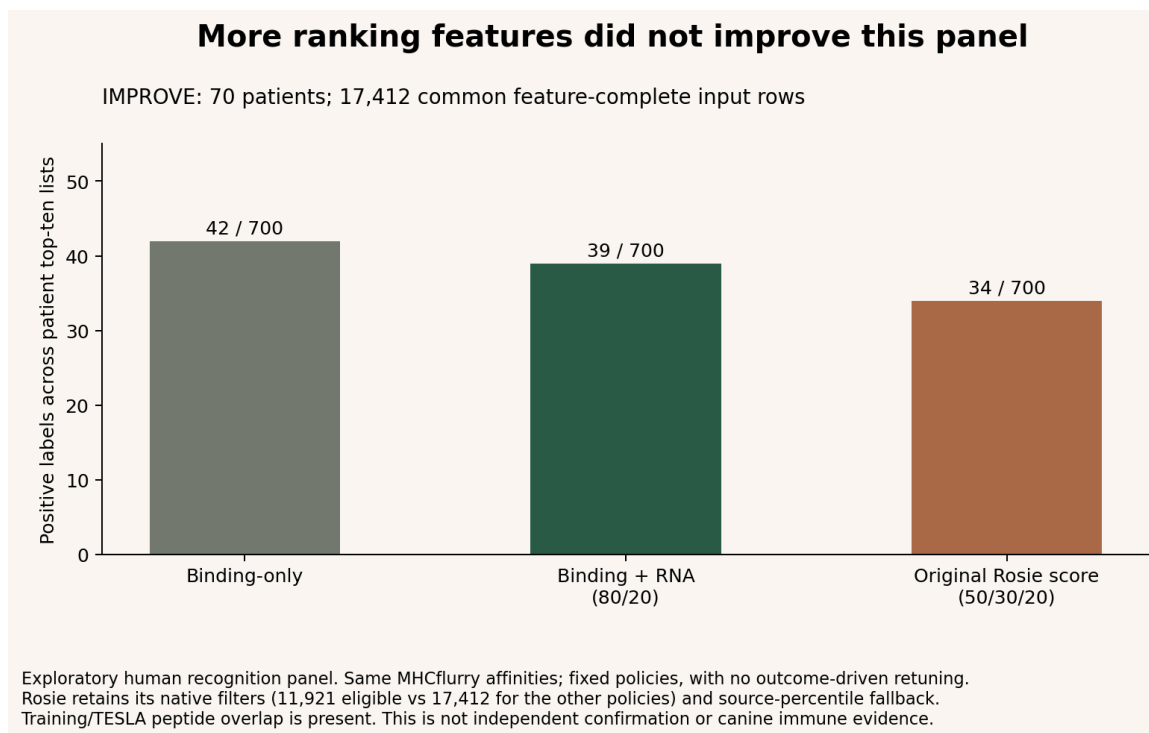


Figure 3. IMPROVE exploratory comparison on the common-feature population. Binding-only recovered more top-ten positive labels than the fixed expression policy or configured original Rosie. This comparison does not reproduce the historical Rosie predictor ensemble.

3.6 Human measured-binding association

Of 608 source rows, 496 were eligible for measured-positive binding comparison, representing 493 distinct peptide-HLA pairs across 11 alleles. Exclusions comprised 105 missing/nonfinite measurements and seven unresolved zero values. Correlations varied by allele; source NetMHCpan had the higher association in eight alleles and executed MHCflurry in three. This descriptive comparison does not establish statistical superiority.

Table 6. Within-allele Spearman correlation with measured positive IC50

HLA allele	Rows	Source NetMHCpan ρ	Executed MHCflurry ρ
A*01:01	27	0.691	0.636
A*02:01	197	0.684	0.656
A*03:01	49	0.657	0.616
A*11:01	35	0.772	0.691
A*24:02	29	0.483	0.583
A*68:01	30	0.487	0.466
B*07:02	18	0.387	0.482
B*08:01	37	0.547	0.549
B*27:05	22	0.782	0.515

HLA allele	Rows	Source NetMHCpan p	Executed MHCflurry p
B*44:02	45	0.698	0.613
B*57:01	7	0.679	0.607

Calculated in the recorded Retriever v1 beta evaluation; values rounded here to three decimals. Rows include repeated peptide-allele observations. These associations do not assess affinity calibration, stability or canine binding.

3.7 Repeatability outcomes

The ten-case construction, exclusions and linked RNA outputs repeated exactly. Selected raw-case downstream construction and mutant-RNA counting also repeated exactly using the existing VCF and BAM. Model and policy repeats met their recorded tolerances and selection checks.

A private-source replay in a separate Python environment reconstructed IMPROVE curation, scoring and arithmetic using recorded predictions. These checks support repeatability at the tested stages. They do not establish a second full raw DNA/RNA run, new model inference in the replay, or independent reproduction by another researcher.

4. Discussion

Sequence-construction agreement and repeatable evidence linkage provide a qualified foundation for further work. They address whether supported inputs produce consistent sequences, rather than whether every true mutation is detected or every proposed peptide is displayed. The selected raw case demonstrates execution with real inputs, while unresolved aliases prevent mutation-accuracy estimation against the published call set.

Mutant RNA provides information beyond gene expression. In the selected case, seven DNA variants met the fragment threshold, but only one generated supported candidate windows. The large low-coverage group limits exclusion decisions. The result motivates matched molecular measurements rather than an assumption that missing support is a true negative.

The ranking evidence does not justify a general advantage from adding expression or original Rosie score components. The small S7 result and larger negative IMPROVE comparison differ in populations and configurations. Their joint implication is that feature complexity requires endpoint-specific evaluation; superiority cannot be inferred from biological plausibility alone.

Canine ligand-ranking results are limited by artificial backgrounds and model training exposure. They do not close the gap between untyped mammary mutations and patient-specific presentation. A subsequent canine study should connect qualified DLA genotype, tumour RNA, peptide-DLA binding/stability and feasible functional assays. Peptide-induced response and recognition or killing of tumour cells require separate controls and endpoints.

5. Limitations

- Canine molecular evaluation covers ten source-labelled cases and one selected raw case, with supported sequence classes only. Shared reference/annotation inputs limit independent software agreement.
- Legacy alias identity, original capture intervals and a biological mutation truth set are unresolved. Mutation precision and sensitivity were not established.
- Patient-specific DLA typing and allele-specific DLA expression are absent in the mammary evaluation. No patient-specific target nomination or measured mutant-peptide display is reported.
- RNA support uses conservative exact context and an exploratory operational threshold. Nearby variants, alignment bias, missing coverage and specimen differences limit interpretation.
- Canine presentation uses artificial shuffles and includes training exposure. Donor-level multiallelic results do not validate each allele separately.
- Human panels contain selected assay candidates, sparse positives and resource/training overlap. Human recognition does not establish canine recognition; binding association does not establish calibrated affinity or stability.
- Exploratory comparator analyses and native eligibility differences limit causal interpretation of score components. No general superiority, vaccine benefit or treatment response was demonstrated.
- Repetition was performed by the implementer. Full raw execution was not repeated; separate-environment replay used cached predictions. Public aggregate outputs do not enable a complete rerun of private software.

6. Conclusion

Retriever v1 beta produced checked peptide sequences and repeatable component outputs within the declared scope. Human ranking comparisons were mixed, and the larger evaluated comparison did not favour the fixed expression policy or configured original Rosie over binding-only. Missing patient DLA and linked canine experimental labels are specific requirements for the next study. The present evidence supports continued research and scientific scrutiny, not biological or clinical validation.

7. Data, software and supplementary methods availability

The original studies [1–4] remain the acquisition sources and govern reuse of their data. The unchanged v0.1 public supplement [6–8] contains component protocols, aggregate results, figure values, cohort flow and provenance hashes. It excludes original study tables, native/model binaries, candidate sequence exports and proprietary application/UI source.

Methods supplement: source-labelled sequence qualification; raw-case qualification; matched-RNA protocol and counting specification; DLA MHCflurry evaluation; TESLA S7; pVACbind; IMPROVE original-Rosie comparison; and human binding association. The supplement's provenance index maps result projections to recorded source fingerprints.

Private implementation access for independent scientific reproduction requires a scoped arrangement. Such access is not yet established. The present public files support inspection of methods and aggregate evidence, but are insufficient for a complete independent rerun. Any journal submission requires a separate assessment of its data and code availability requirements.

8. Preparation, declarations and review status

Prepared by Shashank Padala for N1 Therapeutics with agent-assisted engineering, analysis and drafting. Scientific authorship, contribution statements and approval of a submission manuscript remain to be agreed with actual contributors. No coauthor, reviewer, institution or partner endorsement is implied.

Competing interests: Shashank Padala is the founder of N1 Therapeutics, which is developing Retriever. Funding and any additional competing interests have not been supplied for this manuscript and require author confirmation. This report analyzes public study resources and reports no newly collected animal samples or treatment experiment; a submission-specific ethics and data-use statement remains to be finalized.

Research report v0.2 is a public manuscript-style report prepared for scientific review. It has not been submitted to or accepted by a journal. No independent reviewer reproduction is recorded. The scientific review register [9] accepts version-specific comments and corrections.

Revision history: v0.1 was published October 18, 2026 as an explanatory evidence article. Version 0.2, October 19, restructures the same evaluated evidence into methods, results, discussion, limitations and declarations, adds tabular presentation, and preserves the original v0.1 supplement. No additional experiments or analyses were performed for this revision.

References and supporting materials

- [1] [Canine mammary tumour sequencing resource \(2019\)](#)
- [2] [Canine DLA immunopeptidomics study \(2026\)](#)
- [3] [TESLA consortium study \(2020\)](#)
- [4] [IMPROVE study and feature resource \(2024\)](#)
- [5] [MHCflurry 2.0.6 software and model distribution](#)
- [6] [Versioned evidence supplement: methods, results and provenance](#)
- [7] [Supplement manifest and file hashes](#)
- [8] [Download the original evidence archive](#)
- [9] [Scientific review and corrections register](#)
- [Companion Phase 1 programme delivery report](#)