

Bundibugyo virus carries a genus-exclusive four-position VP24 configuration and a predicted cell-type ordering of interferon restriction

Wenhao Huang

penicillinw@foxmail.com

Nanning No.3 High School <https://orcid.org/0009-0005-7064-2038>

Research Article

Keywords: Bundibugyo virus, VP24, importin- α , interferon, cell-type restriction, sequence conservation

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11329161/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

***Bundibugyo virus* carries a genus-exclusive four-position VP24 configuration and a predicted cell-type ordering of interferon restriction**

Wenhao Huang ^{a,*}

^a Nanning No.3 High School, Nanning, Guangxi, China

* Corresponding author.

E-mail address: penicillinw@foxmail.com (W. Huang).

Abstract

Bundibugyo virus (BDBV) has caused outbreaks with lower case-fatality than *Ebola virus* (EBOV), and its VP24 carries the one BDBV-specific functional difference mapped to date. That mapping used an engineered virus, so whether BDBV itself carries the configuration, and what the defect implies beyond one cell line, were unknown. We combined a sequence census of deposited BDBV VP24 records, a cell-type-resolved human expression atlas and a reanalysis of public infection data. Every full-length BDBV VP24 record and every complete 2026 outbreak genome carried the four-position configuration (S83/Q135/H140/A141), and no other orthoebolavirus did; a wider re-read found two unconfirmed single-position departures among 860 callable copies. Across 154 human cell types, baseline interferon (IFN)-system expression defined a cell-type-resolved score, Δ , read as the IFN-stimulated level a cell attains, and a competing importin- α account (the rival). In the one panel carrying a functional phenotype, only that post-IFN level held under every probe-aggregation rule (AUC, area under the receiver-operating-characteristic curve: 0.83 to 0.90, against 0.67 to 0.80 for the seventeen-gene module's baseline and 0.50 to 0.77 for fold induction). Applied to the defect, that score places urothelium, the airway ciliated cell and circulating myeloid cells highest and the retinal and testicular compartments lowest. We state that ordering as an ordering hypothesis: it converts the cost of a fixed viral defect into a cell-type-resolved, falsifiable prediction, and the framework states the test that would refute it.

Keywords: *Bundibugyo virus*; VP24; importin- α ; interferon; cell-type restriction; sequence conservation

Introduction

Ebolavirus disease severity is species-dependent, yet the viral determinants of that difference remain largely unmapped outside *Ebola virus* (EBOV). *Bundibugyo virus* (BDBV), a member of the genus *Orthoebolavirus*, is the clearest current test case. In May 2026 the World Health Organization declared the Ebola disease outbreak caused by BDBV in the Democratic Republic of the Congo and Uganda a public health emergency of international concern [1]. By 23 September 2026 the Democratic Republic of the Congo had reported 7,890 confirmed cases and 3,799 deaths, a crude case-fatality ratio of 48.1 per cent [2]. No vaccine or specific treatment is approved for BDBV, and this is the largest Ebola disease outbreak recorded in the country irrespective of virus species [2]. The phenotype of BDBV relative to EBOV is already documented, yet most genus-wide studies have not included it. It produces a more protracted, less lethal course than EBOV in non-human primates [3]. Its two previous outbreaks, in Uganda in 2007 and in the Democratic Republic of the Congo in 2012, had crude case-fatality ratios of 30 and 50 per cent [2]. Several comparative structure and function studies in the genus have not included BDBV, so the virus is still often described by what it lacks relative to EBOV, and not by what it carries. The one BDBV-specific functional difference that has been mapped concerns VP24, the protein that competes with phosphorylated STAT1 for a non-classical nuclear-localisation-signal site on KPNA transporters [4], thereby blunting interferon (IFN) signalling. BDBV VP24 binds KPNA1, KPNA5 and KPNA6 with roughly ten-fold lower affinity than EBOV or *Reston virus* VP24, and reduced binding tracks with reduced IFN-induced gene

suppression and a shorter VP24 half-life [5]. When BDBV-type residues were introduced into EBOV at four KPNA-interface positions, the resulting virus no longer inhibited IFN- λ 1, IFN- β or *ISG15* [6]. BDBV VP24 expressed on its own, by contrast, inhibited IFN- β and IFN- λ 1 promoter activation as effectively as EBOV VP24 [7], and inhibited IFN- α -induced *MX1* (MxA protein) and *IFITM3* like most filovirus VP24 proteins [8]. These results leave an unusual gap. The phenotype that distinguishes BDBV is described as a property of an engineered mutant, conditional on the infection context, and defined by a residue set that was chosen by the experimenters. However, it remains unknown whether BDBV itself carries that complete configuration. It is equally unknown whether the configuration is a species-specific trait of BDBV or an arrangement that other members of the genus also carry, and in which cells a retained IFN responsiveness matters most. In the present study, we used a sequence census across the genus, a cell-type-resolved human expression atlas and a reanalysis of public infection data to examine how a virus-intrinsic IFN-escape defect is priced across host cell types. We treat the host variable as a multiplier on a defect the virus already carries. A virus that cannot fully shut down IFN signalling should be penalised in proportion to the IFN-stimulated level the cell attains, and penalised little in cells that are already weakly responsive. BDBV antigen has been reported in immune-privileged tissues of infected macaques that survived [3], and viral RNA was detectable in semen beyond the time at which it cleared from blood in a treated human case [9]. We therefore determine whether the four-position configuration is the natural BDBV state and whether it is exclusive to the species. We then build a cell-type-resolved score of baseline IFN tone from a disease-agnostic human atlas and use it as a multiplier on the VP24 defect. Finally, we state the observations that would refute the resulting gradient and report the first consistency check available in public data.

Methods

Sequence analysis and conservation of VP24

Reference protein sequences for BDBV, EBOV, SUDV, TAFV and *Reston virus* were taken from UniProt and aligned with Clustal W 2.1. Residues were compared at positions defined by the experimental structure below. The 2026 outbreak was read from NCBI Nucleotide twice, by independent routes. The first route locates the VP24 coding region by exact 12-mer diagonal voting against the deposited reference genome (NC_014373.1) with a search window of 300 nucleotides on either side. It calls a position only if the extracted coding sequence matches the reference length, contains no internal stop codon and translates at $\geq 95\%$ identity to the reference protein; this route resolves three of the four positions. The second reads the complete genomes deposited from the outbreak directly. The reference VP24 coding sequence (756 nucleotides, 251 amino acids) is taken from NC_014373.1, its first and last 30 nucleotides are matched exactly and uniquely on either strand with a spacing equal to the coding length minus 30. The extracted coding sequence is translated and required to be full length without an internal stop. A codon carrying an ambiguity code is recorded as not determinable and not as a difference. Thirty-five of the thirty-six records satisfy the double-anchor criterion; the remaining one, which carries the most ambiguity, is called from the leading anchor alone and is flagged as such. The calls are unchanged when the whole procedure is repeated against a 2012 reference genome (KC545393.1), no target codon carries an ambiguity code within two codons of it, and one isolate deposited twice accounts for the difference between thirty-six records and thirty-five independent isolates. These are consensus genomes without read-level data, and they enter the same public submission stream that the aggregation instance indexes. The second route therefore re-reads the same deposits along an independent path, without sampling an independent set, and no statement here rests on within-host variation or on read depth. Positions failing any criterion were recorded as unresolved together with the reason; ambiguity, not sequence difference, accounted for every unresolved call. Two systematic censuses support the residue calls. First, all BDBV VP24 protein records in NCBI were retrieved with the query *Bundibugyo virus*[Organism] AND VP24[All Fields] on 26 September 2026. Every record was classified as full length (251 amino acids) or partial, and residues were read at positions 83, 135, 140, 141, 184 and 217 of the alignment to the EBOV reference protein translated from the VP24 coding sequence of NC_002549.1. Full length BDBV VP24

and EBOV VP24 are both 251 residues and align without gaps, so BDBV N135Q, R140H and V141A and EBOV P83 are reported in EBOV numbering throughout. Second, the same alignment was applied to all six orthoebolavirus species recognised by the ICTV, each queried separately in the NCBI protein database. Records were retained only if the sequence header identified a VP24 or membrane-associated protein. Chains returned for PDB entries whose entry-level annotation mentions VP24 but whose sequence is another viral protein, such as the nucleoprotein, were excluded, and residues that could not be called were recorded as unresolved. Every record returned was retrieved, and each species archive was searched exhaustively. The census and the cross-atlas search were closed on 26 September 2026; later dates in this article are re-reads of the same instance and are identified as such. A sequence was scored only if it was at least 200 residues long, was identified as a VP24 or membrane-associated protein, and yielded a call at all four positions; identical sequences were aligned once and reused. Records failing the length or identity criterion, and positions that could not be called, are counted in the denominators of the census (Fig. 1a) and excluded from its numerators. The same four positions were then read in three independent archives - UniProt, BV-BRC and ENA - for all six species, with the same reference (UniProt Q05322 for *Ebola virus* VP24 and B8XCN4 or R4QRB9 for *Bundibugyo virus* VP24, 251 residues). Complete genomes were prioritised when a species archive exceeded the retrieval cap.

The same four positions were also read in Pathoplexus LAPIS (<https://lapis.pathoplexus.org>), a real-time instance that aggregates deposited genomes and exposes both sequence records and per-position coverage. Its aminoAcidMutations endpoint was queried for three species on 26 September 2026 with a limit of 5000 rows: ebola-zaire (data version 1790096708), ebola-sudan (data version 1790096655) and ebola-bdbv (data version 1790096663). Coverage in that endpoint is the number of indexed sequences carrying a callable amino acid at a position, not the number of records in the instance, and the values quoted in the Results are these per-position coverage values. No variant was returned at positions 83, 135, 140 or 141 in any of the three instances at that snapshot. The ebola-bdbv instance was re-read on 30 September 2026 (data version 1790610472).

Interface geometry and pilot simulations

The interface geometry was computed from the experimental VP24-KPNA5 complex (PDB 4U2X), downloaded and verified through the RCSB data API. Chains A, B and C are EBOV VP24 and chains D, E and F are human importin subunit α -6 (KPNA5). Interface residues were defined by inter-chain heavy-atom contacts within 4.5 Å between cognate pairs (A–D, B–E, C–F). A wider 8 Å shell was used only to establish which positions are not in contact, and the VP24 chain was placed on full-length numbering by exhaustive offset maximisation of sequence identity. The pilot molecular dynamics reported in the Discussion used the same VP24 chain with P83S introduced by truncating the side chain at CB and rebuilding with pdbfixer. The system was built with amber14-all and GBn2 implicit solvent at a 1.0 nm cutoff with hydrogen-mass repartitioning, minimised for 500 steps and propagated at 300 K in one-nanosecond segments on the OpenCL platform. Local RMSD was computed over residues 77–90 and the native-contact fraction used a 0.8 nm heavy-atom criterion with a 1.5-fold allowance.

IFN-tone score

Cell-type-resolved expression values in nCPM for 154 human cell types were taken from the single-cell-type reference table distributed with the Human Protein Atlas. Values were transformed as $\log_{10}(x + 1)$ and z-scored across cell types for each gene. Three components were averaged with equal weight: type-I capacity, the mean z-score of *IFNAR1* and *IFNAR2*; type-III capacity, the mean z-score of *IFNL1* and *IL10RB*; and baseline IFN-stimulated priming, the mean z-score of *ISG15* and *MX1*. *IFITM3* was excluded in advance because of extreme lineage-dependent baseline variation.

Δ is the equal-weight mean of the six z-scores that make up those three components, each standardised per gene across the 154 cell types using the population standard deviation (Eq. 1).

$$y_{g,c} = \log_{10}(nCPM_{g,c} + 1)$$

$$z_{g,c} = \frac{y_{g,c} - \mu_g}{\sigma_g}$$

(1)

$$\Delta(c) = \frac{1}{6}(z_{IFNAR1,c} + z_{IFNAR2,c} + z_{IFNLR1,c} + z_{IL10RB,c} + z_{ISG15,c} + z_{MX1,c})$$

The formula and the resulting ranking were frozen before any external infection dataset was inspected.

Cross-atlas reproduction of the ranking

An independent test of the ranking used nineteen single-cell datasets cached locally, each of which annotates every cell with a Cell Ontology term, together with the Cell Ontology hierarchy (18,914 terms). In that hierarchy *is_a* and *part_of* are both treated as ancestry, and a qualified *is_a* line is read as its bare identifier, since the qualifier records inference provenance, not a different parent. Each of the 154 reference cell types was mapped to a Cell Ontology term by normalised exact label match, by ontology synonym, or through a curated table used where the ontology label differs from the reference name by granularity. 97 types mapped, 60 exactly, 10 by synonym and 27 through the curated table, and 39 of the mapped types contributed at least one reference-atlas pair. For each dataset and matched term, pseudobulk expression was formed from raw integer counts and not from the stored normalised matrix, converted to counts per million, and passed through the same three components and the same equal weighting. The reference values correlated against were rebuilt from the same table, not read from the tabulated score; the two agree at a rank correlation of 0.993 and differ by 0.03 in the pooled value. A pair was formed between a reference cell type and a dataset only where that dataset carried the mapped term or one of its descendants, in which case the dataset value is the library-size-weighted pool of those labels. Agreement is the Spearman correlation over 162 pairs, with a permutation *p* value over 4,000 label shuffles and the smallest correlation detectable at 80 per cent power at that *n*. The primary summary standardises each dataset over its full label set. Alternatives that standardise over matched labels only, that reduce to one row per distinct label set, that impose 200- and 500-cell floors, and that drop each dataset in turn are reported beside it. The probe-aggregation grid behind the level-against-induction contrast is described in Supplementary Methods and Supplementary Note 1, and the leave-one-cell-out and leave-one-atlas-out resamplings are reported with the tests they qualify.

External datasets

Methods for the analyses that test the score against external data appear in the Supplementary Methods and Supplementary Notes 1 to 6. They cover the three panels of the proxy test, the matched panels with their probe-aggregation grid, the filovirus functional anchor, the tissue-level extension with its composition analysis, a primary polarised epithelial panel released at the close of this study's census, the additional coverage and residue-call detail, and the splenic reconstruction.

Importin- α axis

The importin- α axis used the same reference table, $\log_{10}(x + 1)$ transformation and z-scoring across cell types, and was defined as the equal-weight mean z-score of *KPNA1*, *KPNA5* and *KPNA6*, the importin- α genes bound by VP24 in the available structures (Eq. 2). Two pre-specified criteria were applied before any quadrant was inspected: the axis had to be separable from Δ , with an absolute Spearman correlation below 0.6, and it had to have an interquartile range of at least 0.5 z units. Cell types were then classified by their position relative to the median of each axis. A random three-gene axis reaches an absolute Spearman correlation of 0.27 with the IFN-tone score, the 69.5th centile of its null.

Two alternative axes were computed as robustness checks: the maximum of the three importin- α z-scores, and the mean z-score of the PY-*STAT1* cargo genes *STAT1*, *STAT2* and *IRF9*.

$$\begin{aligned}
K(c) &= \frac{1}{3}(z_{\text{KPNA1},c} + z_{\text{KPNA5},c} + z_{\text{KPNA6},c}) \\
K_{\text{max}}(c) &= \max(z_{\text{KPNA1},c}, z_{\text{KPNA5},c}, z_{\text{KPNA6},c}) \\
K_{\text{STAT}}(c) &= \frac{1}{3}(z_{\text{STAT1},c} + z_{\text{STAT2},c} + z_{\text{IRF9},c})
\end{aligned} \tag{2}$$

Because substring matching of cell-type names does not recover the enteric, urinary, hepatic, airway and myeloid sets, compartment membership was defined by explicit cell-type name lists recorded with the analysis code.

Folding the two accounts into a single score was rejected, because the result depends on the combination rule; writing d and k for Δ and for the importin- α axis standardised across the 154 cell types, the candidate rules are those of Eq. 3.

$$\begin{aligned}
C_{\text{add}}(c) &= d(c) + k(c), \quad C_{\text{mult}}(c) = d(c) \cdot k(c) \\
C_{\text{min}}(c) &= \min[d(c), k(c)], \quad C_{\text{thr}}(c) = \min[d(c), \tau], \tau = Q3(d) \\
C_{\text{sat}}(c) &= \frac{d(c) - d_{\text{min}}}{K_s + d(c) - d_{\text{min}}}, \quad d_{\text{min}} = \min_c d, K_s = Q3(d - d_{\text{min}})
\end{aligned} \tag{3}$$

Ranking cell types by the equal-weight sum of the two standardised axes retains 10 of the 20 cell types that the IFN axis alone places highest, and ranking by their minimum, a weakest-link rule, retains 6. The product of the two axes is the most divergent form: it retains 5 of the 15 highest-scoring cell types of the IFN axis alone, 7 of the top 20, and moves 72 of the 154 cell types across the median, against 9 of the top 15 for the additive rule and 4 for the weakest-link rule, with the product and additive orderings correlated at only 0.13. A threshold or saturating transform of Δ leaves the IFN-axis ordering itself almost unchanged (ρ 0.99 and 1.00), so what matters is how the two axes are combined and not the shape given to Δ alone; on the eleven-unit functional panel the four forms give AUCs (first-probe rule) of 0.167 for the product, 0.833 for the sum, 0.867 for the threshold and 0.867 for the saturation. The two accounts are therefore reported separately, the IFN axis alone is the prediction of record, and the importin- α account is retained as the rival; both rules were evaluated on the same 154 cell types and the counts are reproducible from the committed table.

Viral-output reanalysis

The four supplementary workbooks of the multi-virus infection series (GEO GSE114905) were downloaded from the GEO series file directory and parsed. Per-virus, per-time-point totals were obtained by summing the read counts of the seven viral genes; per-gene shares were computed within each virus and time point. Multiplicity of infection and the single-infection design were taken from the published Methods of the source study.

Statistical analysis

All tests were two-sided unless stated otherwise; the exact permutation values quoted for the functional panel are one-sided over the enumerated label assignments, and are the smallest attainable over that enumeration. Values that could not be read or called were recorded as not determinable together with the reason, excluded from numerators and retained in denominators. An undeterminable value was thereby distinguished from a negative result, and only the latter counted as evidence against the presence of a feature. Associations are reported as Spearman correlations, discrimination as areas under the curve, and paired comparisons with exact sign-flip or Wilcoxon tests as stated with each analysis. Multiple comparisons were handled by a family-wise correction over the contrast sweep, with the corrected value reported beside the uncorrected one ($p = 0.056$ over the nine-unit panel, 0.063 over the eleven-unit panel; Supplementary Note 1). The compartment compositions were tested by exact multinomial goodness-of-fit against the atlas-wide quadrant composition, with Benjamini–Hochberg correction across the 25 compartments that have at least one member. Label-permutation p values are reported over 4,000 shuffles for the cross-atlas test and over 5,000 draws for the splenic comparison. Power is reported as the smallest correlation detectable at 80 per cent power at the observed n , which is 0.92 for the six matched groups of the protein-layer comparison and about 1.0 at the observed effect size for the cross-atlas test. The

statistics reported in the main text are the areas under the curve for the three candidate readings of the host axis, Spearman correlations with their exact permutation p values, and the leave-one-cell-out and leave-one-atlas-out resamplings. The primary polarised epithelial panel uses exact paired tests. The analysis code, the software and library versions, the entry point of each script, the random seed used for Monte Carlo permutation and the number of permutations are listed with the deposited analysis scripts.

Software and versions

Sequence alignments used Clustal W 2.1. Structural figures used PyMOL 3.1.1. The molecular-dynamics pilot used OpenMM 8.6 with pdbfixer, the amber14-all and implicit/gbn2 force-field files, and the OpenCL platform. The analysis used Python 3.14.6 with NumPy 2.5.2, pandas 3.0.5, SciPy 1.18.0, matplotlib 3.11.1, Biopython 1.88 and scikit-learn 1.9.0. The recorded versions of these packages are listed in the deposited environment file.

Results

Bundibugyo virus carries a genus-exclusive four-position VP24 configuration in every sequenced outbreak

The engineered mutant that lost IFN- λ 1, IFN- β and *ISG15* inhibition carried BDBV-type residues at four KPNA-interface positions [6]. EBOV P83, N135, R140 and V141 correspond to BDBV S83, Q135, H140 and A141, so the mutant reproduces the natural BDBV interface. We therefore enumerated every BDBV VP24 protein record in NCBI. The query returned 32 records: 31 are full length and every one carries S83, Q135, H140 and A141, while the remaining record is a 95-residue fragment that does not span the region (Fig. 1a). The records span both earlier outbreaks, differ from EBOV at 35 or 36 of 251 positions, and collapse to two sequences that differ only at residue 24. The 2026 outbreak supplies a third time point at the nucleotide level: across the 36 complete genomes deposited from it, which represent 35 independent isolates from four submitting laboratories, every one carries S83, Q135, H140, A141 and R184. Three of those genomes carry ambiguity inside VP24 away from these positions, none within two codons of them, and none differs at a position that could be called. Reading the archive directly recovers position 83 from the nucleotides as well, where it is serine in all 31 full-length protein records. Across the two archives, retrieved at different times and along different extraction routes, the configuration is invariant in every BDBV outbreak for which sequence exists, an interval of nineteen years. A re-read of the same surveillance instance on 30 September 2026, after this census, returned 886 BDBV VP24 copies. Of these, 860 were callable at all four positions: 858 carry the complete configuration and two, both from the 2026 outbreak, each show a single-position departure (S83L, A141V), with no record departing at more than one. The remaining 26 could not be called at one or more of the four positions and are reported as unresolved. Neither departing copy is one of the 36 complete genomes read above; both are consensus-only records from the 2026 outbreak.

The binding study that first compared the species writes out the EBOV and BDBV residue at each position [5], and the mutagenesis study gives its virus BDBV residues at all four [6]. A concurrent preprint reports that VP24 varies minimally across the 2007, 2012 and 2026 outbreaks without naming the positions [10]. We are not aware of a study that states the four-position configuration as a single, species-level feature of BDBV.

To test whether the configuration is exclusive to BDBV, we aligned the VP24 protein of all six species recognised in the genus against the EBOV reference and read the four positions (Fig. 1a). Across 3,490 assessable VP24 sequences, the complete configuration S83/Q135/H140/A141 occurs only in BDBV, and in every BDBV sequence in which the region could be read. Two of the four positions are exclusive to BDBV on their own. Position 83 is serine only in BDBV, the other species carrying proline or, in *Sudan virus*, glutamine. Position 140 is histidine only in BDBV, the other five species carrying arginine, glutamine or serine there. The remaining two positions are not exclusive: only *Tai Forest virus* carries Q135 and A141, as P83/Q135/Q140/A141, and no other species carries either. *Sudan virus* (Q83/N135/R140/V141) shares none of the four with BDBV, and no species

shares three. The same partition holds outside the nucleotide archive. In UniProt, BV-BRC and ENA, every BDBV record in which the four positions can be read carries all four; the per-source denominators, including the one BV-BRC record that could not be read, are given in the Fig. S1b caption. No non-BDBV entry screened carries the combination. The aggregation instance (26 September 2026) records no amino-acid variant at the four positions in any non-BDBV VP24 copy it indexes (VP24 coverage 11,561 for Ebola virus, 579 for Sudan virus and 849 for BDBV). The configuration is therefore not a strain-level variant, and the engineered mutant reconstructs an interface that exists nowhere else in the genus (Fig. S1b).

All six species were retrieved exhaustively. Coverage of the four positions is 31 of 31 BDBV records, 3,253 of 3,271 EBOV records, 146 of 147 *Sudan virus* records and every *Tai Forest*, *Reston* and *Bombali* record (Fig. 1a; Table 1). The shortfall is fragments and records in which at least one position could not be called; those were left unscored as unresolved, so an unreadable position is never a negative one. All four positions were invariant within EBOV itself, so the engineered mutant introduced the residues of a different species, not rare EBOV variants. Read as an upper bound on what the smaller archives can exclude, the one-sided 95% exact confidence limit on the frequency of the complete configuration in a species in which none was seen is 0.09% for EBOV, 2.0% for Sudan virus, 10.5% for Reston virus, 25.9% for Bombali virus and 52.7% for Tai Forest virus, each computed on that species' own callable count (3,253, 146, 27, 10 and 4). A two-position description of residues 135 and 140 is insufficient, because the definition of the feature includes the infection context: the same protein inhibited IFN induction when expressed alone but not in an infected cell carrying the four positions. The species-level feature is therefore the complete four-position configuration, in every sequenced BDBV outbreak and in no other species assessed, not the two positions the binding comparison originally highlighted.

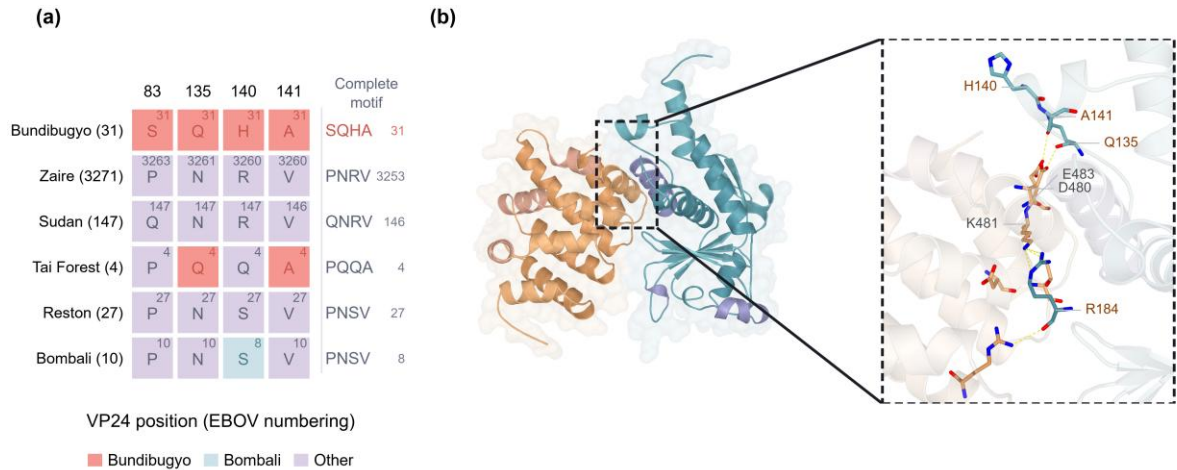


Fig. 1. Bundibugyo virus alone carries the complete four-position VP24 configuration. (a) Residue calls at EBOV VP24 positions 83, 135, 140 and 141 across the six orthoebolaviruses; each cell gives the major call and its record count, the row label gives sequences assessed, and the right column gives the assembled motif. Of 3,522 retrieved VP24 records, 32 fragments outside the four-position window were excluded from the denominator rather than scored as negative, leaving 3,490 sequences assessed; 19 left at least one position uncalled, so 3,471 carried a call at all four positions, 3,440 of them from species other than Bundibugyo. (b) Bundibugyo VP24 homology model on the KPNA5 chain of PDB 4U2X, showing the four KPNA5-contacting variant positions (N135Q, R140H, V141A and Q184R) and the residues they touch; dashed yellow lines mark the six inter-chain heavy-atom N/O...N/O contacts below 3.6 Å between the four variant positions and KPNA5 (two from Q135 and four from R184), and P83 lies 11.7 Å from the nearest KPNA5 atom in the model and 11.4 Å in the experimental coordinates; no KPNA5 atom lies within 8 Å of it in either structure, and its closest atom in the crystal, 5.3 Å away, belongs to a second VP24 chain rather than to KPNA5. Residue calls were read at the four positions after gap-free alignment to EBOV VP24, and structural distances use heavy atoms.

Receptor capacity and priming combine into a cell-type-resolved baseline IFN-tone score

To convert the defect into a cell-type-resolved cost, we built a host-side multiplier from a single-cell-type RNA reference of 154 human cell types. From that reference we extracted the type-I receptor complex (*IFNARI*,

IFNAR2), the type-III receptor complex (*IFNLR1*, *IL10RB*) and two baseline IFN-stimulated transcripts (*ISG15*, *MX1*), and combined them into one score, Δ , after log transformation and z-scoring across cell types (Fig. 2a, b). *IFITM3* was excluded because its baseline varies by more than an order of magnitude between lineages for reasons unrelated to IFN tone; including it would have made the score a lineage marker. The type-III receptor shows the expected epithelial bias, with Ocular surface epithelium (113.5 nCPM, normalised counts per million), Urothelium (102.5), Enterocytes (101.8) and Epididymal efferent-duct ciliated cells (91.5) at the top. The bias is not absolute: Plasmacytoid dendritic cells (110.4), Plasma cells (46.7) and Hepatocytes (38.6) are also high. Four of these display names differ from the labels in the source table, which reads ocular epithelial cells, urothelial cells, pDCs and epididymal efferent duct ciliated cells. Retinal pigment epithelium (8.5), Sertoli cells (1.7), Leydig cells (3.3), Endothelium (1.0–2.3), Monocytes (1.6), Oligodendrocytes (0.4) and Neurons (6–9) are low on this term alone. Monocytes rank high on Δ through the priming term, not through the type-III receptor. The score therefore behaves as known receptor biology predicts and adds a priming term that is not receptor-driven. The highest- and lowest-ranking cell types are listed in Table 2.

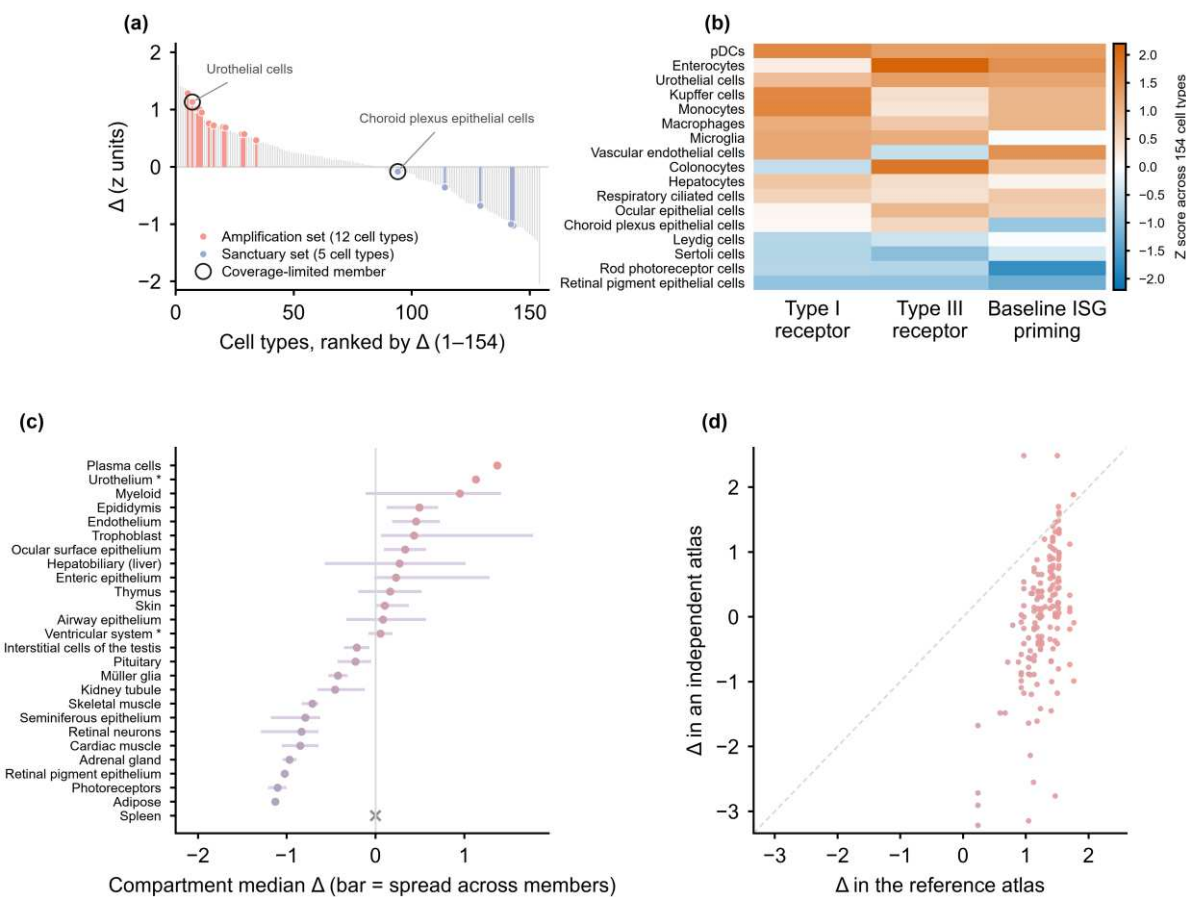


Fig. 2. A cell-type-resolved IFN-tone axis partitions the reference atlas more sharply than organs. (a) The frozen score Δ for all 154 human cell types, ranked, one stem per cell type; salmon marks the 12-member Amplification set and periwinkle the 5-member Sanctuary set, the two sets of cell types named in Results; open rings mark the two coverage-limited members (Urothelial cells and Choroid plexus epithelial cells). (b) The three Δ components for 17 representative cell types: Type I receptor capacity, Type III receptor capacity and Baseline ISG priming. (c) Median Δ per anatomical compartment (member range), ranked by member median; members span a lavender bar, medians are coloured on the frozen Δ scale, the spleen has no member cell type in the reference atlas (grey $\times$, 0 of 3 members; Supplementary Note 5), and an asterisk marks the Urothelium and Ventricular system, coverage-limited in independent atlases (Supplementary Note 4). (d) The ordering in independent atlases; each point is one matched pair, an ontology-matched label shared by an independent atlas and the reference atlas, coloured by its reference-atlas Δ . Δ is the equal-weight mean of three components; each component is the mean z score of its genes across the 154 reference cell types, computed from nCPM after $\log_{10}(x + 1)$ transformation. Panel (d) reports Spearman rho = 0.519,

permutation $p = 0.0002$ and $n = 162$ matched pairs across 18 atlases, with $\rho = 0.452$ when each atlas is standardised over only its matched labels.

A rival importin- α account predicts the opposite ordering for a defined set of cell types

The defect acts through one route: VP24 competes with phosphorylated STAT1 for the nuclear-localisation-signal binding site on importin- α , and the BDBV protein competes poorly. A score built only from IFN receptors and baseline IFN-stimulated transcripts therefore omits the host component of the pathway the defect uses: a cell type with high IFN capacity but little importin- α has few of the sites the competition is about, and should pay little cost if competition is the only route. That reading is the rival to the one-axis account, not a refinement of it, and its axis was pre-specified as the mean z-score of *KPNA1*, *KPNA5* and *KPNA6* in the same reference. Across 29 healthy human tissues profiled for both layers, *KPNA1* mRNA and protein correlate at 0.18 and *KPNA6* at 0.48, while *KPNA5* protein is undetected in 24 of 29. Across the 154 cell types the Spearman correlation between Δ and the importin- α axis is -0.27, inside the pre-specified bound of $|\rho|$ below 0.6, and the axis spans an interquartile range of 0.90 z units. The three genes are not interchangeable (*KPNA5* and *KPNA6* correlate at only +0.27), so the axis is the average of a heterogeneous set, not a property of one importin. Splitting the atlas at the median of each axis gives 30 cell types high in both, 47 high in Δ but low in importin- α , 47 the reverse and 30 low in both; the 47 in each disagreement quadrant discriminate between the two accounts.

The split is a convention, not a property of the data: tertile, quartile, mixture or k-means cuts move the quadrant counts, but the Amplification set's membership at the two extremes is unchanged in every cut of the frozen score, and the six extremes survive the wider grid of gene sets, transformations, standardisations and weightings, where the share of perturbed assignments reaches 11.3% for macrophages and stays below 3% for enterocytes and urothelial cells (Supplementary Note 4). The high- Δ , low-importin quadrant holds the cell types whose predicted direction reverses if STAT1 nuclear import is the only mechanism. Ten of the twelve Enteric epithelial cell types fall there, including Enterocytes (Δ rank 5), Colonocytes (20) and Goblet cells (23), together with Kupffer cells (9), Hofbauer cells (8) and three of the four hepatic cell types. Urothelial cells (7), Plasmacytoid dendritic cells (2), Neutrophils (6), Monocytes (10), Macrophages (11), Microglia (14) and Respiratory ciliated cells (28) score high on both axes and retain the full predicted penalty. This does not overturn the gradient; it defines where the two accounts part company. The enteric epithelium remains the most IFN-responsive epithelium in the atlas, but on the rival account that responsiveness does not translate into a BDBV-specific disadvantage there. The urothelium, whose compartment median ranks second (Fig. 2c), is untouched by this qualification and becomes the cleanest predicted amplification compartment in the reference atlas. The assignment of every compartment under both accounts is given separately (Fig. 3b).

The one functional test available does not support a role for the axis. In the IFN panel whose record states which members become restricted to HIV-1 after IFN and which remain permissive [11], the importin- α axis separates the two groups no better than chance (Fig. 3c), with an AUC of 0.47 to 0.50 depending on how the three importin- α genes are aggregated, against 0.87 for the three-component score in the same units. The difference did not reach significance in an exact paired label-swap comparison over the same eleven units (0.37; two-sided $p = 0.23$, one-sided $p = 0.11$) or over the nine units that exclude both partially resistant lines (0.40; two-sided $p = 0.29$). We therefore retain the importin- α axis as the rival. It reverses the predicted penalty for an enumerable set of cell types, so it is carried into the falsification conditions as the rival to be discriminated, and the prediction of record remains the IFN axis alone.

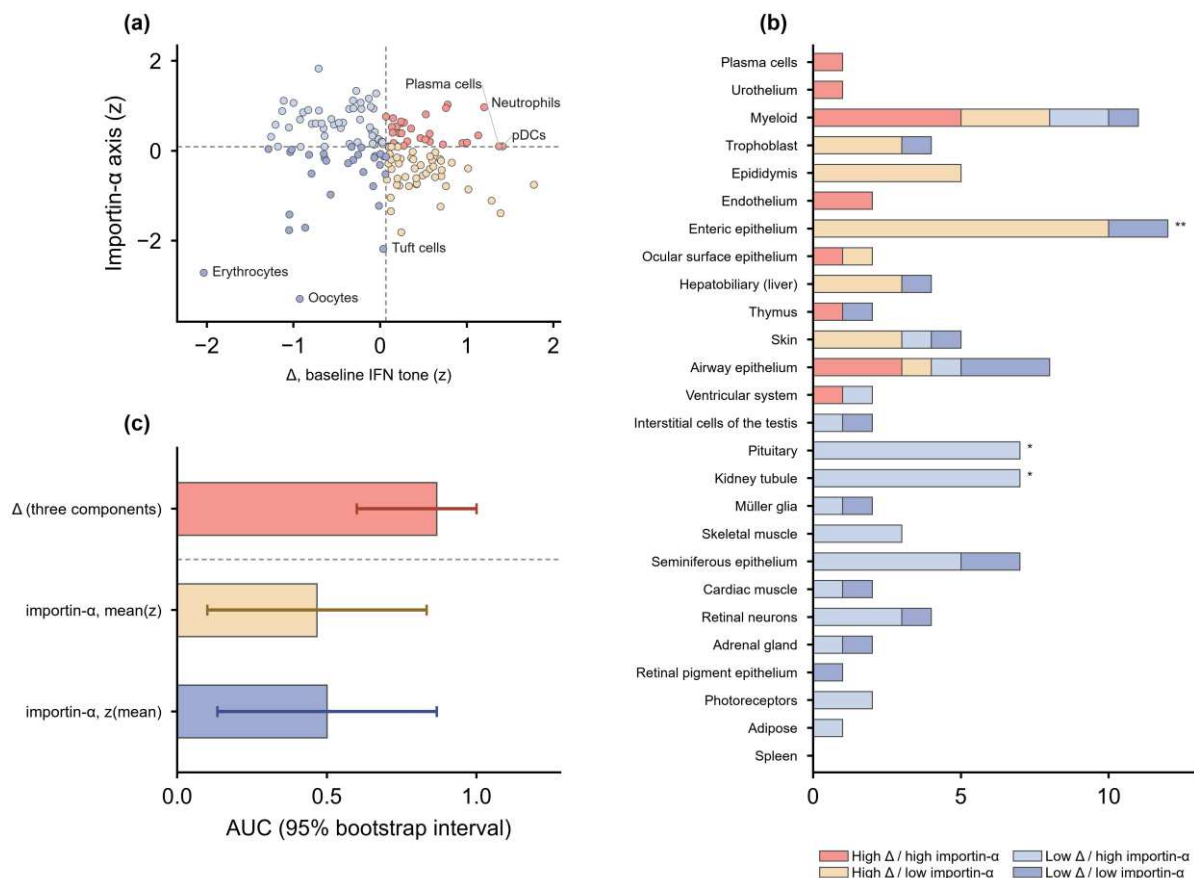


Fig. 3. A rival importin- α axis reclassifies selected compartments. (a) The 154 reference cell types on the Δ and importin- α axes; dashed lines are the two axis medians; high and low follow a two-axis median split (quadrant sizes 30, 47, 47 and 30); points are coloured by quadrant; the three most extreme members of the high- Δ /high-importin- α quadrant and of the low- Δ /low-importin- α quadrant are named. (b) Every compartment of the coverage audit, members stacked by quadrant, ranked by mean Δ ; asterisks mark compartments departing from the atlas-wide quadrant composition (expected composition from the same atlas; two-sided exact multinomial test; Benjamini–Hochberg across 25 compartments with members: * $q < 0.05$, ** $q < 0.01$, *** $q < 0.001$); the spleen has no member cell type in the atlas. (c) Discrimination in the 11-unit functional panel (five restricted, four non-restricted and two partially resistant): Δ , importin- α mean(z) and z(mean). The Spearman correlation between the two axes is -0.27 against a pre-specified $|\rho| < 0.6$, and the importin- α axis spans an interquartile range of 0.90 z units. Panel (c) gives Δ 0.87 (95% bootstrap interval 0.60–1.00), importin- α mean(z) 0.47 (0.10–0.83) and importin- α z(mean) 0.50 (0.13–0.87); the dashed line marks chance (0.5); intervals are 2.5th–97.5th percentiles of 5,000 stratified bootstrap draws (seed 20260930). Panel (a) reports no inferential statistic. AUC, in-sample area under the receiver-operating-characteristic curve for restricted versus non-restricted units; Δ , cell-type-resolved baseline IFN tone; mean(z), mean of the three per-gene z scores; z(mean), z score of the three-gene module mean.

Same-cell panels link baseline tone to the IFN-stimulated level

Two same-cell panels test whether the score can be read as responsiveness; a third, a cancer-cell-line panel [12], is reported with the others in Supplementary Notes 1 and 2 (Fig. S2, S3). The first, in which every member responds, shows that baseline tone explains essentially none of the variation in the response (-0.08 , exact permutation $p = 0.84$) and tracks the induced level instead ($+0.88$, $p = 0.003$, $n = 9$). The second, in primary human immune cells from six healthy donors across five populations [13], reverses the relation for the additional response, and the reversal reproduces in an independent cohort from a different study [14] (Monocytes -1.00 in each; exact $p = 0.003$ and 0.017 ; Fig. S3c, d). The pattern is convergence, not desensitisation. The primary epithelium the prediction concerns is covered by two further instruments analysed in the supplement, a differentiated airway epithelium series [15] and a keratinocyte series. The score is therefore read as a baseline-

tone axis whose operational reading is the IFN-stimulated level a cell attains rather than the size of the induction it mounts. The partition below is therefore stated as an ordering hypothesis.

The ordering survives the variants of the gene set and of the score tested here, but not every weighting. Recomputing Δ under five variants, which drop any single one of its six genes, replace the two-gene priming term with a seventeen-gene IFN-stimulated module, rank instead of z-score, take the first principal component or omit the log transformation, gives Spearman correlations against the frozen score of 0.925 to 0.984, and no variant retains fewer than 11 of the 15 highest-scoring cell types. Reweighting the three components with the type I against type III effect measured in primary nasal epithelium, or with weights fitted on the functional panel, moves the correlation to 0.736 to 0.965 and leaves 6 to 12 of the 15 in place. The fitted weighting loses on leave-one-out cross-validation (out-of-sample rank correlation 0.00 to -0.07), so equal weighting is retained as the defensible choice at this sample size.

The ranking also reproduces outside its reference atlas once cell types are matched by harmonised ontology labels rather than by name. Mapping the 154 reference cell types onto Cell Ontology terms and rebuilding the same three components from raw counts in eighteen of the nineteen independent single-cell datasets screened gives Spearman 0.519 over 162 matched labels (permutation $p = 0.0002$), 0.452 when each dataset is standardised over only its matched labels, and 0.493 to 0.555 across leave-one-atlas-out resamplings, so no single dataset drives it; kidney is the only negative value (Fig. 2d). Baseline tone is therefore a reproducible ordering across atlases and gene sets. In primary, differentiated, air-liquid-interface human nasal epithelium, type I IFN induced the IFN-stimulated module more strongly than type III IFN at equal mass concentration, the only primary-epithelium measurement of both arms in the same cells (Fig. S4a-d; Supplementary Note 3). Across these tests, Δ behaves as a single operational quantity: the IFN-stimulated level a cell attains. The ordering built on it is stable across gene sets, score variants and independent atlases, and it is in that form that the score is carried into the cell-type analysis below.

Multiplying the VP24 defect by baseline IFN tone ranks cell types within an organ

Multiplying the VP24 defect by Δ predicts a cell-type-specific relative disadvantage for BDBV, and that it is smallest in the compartments where filovirus persistence has been documented. Δ enters as a multiplier on a defect that is the same in every cell type, so the ordering below is the ordering of Δ itself. The extremes of the IFN axis alone (Fig. 2a) place Enterocytes, Urothelium, Kupffer cells, Monocytes, Macrophages, Microglia, Vascular endothelial cells, Colonocytes, Epididymal efferent-duct epithelium, Respiratory ciliated cells, Ocular surface epithelium and Hepatocytes at the high end. Retinal pigment epithelium, rod photoreceptors, Sertoli cells and Leydig cells sit at the low end; the choroid-plexus epithelium sits below the median at rank 94. The compartment medians are given separately (Fig. 2c). On the rival account the penalty would be confined to the 30 cell types that score high on both axes. Urothelial cells, Plasmacytoid dendritic cells, Neutrophils, Monocytes, Macrophages, Microglia, Respiratory ciliated cells and Type 1 alveolar cells would retain it. The Enteric epithelium and the hepatocyte lineage, by contrast, are the compartments where that penalty would be cancelled, because their IFN responsiveness is high and their importin- α transcripts are the lowest in the atlas. The ordering is stated per cell type and not per organ, because the compartments are internally heterogeneous, and it is falsifiable at that resolution. The ordering separates cell types within the same organ: surface epithelium is high-scoring while the retinal pigment epithelium and the photoreceptors are the lowest-scoring cell types in the atlas, the efferent-duct epithelium is high-scoring while Sertoli and Leydig cells are low, and microglia score high while astrocytes, neurons and choroid-plexus epithelium do not. Each pair therefore predicts where a virus with this defect should and should not accumulate, within the same organ. The two accounts separate one of these contrasts and part company over a second. Enterocytes and Colonocytes are the most IFN-responsive epithelium in the atlas yet carry the lowest importin- α levels of any epithelial set. That is the configuration in which the two accounts make opposite predictions, and therefore the configuration that can discriminate between them. One compartment is not

uniform: choroid-plexus epithelium scores below the median (rank 94 of 154) while ependymal cells score above it (rank 61). The prediction of record is the IFN-tone axis alone; a rival importin- α account that would reverse the penalty for an enumerable set of cell types is stated separately as the rival to be discriminated (Fig. 3a). The score therefore converts a virus-intrinsic defect into a relative, cell-type-resolved cost. In the one multiplicity-matched dataset available, a hepatocyte-lineage line, BDBV produced less viral mRNA than EBOV, the direction the IFN axis alone predicts (Fig. S2c, d and Supplementary Note 6).

Discussion

BDBV carries the four-position VP24 configuration natively, in every full-length protein record deposited for the species and in every complete 2026 outbreak genome read here. No other orthoebolavirus species carries the combination, and two of the four residues are exclusive to BDBV across the genus. A configuration first characterised as an engineered construct is therefore a natural feature of this species. A host-side score that acts as a multiplier turns that cellular phenotype into a cell-type-resolved prediction of where the same virus should be most and least constrained. A rival importin- α account, weakly correlated with Δ , that would reverse the penalty for a defined set of cell types is carried beside it as the rival. The interface geometry (Fig. 1b and Fig. S1a) is the first of two structural points. In the experimental VP24-KPNA5 complex (PDB 4U2X) four of the five positions at which BDBV differs from EBOV contact KPNA5 (N135Q, R140H, V141A and Q184R), and Q184R carries the second-largest contact footprint (Fig. S1a). A two-position account of the BDBV interface therefore omits a position that contacts KPNA5 extensively. The functionally required position 83 is not an interface residue at all: P83 lies 11.4 Å from the nearest KPNA5 atom in the experimental complex, has no KPNA5 atom within an 8 Å shell (its closest atom belongs to a second VP24 chain, 5.3 Å away), and sits in a tight turn whose backbone it constrains as a proline. The economical reading of position 83 is therefore that proline-to-serine acts through local conformation or protein stability, which is consistent with the shorter VP24 half-life reported for the BDBV protein [5]. The prediction is qualitative in an unusual and useful way: it does not say that BDBV is weaker, which is the published claim. It says that the weakness is unevenly priced, and that the tissues where it costs least overlap with the compartments where persistence has been documented for filoviruses [3,9]. Two features distinguish it from the expression-based tropism models that have failed in this field. First, the host quantity multiplies a virus-intrinsic defect: a cell with high IFN responsiveness is predicted to extract a larger penalty once infection has occurred. Second, the prediction is stated per cell type within an organ, which makes it falsifiable at a resolution that organ-level pathology cannot reach. The strongest rival reading is that the gradient is an artefact of the score, because baseline expression is only a proxy for the responsiveness that prices the defect. In primary human immune cells from healthy donors the score predicts the additional IFN response with the wrong sign while tracking the level the cells reach, so the score is read as a defined axis of baseline IFN tone and the post-IFN level is its operational phenotype. A score built to track the additional response a cell can still mount instead of the level it attains correlates with the frozen score at -0.21 across the 154 cell types and would reverse the assignment of six of the seven compartments named above, leaving circulating myeloid cells as the only named compartment that stays high on both; the level-priced reading is therefore the prediction of record, and that comparison is one of the analysis outputs listed under Data availability. The cell types in which the two accounts disagree are the ones that discriminate between them, and excluding a pure score artefact would require a cell type that scores high on both axes to show a larger relative penalty than one that scores low on both when both are measured. Finally, the species comparison is asymmetric: *Reston virus* VP24 binds KPNA with EBOV-like affinity [5] while being non-pathogenic in humans. The cross-species comparison was exhaustive for the records returned on the search date. The tissue-level extension did not recover agreement with the only multi-organ infection series available [16].

If the ordering is a property of the virus rather than of the score, it carries clinical consequences that need no new data. The enteric epithelium, where the IFN axis places the greatest penalty, is predicted to answer infection with an epithelial response rather than unchecked replication: gastrointestinal signs are prominent in severe BDBV

disease [17], filovirus infection disrupts epithelial barrier function and ion transport in human intestinal organoids [18], and barrier support and anti-inflammatory management become candidate adjuncts in trial design. The testicular and posterior ocular compartments lie at the other end: persistence has been documented for BDBV in semen [9], the seminiferous epithelium is immune-privileged [19], *Ebola virus* persists in semen after clearance from blood [20], and the retinal pigment epithelium raises its IFN-stimulated module once infected [21], while *Ebola virus* also persists in ocular fluid during convalescence [22]. The defect is therefore priced least behind the blood-testis and blood-retinal barriers, which predicts extended semen monitoring and counselling, and ocular fluid sampling across the anterior and posterior segments in uveitis after acute disease [21,22]. These are consequences of an ordering hypothesis.

Two cell types at the top of the axis lie outside the named sets and carry their own clinical readings. The highest-scoring cell type in the atlas is the extravillous trophoblast, the fetal cell that invades the decidua and contacts maternal blood, although the trophoblast compartment ranks only sixth by median because its four members span most of the axis. Filovirus disease in pregnancy carries an absolute risk of maternal death of 67.8%, fetal loss of 76.9% and neonatal death of 98.5% [24], and the one BDBV pregnancy reported in detail ended with maternal and infant death after a live birth, with viral antigen in the syncytiotrophoblast [23]. Trophoblasts are protected by type III interferon [25], the arm that the BDBV-type configuration fails to antagonise [6], so the maternal-fetal interface is where the defect should be priced most heavily and where the gradient can be tested in placental tissue. The oesophageal apical cell, the differentiated layer that ingested virus meets first, ranks third (suprabasal and basal cells rank 36th and 46th), and sore throat or difficulty swallowing was reported in 70.6% of patients treated in the 2012 BDBV outbreak [26] and is more frequent in fatal than in surviving cases across Ebola virus disease cohorts [27]; the local reading is that an epithelial response the virus cannot blunt adds to odynophagia and reduced intake, which is a hypothesis about local pathology rather than about virulence.

Limitations

Several boundaries apply. The framework states a relative-cost ordering, not a tropism prediction: it is silent on which cells are infected first, it does not predict absolute permissiveness from expression alone, and it does not extend from IFN handling to virulence. The sequence census carries three limits. Exclusivity of the four-position configuration is tightly bounded only for EBOV, *Sudan virus* and *Reston virus*, and only weakly bounded for *Bombali virus* and *Tai Forest virus*, so absence from those two species is not established. Within BDBV, two distinct protein sequences carry no power to detect invariance, so the discriminating evidence is the comparison across species rather than within the species (Supplementary Note 4). The two 2026 departures are consensus-only records from a single outbreak: they cannot be evaluated for depth, allele fraction, strand balance or independent reproducibility, so neither recurrence nor a direction of change is established, and the wider re-read samples the same deposit stream rather than an independent cohort. A pilot molecular-dynamics comparison of position 83 was uninformative and is not used as support.

The host axis is measured; the restriction it is asked to predict is not. The score is a baseline-tone axis and not a validated proxy for IFN responsiveness: it tracks the level a cell attains, and in the one panel carrying a functional phenotype that level could not be separated from the baseline it is computed from. That panel reports HIV-1 restriction rather than filovirus restriction, and its level advantage does not survive the family-wise correction for the contrast sweep ($p = 0.056$ over the nine-unit panel, 0.063 over the eleven-unit panel; Supplementary Note 1). Neither same-cell panel contains the primary epithelium the prediction concerns; the two instruments that exist for that material, a differentiated airway epithelium series [15] and a keratinocyte series, do not test filovirus restriction, and the cancer-cell-line panel profiled with them [12] was unusable as a graded test because its response was binary. The rival importin- α account is likewise unvalidated: its transcript layer is a proxy for protein capacity (*KPNA1* 0.18, *KPNA6* 0.48, *KPNA5* protein undetected in 24 of 29 tissues), the two axes are only weakly correlated, so the pre-specified separation criterion is permissive rather than evidence of independence,

and the same functional panel separates restricted from permissive cells no better than chance (AUC 0.47 to 0.50). The enteric epithelium is the compartment whose prediction depends most on which layer is read, because that resource annotates KPNA5 protein as enriched in intestine while its transcripts are the lowest of any epithelial set. Coverage bounds how far the ordering can be transported. Cross-atlas agreement rests on ontology labels matched for 97 of the 154 reference cell types, no local atlas carries urothelial, retinal pigment epithelial, photoreceptor, Sertoli or choroid-plexus cells, and the coverage repair that added them does not raise agreement (Supplementary Note 4). The urothelial placement is not recovered in an independent ureter atlas, the choroid plexus remains uninformative rather than negative, and the retinal and testicular placements rest on few labels, so the low end is better established than the high end; a positive control for the per-compartment rule also failed, with enterocytes reaching only +0.28 standard deviations in a 44-label atlas. The tissue-level extension did not recover agreement with the only multi-organ infection series available [16] across 45 combinations of score form and readout, the sign is unstable across conventions, and deconvolution shows that immune infiltration does not explain the failure (0.35 to 5.3% across the seven tissues other than lymph node, 14.6% in lymph node), so the ordering has not been transported to tissue-level observables. In the protein layer, three of the four receptor genes are quantified in at most one of 27 cell types from a single donor, so the receptor arm cannot be tested. Two literature comparisons bound the framework without settling it: the keratinocyte series [28] crosses IFN classes at doses differing by an order of magnitude, so its ordinal reading is conditional, and a cross-sample clinical IFN trial (GEO GSE107549) separated subjects no better than a housekeeping module. The data that would decide the largest remaining question, whether the defect is priced by the tone a cell already carries or by the response it can still add, do not exist: the decisive test is a matched comparison of BDBV and EBOV in one host cell type with host transcriptome readout and viral load as a covariate, and of the four GEO series carrying BDBV a human-restricted query returns none. The closest available design, intestinal organoids exposed to EBOV and *Marburg virus* [18], is unkeyed to its conditions. The framework is testable in that design, and its prediction is that cell types high on both axes should show a larger relative penalty than those low on both, the highest-value pair being enterocytes against urothelial cells. Neither the extravillous trophoblast nor the oesophageal apical cell was covered by the cross-atlas reproduction, so both placements rest on the reference table alone.

Conclusion

BDBV carries natively the four-position VP24 configuration that was engineered to abolish IFN- λ 1, IFN- β and *ISG15* antagonism. That configuration is fixed across the 2007, 2012 and 2026 outbreaks, apart from two 2026 sequences that each show an unconfirmed single-position departure. It is exclusive to the species among the six assessed, and two of its four positions are found in no other species. A cell-type-resolved score of baseline IFN tone, read operationally as the IFN-stimulated level a cell attains, yields an ordering that is stated as an ordering hypothesis, and a weakly correlated importin- α account that would reverse the penalty for defined cell types is carried as its rival. On that axis the partition places the cost of the defect in Urothelial cells, the airway ciliated cell and in circulating myeloid cells, and places it lowest in the retinal and seminiferous compartments. The framework is falsifiable by specified experiments.

Supplementary material

Fig. S1-S4, Supplementary Notes 1-6, Supplementary Methods and Supplementary Table S1 accompany this article in the supplementary PDF. Supplementary Table S1 gives the baseline IFN-tone score for all 154 reference cell types, of which Table 2 lists the extremes. All supplementary items are cited in the main text.

CRedit authorship contribution statement

Wenhao Huang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review and editing, Visualization.

Declaration of competing interest

The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical statement

This study analysed only pre-existing, publicly available and de-identified data. No human participants, human tissue, animals or identifiable personal data were involved, so the work is exempt from ethical approval as research using only public data, and no informed consent to participate was required.

Data availability

All primary data analysed here are public and are identified by repository accession in Methods and References. These comprise protein sequences from UniProt (VP24 references Q05322, B8XCN4 and R4QRB9), the experimental complex from the Protein Data Bank (4U2X), and the human cell-type expression reference from the Human Protein Atlas (v25.1). The GEO series analysed here are GSE21158, GSE46599, GSE114905, GSE226106, GSE327707, GSE306664, GSE342661, GSE309699, GSE100839, GSE300073, GSE298600 and GSE107549, of which GSE226106 is the tissue series behind the tissue-level analysis reported in Results; each accession can be retrieved from the NCBI Gene Expression Omnibus at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=<accession>>. Further sources are VP24 protein records from NCBI Protein and 2026 outbreak genomes from NCBI Nucleotide (NC_002549.1, NC_014373.1, KC545393.1, FJ217161). A real-time surveillance instance that aggregates deposited genomes (Pathoplexus LAPIS) was also used. The cross-atlas comparison screened nineteen datasets, eighteen of which contributed the 162 matched pairs of the primary analysis; seven further atlas resources added for coverage are a coverage repair and not a second version of the same analysis. The sequence census is reported with per-record accessions, collection dates and residue calls, so that every count stated here can be re-derived from the archive; the per-record accessions and residue calls for the 2026 outbreak genomes are held in the repository directory `02_primary_sequence_data/outbreak_2026_genomes/`, alongside the call tables listed in the file manifest. The derived tables, the analysis scripts that generate every number reported here and the per-figure source data are openly available in the analysis data repository at <https://github.com/Penicillinw/BDBV-VP24-Data> and archived at <https://doi.org/10.5281/zenodo.23125404>. That repository also carries a file manifest, per-file SHA-256 checksums, a data dictionary and reproduction notes. The versions of the software used to run the analyses, the entry point of each script and the random seed are listed with the deposited analysis scripts. Third-party raw data are not redistributed there; each such file is registered with its accession, retrieval date and checksum so that it can be obtained from the original resource.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author used generative AI tools in order to draft and check analysis code and to edit the language of the manuscript. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Abbreviations

AUC, area under the curve; BDBV, *Bundibugyo virus*; BOMV, *Bombali virus*; EBOV, *Ebola virus*; ENA, European Nucleotide Archive; GEO, Gene Expression Omnibus; HPA, Human Protein Atlas; IFN, interferon; ISG, interferon-stimulated gene; KPNA, karyopherin alpha; LAPIS, Lightweight API for Sequences; nCPM, normalised counts per million; PDB, Protein Data Bank; RESTV, *Reston virus*; SUDV, *Sudan virus*; TAFV, *Tai Forest virus*; VP24, viral protein 24; WHO, World Health Organization. Additional abbreviations: BV-BRC, Bacterial and Viral Bioinformatics Resource Center; ICTV, International Committee on Taxonomy of Viruses; NCBI, National Center for Biotechnology Information; NK, natural killer; RCSB, Research Collaboratory for Structural Bioinformatics; Δ , the cell-type-resolved IFN-tone score.

Acknowledgements

The author thanks the developers and maintainers of the public resources on which this study depends — UniProt, the Protein Data Bank, the Human Protein Atlas, the NCBI Gene Expression Omnibus, NCBI Nucleotide and Pathoplexus — for making their data openly available. Generative AI tools were used for language editing and for drafting and checking analysis code; all scientific content, data analysis and conclusions were verified and are the responsibility of the author.

References

1. World Health Organization. Epidemic of Ebola disease caused by *Bundibugyo virus* in the Democratic Republic of the Congo and Uganda determined a public health emergency of international concern. Geneva: WHO; 17 May 2026. <https://www.who.int/news/item/17-05-2026-epidemic-of-ebola-disease-in-the-democratic-republic-of-the-congo-and-uganda-determined-a-public-health-emergency-of-international-concern> (accessed 4 October 2026).
2. World Health Organization. Ebola disease caused by *Bundibugyo virus* - Democratic Republic of the Congo. Disease Outbreak News, 25 September 2026 (DON618; data as of 23 September 2026). Geneva: WHO; 2026. <https://www.who.int/emergencies/disease-outbreak-news/item/2026-DON618> (accessed 4 October 2026).
3. Fenton KA, Pigeaud DD, Turcinovic J, Prasad AN, Agans KN, et al. Pathogenesis and natural history of the Bundibugyo species of *Orthoebolavirus* in nonhuman primates. bioRxiv. 2026. <https://doi.org/10.64898/2026.08.10.743937>. Preprint; not peer reviewed.
4. Xu W, Edwards MR, Borek DM, Feagins AR, Mittal A, et al. *Ebola virus* VP24 targets a unique NLS binding site on karyopherin α 5 to selectively compete with nuclear import of phosphorylated STAT1. Cell Host Microbe. 2014;16(2):187-200. <https://doi.org/10.1016/j.chom.2014.07.008>
5. Schwarz TM, Edwards MR, Diederichs A, Alinger JB, Leung DW, et al. VP24-karyopherin α binding affinities differ between ebolavirus species, influencing interferon inhibition and VP24 stability. J Virol. 2017;91(4):e01715-16. <https://doi.org/10.1128/jvi.01715-16>
6. Ramanathan P, Tigabu B, Santos RI, Illykh PA, Kuzmina N, et al. *Ebolavirus* species-specific interferon antagonism mediated by VP24. Viruses. 2023;15(5):1075. <https://doi.org/10.3390/v15051075>
7. He FB, Khan H, Huttunen M, Kolehmainen P, Melén K, et al. Filovirus VP24 proteins differentially regulate RIG-I and MDA5-dependent type I and III interferon promoter activation. Front Immunol. 2021;12:694105. <https://doi.org/10.3389/fimmu.2021.694105>
8. Khan H, Tripathi L, Kolehmainen P, Lundberg R, Altan E, et al. VP24 matrix proteins of eight filoviruses downregulate innate immune response by inhibiting the interferon-induced pathway. J Gen Virol. 2023;104(8):001888. <https://doi.org/10.1099/jgv.0.001888>
9. Uhrig A, Pfäfflin F, Tober-Lau P, Corman VM, Stegemann MS, et al. A case of *Bundibugyo virus* disease treated with monoclonal antibodies and remdesivir. Nat Med. 2026. <https://doi.org/10.1038/s41591-026-04663-5>.

10. Park YC. Comparative whole-ORF evolutionary analysis of the seven canonical proteins of *Bundibugyo ebolavirus* across the 2007, 2012 and 2026 outbreaks. Research Square. 2026. <https://doi.org/10.21203/rs.3.rs-10801608/v1>. Europe PMC PPR1304528. Not peer reviewed.
11. Goujon C, Moncorgé O, Bauby H, Doyle T, Ward CC, et al. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature*. 2013;502(7472):559-62. Data: GEO GSE46599, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46599> (accessed 4 October 2026). <https://doi.org/10.1038/nature12542>
12. Siegrist F, Ebeling M, Certa U. The small interferon-induced transmembrane genes and proteins. *J Interferon Cytokine Res*. 2011;31(1):183-97. Data: GEO GSE21158, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21158> (accessed 4 October 2026). <https://doi.org/10.1089/jir.2010.0112>
13. [dataset] Wood O, Braithwaite AT, Jackson-Wood MA. Sorted bulk RNA-seq profiling of human peripheral blood mononuclear cell populations across diverse in vitro immune stimuli. NCBI Gene Expression Omnibus; 2026. Accession GSE327707. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE327707> (accessed 4 October 2026).
14. [dataset] Moss N, Sakai C, Kaul SN, Layton ED, Bouvatte P, et al. Revealing differential type I and type II IFN responses in different immune cells in various human diseases by a new IFNs transcriptomic atlas. NCBI Gene Expression Omnibus; 2026. Accession GSE306664. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE306664> (accessed 4 October 2026).
15. Tay DJW, Tay HY, Chu JJH, Tan KS. Differential host responses by the type I and III interferons in primary differentiated human nasal epithelial cells are a determinant of antiviral activity. *Front Pharmacol*. 2026;17:1895412. Data: GEO GSE342661, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE342661> (accessed 4 October 2026). <https://doi.org/10.3389/fphar.2026.1895412>
16. Normandin E, Triana S, Raju SS, Lan TCT, Lagerborg K, et al. Natural history of Ebola virus disease in rhesus monkeys shows viral variant emergence dynamics and tissue-specific host responses. *Cell Genom*. 2023;3(12):100440. Data: GEO GSE226106, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE226106> (accessed 4 October 2026). <https://doi.org/10.1016/j.xgen.2023.100440>
17. MacNeil A, Farnon EC, Wamala J, Okware S, Cannon DL, et al. Proportion of deaths and clinical features in *Bundibugyo Ebola virus* infection, Uganda. *Emerg Infect Dis*. 2010;16(12):1969-72. <https://doi.org/10.3201/eid1612.100627>
18. Flores EY, Hume AJ, Olejnik J, Mithal A, D'Amico A, et al. Filovirus infection disrupts epithelial barrier function and ion transport in human iPSC-derived gut organoids. *PLoS Pathog*. 2025;21(11):e1013698. Data: GEO GSE300073, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE300073>, and GSE298600, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE298600> (both accessed 4 October 2026; human intestinal organoids, EBOV and Marburg virus, mock controls at 1 and 3 days). <https://doi.org/10.1371/journal.ppat.1013698>
19. Zhao S, Zhu W, Xue S, Han D. Testicular defense systems: immune privilege and innate immunity. *Cell Mol Immunol*. 2014;11(5):428-37. <https://doi.org/10.1038/cmi.2014.38>
20. Deen GF, Broutet N, Xu W, Knust B, Sesay FR, et al. Ebola RNA persistence in semen of *Ebola virus* disease survivors — final report. *N Engl J Med*. 2017;377(15):1428-37. <https://doi.org/10.1056/NEJMoa1511410>
21. Smith JR, Todd S, Ashander LM, Charitou T, Ma Y, et al. Retinal pigment epithelial cells are a potential reservoir for *Ebola virus* in the human eye. *Transl Vis Sci Technol*. 2017;6(4):12. Data: GEO GSE100839, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100839> (accessed 4 October 2026; ARPE-19 cells, mock and EBOV at 24 hours). <https://doi.org/10.1167/tvst.6.4.12>
22. Varkey JB, Shantha JG, Crozier I, Kraft CS, Lyon GM, et al. Persistence of *Ebola virus* in ocular fluid during convalescence. *N Engl J Med*. 2015;372(25):2423-7. <https://doi.org/10.1056/NEJMoa1500306>
23. Muehlenbachs A, de la Rosa Vázquez O, Bausch DG, Schafer JJ, Paddock CD, et al. Ebola virus disease in pregnancy: clinical, histopathologic, and immunohistochemical findings. *J Infect Dis*. 2017;215(1):64-9. <https://doi.org/10.1093/infdis/jiw206>
24. Kayem ND, Benson C, Aye CYL, Barker S, Tome M, et al. Ebola virus disease in pregnancy: a systematic review and meta-analysis. *Trans R Soc Trop Med Hyg*. 2022;116(6):509-22. <https://doi.org/10.1093/trstmh/tra180>
25. Bayer A, Lennemann NJ, Ouyang Y, Bramley JC, Morosky S, et al. Type III interferons produced by human placental trophoblasts confer protection against Zika virus infection. *Cell Host Microbe*. 2016;19(5):705-12. <https://doi.org/10.1016/j.chom.2016.03.008>
26. Kratz T, Roddy P, Tshomba Oloma A, Jeffs B, Pou Ciruelo D, et al. Ebola virus disease outbreak in Isiro, Democratic Republic of the Congo, 2012: signs and symptoms, management and outcomes. *PLoS One*. 2015;10(6):e0129333. <https://doi.org/10.1371/journal.pone.0129333>
27. Moole H, Chitta S, Victor D, Kandula M, Moole V, et al. Association of clinical signs and symptoms of Ebola viral disease with case fatality: a systematic review and meta-analysis. *J Community Hosp Intern Med Perspect*. 2015;5(4):28406. <https://doi.org/10.3402/jchimp.v5.28406>

28. Elliff J, Van Ert H, Sevcik K, Djurkovic MA, Rudd D, et al. Interferons inhibit *Ebola virus* infection of human keratinocytes. *Viruses*. 2025;17(12):1577. Data: GEO GSE309699, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE309699> (accessed 4 October 2026). <https://doi.org/10.3390/v17121577>

Tables

Table 1. Coverage of the four VP24 positions across the six orthoebolavirus species.

Species	Records retrieved	Sequences assessed	Called at all four positions	Motif	BDBV motif	Residues shared
BDBV	32	31	31	SQHA	Yes	S/Q/H/A
EBOV	3302	3271	3253	PNRV	No	
SUDV	147	147	146	QNRV	No	
RESTV	27	27	27	PNSV	No	
TAFV	4	4	4	PQQA	No	Q/A
BOMV	10	10	10	PNSV	No	

Motif, the concatenated residue calls at EBOV VP24 positions 83, 135, 140 and 141 after gap-free alignment; BDBV motif, whether that row carries the Bundibugyo configuration S83/Q135/H140/A141; Residues shared, the positions at which the species carries the BDBV residue. Records retrieved counts every VP24 record returned by the species query, sequences assessed excludes fragments that do not span the four-position window, and called at all four positions excludes records in which at least one position could not be called.

Table 2. Baseline IFN-tone score by human cell type.

Segment	Rank	Cell type	Δ
Top 20	1	Extravillous trophoblasts	1.7743
Top 20	2	Plasmacytoid dendritic cells	1.414
Top 20	3	Esophageal apical cells	1.3885
Top 20	4	Plasma cells	1.3707
Top 20	5	Enterocytes	1.2837
Top 20	6	Neutrophils	1.2004
Top 20	7	Urothelial cells	1.1309
Top 20	8	Hofbauer cells	1.0186
Top 20	9	Kupffer cells	1.0144
Top 20	10	Monocytes	1.0018
Top 20	11	Macrophages	0.9482
Top 20	12	Prostatic club cells	0.8258
Top 20	13	B-cells	0.7781
Top 20	14	Microglia	0.7595
Top 20	15	Migrating cytotrophoblasts	0.7587
Top 20	16	Vascular endothelial cells	0.7253
Top 20	17	Pancreatic duct cells	0.7085
Top 20	18	Epididymal basal cells	0.7053
Top 20	19	Fallopian secretory cells	0.7031
Top 20	20	Colonocytes	0.6965

Segment	Rank	Cell type	Δ
Bottom 20	135	Megakaryocyte progenitors	-0.8223
Bottom 20	136	Fibro-adipogenic progenitors	-0.8304
Bottom 20	137	Megakaryocytes	-0.8655
Bottom 20	138	Adrenal cortex cells	-0.8899
Bottom 20	139	Megakaryocyte-erythroid progenitors	-0.9184
Bottom 20	140	Retinal horizontal cells	-0.9251
Bottom 20	141	Oocytes	-0.93
Bottom 20	142	Rod photoreceptor cells	-0.9999
Bottom 20	143	Retinal pigment epithelial cells	-1.0246
Bottom 20	144	Erythrocyte progenitors	-1.0374
Bottom 20	145	Adrenal medulla cells	-1.0481
Bottom 20	146	Platelets	-1.0529
Bottom 20	147	Cardiomyocytes	-1.0551
Bottom 20	148	Late primary spermatocytes	-1.115
Bottom 20	149	Adipocytes	-1.1312
Bottom 20	150	Early primary spermatocytes	-1.1793
Bottom 20	151	Cone photoreceptor cells	-1.2095
Bottom 20	152	Thymic myoid cells	-1.2599
Bottom 20	153	Retinal ganglion cells	-1.2906
Bottom 20	154	Erythrocytes	-2.0368
Named in text	21	Epididymal efferent-duct ciliated cells	0.6892
Named in text	23	Goblet cells	0.6336
Named in text	28	Respiratory ciliated cells	0.572
Named in text	29	Ocular surface epithelium	0.5707
Named in text	34	Hepatocytes	0.4669
Named in text	35	Alveolar cells type 1	0.4639
Named in text	61	Ependymal cells	0.1903
Named in text	94	Choroid plexus epithelial cells	-0.08
Named in text	114	Leydig cells	-0.356
Named in text	129	Sertoli cells	-0.6769

Δ , the cell-type-resolved IFN-tone score (Eq. 1), standardised across the 154 human cell types of the reference atlas; ranks run from the highest (1) to the lowest (154). Top 20 and Bottom 20 are the two extremes of that ranking, and Named in text lists further cell types cited in the Results.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialVirology.docx](#)
- [FigureS1.png](#)
- [FigureS2.png](#)
- [FigureS3.png](#)
- [FigureS4.png](#)