Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Natural genus-exclusive four-position VP24 configuration in Bundibugyo virus links host interferon tone to a testable cell-type ordering

Wenhao Huang

penicillinw@foxmail.com

Nanning No.3 High School

Research Article

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

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11661270/v1

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

Additional Declarations: No competing interests reported.

Original Article

Natural genus-exclusive four-position VP24 configuration in Bundibugyo virus links host interferon tone to a testable cell-type ordering

Wenhao Huang [a,*]

[a] *Nanning No.3 High School, No. 5 Qingshan Road, Qingxiu District, Nanning, Guangxi Zhuang Autonomous Region 530021, China*

* Corresponding author.

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

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. DeepSeek-V4.1-Flash-1017 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.

Abstract

Bundibugyo virus (BDBV) causes outbreaks with lower case-fatality than Ebola virus (EBOV); its VP24 carries the only BDBV-specific functional difference mapped to date. That difference was defined in an engineered virus. Here, we integrate a sequence census, a human cell-type expression atlas, and a reanalysis of public infection data to connect the natural BDBV VP24 configuration to host interferon (IFN) tone and derive a testable cell-type ordering of post-entry IFN-mediated restriction. All 31 full-length BDBV VP24 records and all 36 complete 2026 outbreak genomes carried S83/Q135/H140/A141; none of the other five orthoebolavirus species assessed carried the complete set. Two consensus-only single-position departures occurred among 860 callable copies in the surveillance data. Δ, a baseline IFN-tone score, tracks the IFN-stimulated level attained by a cell and predicts the strength of IFN-mediated suppression of viral RNA accumulation in cells infected with BDBV carrying the VP24 defect: high Δ denotes stronger suppression, low Δ weaker. Across 154 human cell types, Δ was higher in surface epithelia and circulating myeloid cells (enterocytes 1.28; monocytes 1.00) than in retinal and testicular cells (retinal pigment epithelium −1.02; Sertoli cells −0.68). A competing importin-α model predicted exemption in the 47 cell types with high Δ but low importin-α, including enteric epithelium and hepatocytes. In the HIV-1 restriction panel, Δ discriminated restricted from permissive cells (area under the curve (AUC) 0.87), whereas importin-α was at chance (AUC 0.47–0.50); functional data therefore descriptively support Δ.

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, provides a direct 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 3 November 2026 the Democratic Republic of the Congo had reported 10,668 confirmed cases and 5,198 deaths, a crude case-fatality ratio of 48.7 per cent, across 67 health zones in eight provinces. Kenya had reported nine cases, including two deaths, and Nigeria had confirmed one imported case [2]. This is also the second-largest Ebola disease outbreak on record. [2].

The phenotype of BDBV relative to EBOV is documented, but most genus-wide comparisons have excluded it. BDBV produces a more protracted, less lethal course than EBOV in non-human primates [3]. Its two previous outbreaks, in Uganda in 2007 and the Democratic Republic of the Congo in 2012, had crude case-fatality ratios of 30 and 50 per cent [2]. The virus is therefore often described by what it lacks relative to EBOV rather than by what it carries. The one BDBV-specific functional difference mapped to date concerns VP24, which competes with phosphorylated STAT1 for a non-classical nuclear-localisation-signal site on KPNA transporters [4] and thereby blunts 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 positions that distinguish the two proteins, the resulting virus no longer inhibited IFN-λ1, IFN-β or ISG15 [6]. BDBV VP24 expressed on its own, by contrast, inhibited IFN-β and IFN-λ1 promoter activation as effectively as EBOV VP24 [7], and inhibited IFN-α-induced *MX1* (MxA protein) and IFITM3 like most filovirus VP24 proteins [8]. The phenotype is context-dependent: isolated expression preserves IFN antagonism, whereas the four-position configuration determines the outcome in infected cells. We therefore determine the natural BDBV configuration, its species specificity, and the cell types that depend most on retained IFN responsiveness.

We used a sequence census across the genus, a cell-type-resolved human expression atlas and a reanalysis of public infection data to ask how a virus-intrinsic IFN-escape defect is priced across host cell types. The host variable enters as a multiplier on a defect the virus already carries, and three cost models make different predictions. If cost scales with the absolute IFN-stimulated level a cell attains, high-tone cells pay most (M1). If it scales with the inducible headroom that remains, the ordering reverses (M2). If it scales with importin-α

capacity, cells with high tone but low importin-α are exempt (M3). We first determine the natural BDBV state of the four-position configuration and its species specificity. We then determine which of M1 to M3 the functional data separate and define the measurements that test the resulting gradient.

Materials and 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 ≥95% identity to the reference protein. 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. 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. Sequence ambiguity accounted for every unresolved position, and each call was recorded together with its specific reason.

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. 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 here counts indexed sequences with a callable amino acid at a position, not records in the instance; the Results quote these per-position 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). 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 log10(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 *IFNLR1* 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).

$$y_{g,c} = \log10\left(nCPM_{g,c} + 1\right)$$
$$z_{g,c} = \frac{y_{g,c} - \mu_g}{\sigma_g} \quad (1)$$
$$\Delta(c) = \frac{1}{6}\left(z_{\mathrm{IFNAR1},c} + z_{\mathrm{IFNAR2},c} + z_{\mathrm{IFNLR1},c} + z_{\mathrm{IL10RB},c} + z_{\mathrm{ISG15},c} + z_{\mathrm{MX1},c}\right)$$

where c is a human cell type and g is one of the six Δ genes; $nCPM_{g,c}$ is the normalised counts-per-million value; $y_{g,c}$ is its log10($nCPM$ + 1) transform; μ_g and σ_g are the mean and population standard deviation of $y_{g,c}$ across the 154 cell types; and $z_{g,c}$ is the resulting per-gene z-score. Δ(c) is the unweighted mean of the six $z_{g,c}$ values.

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. 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, 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; 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 score-tracking 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, log10(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 robustness axes were computed: the maximum of the three importin-α z-scores, and the mean z-score of the PY-STAT1 cargo genes STAT1, STAT2 and IRF9.

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

where $z_{g,c}$ is defined in Eq. 1; $K(c)$ is the unweighted mean of $z_{KPNA1,c}$, $z_{KPNA5,c}$ and $z_{KPNA6,c}$; $K_{\max}(c)$ is their maximum; and $K_{\mathrm{STAT}}(c)$ is the unweighted mean of $z_{STAT1,c}$, $z_{STAT2,c}$ and $z_{IRF9,c}$.

Compartment membership used explicit cell-type name lists recorded with the analysis code, which capture the enteric, urinary, hepatic, airway and myeloid sets.

The two accounts were kept separate; writing d and k for Δ and for the importin-α axis standardised across the 154 cell types, Eq. 3 gives the candidate combination rules.

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

where $d(c) = \Delta(c)$, $k(c) = K(c)$, and *Q3* is the third quartile across the 154 cell types; $\tau = Q3[d(c)]$, $d_{\min} = \min_c d$, and $K_s = Q3[d(c)] - d_{\min}$. The subscripts add, mult, min, thr and sat denote the additive, product, minimum, threshold and saturating combination scores, respectively.

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 (rho 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 primary prediction (M1), and the importin-α account provides the competing prediction (M3); 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 with the reason, excluded from numerators and retained in denominators; only a called negative counted as evidence against 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 recorded versions of these packages are listed in the deposited environment file.

Use of Artificial Intelligence

During the preparation of this work the author used DeepSeek-V4.1-Flash-1017 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.

Results

Bundibugyo virus alone carries the complete four-position VP24 configuration across sequenced outbreaks

The engineered mutant that lost IFN-λ1, IFN-β and ISG15 inhibition carried BDBV-type residues at four positions that distinguish BDBV from EBOV VP24 [6]. EBOV P83, N135, R140 and V141 correspond to BDBV S83, Q135, H140 and A141, so the mutant reproduces the natural BDBV interface. We enumerated all BDBV VP24 protein records in NCBI. The query returned 32 records: 31 are full length and every one carries S83, Q135, H140 and A141; 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. Across the 36 complete genomes deposited from the 2026 outbreak, representing 35 independent isolates from four laboratories, every genome carries S83, Q135, H140, A141 and R184. Three carry ambiguity inside VP24 outside these positions and none within two codons; no callable position differs. Direct reading of the nucleotide archive also recovers position 83 as serine, matching all 31 full-length protein records. Across the two archives, retrieved at different times by different extraction routes, the

configuration is present in every BDBV outbreak with sequence data, an interval of nineteen years. A re-read of the same surveillance instance on 30 September 2026 returned 886 BDBV VP24 copies; 860 were callable at all four positions, 858 carried the complete configuration, and two consensus-only records from the 2026 outbreak each showed a single-position departure (S83L, A141V), with no record departing at more than one. The remaining 26 were unresolved, and neither departure was among the 36 complete genomes. The binding study reports the EBOV and BDBV residue at each position [5], the mutagenesis study reports the BDBV residues at all four [6], and a concurrent preprint reports minimal VP24 variation across the 2007, 2012 and 2026 outbreaks without naming the positions [9]. Earlier binding and mutagenesis studies characterized these residues individually, whereas this study identifies their four-position configuration as a species-level feature of BDBV.

To test the species specificity of the configuration, we aligned VP24 from all six species against the EBOV reference and read the four positions (Fig. 1a). Across 3,490 assessable sequences, S83/Q135/H140/A141 occurs only in BDBV, and in every BDBV sequence that spans the region in the protein archive. Positions 83 and 140 are each exclusive to BDBV: position 83 is serine only in BDBV, whereas the other species carry proline or, in Sudan virus, glutamine; position 140 is histidine only in BDBV, whereas the other five species carry arginine, glutamine or serine. Positions 135 and 141 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 pattern holds in UniProt, BV-BRC and ENA: every readable BDBV record carries all four residues, the per-source denominators are in the Fig. S1b caption, and 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 indexed non-BDBV VP24 copy (coverage 11,561 for Ebola virus, 579 for Sudan virus and 849 for BDBV). 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 consists of fragments and records with at least one uncallable position, which were left unscored. All four positions were invariant within EBOV itself, so the engineered mutant introduced the residues of a different species, not rare EBOV variants. The one-sided 95% exact confidence limit on the frequency of the complete configuration when 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 from that species' callable count (3,253, 146, 27, 10 and 4). The species-level feature requires all four positions: the same protein inhibited IFN induction when expressed alone but not in infected cells carrying the four positions. The complete four-position configuration is therefore the trait, present as the dominant configuration in every sequenced BDBV outbreak and absent from the other species assessed. The configuration is species-level, and the engineered mutant reconstructs an interface found nowhere else in the genus (Fig. S1b).

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). Three of the four positions that define the configuration, 135, 140 and 141, are contacts; position 83 is not. It lies 11.4 Å from the nearest KPNA5 atom and has no KPNA5 atom within an 8 Å shell; its closest atom belongs to a second VP24 chain, 5.3 Å away. Position 83 lies in a tight turn whose backbone a proline would constrain; its replacement by serine acts through local conformation or protein stability, consistent with the shorter half-life of BDBV VP24 [5]. The configuration therefore combines three contact positions with one that acts through conformation.

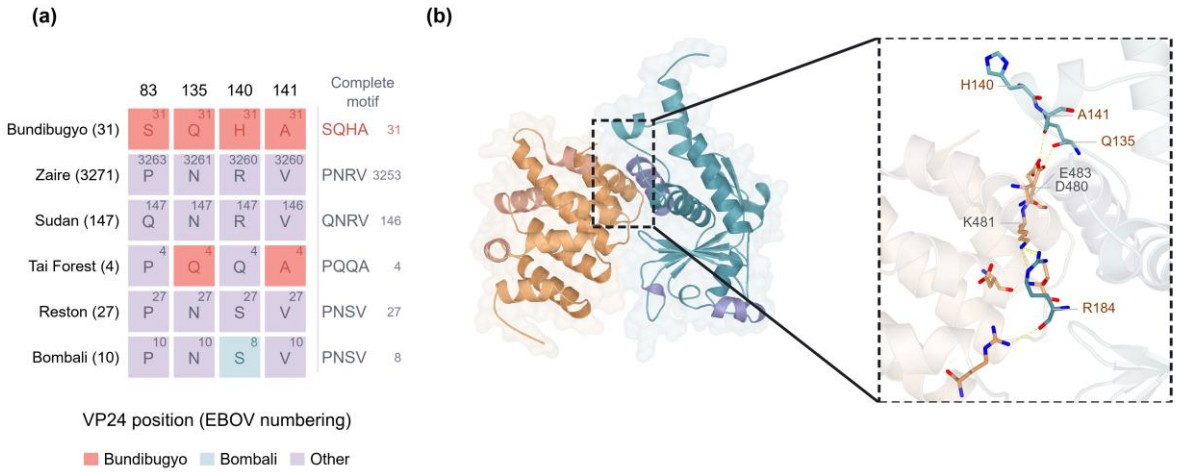


Fig. 1 Bundibugyo virus alone carries the complete four-position VP24 configuration. (a) Residue calls at EBOV VP24 positions 83, 135, 140 and 141 across the six orthoebolaviruses; each cell gives the major call and its record count, the row label gives sequences assessed, and the right column gives the assembled motif. Of 3,522 retrieved VP24 records, 32 fragments outside the four-position window were excluded from the denominator, 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. 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 define a reproducible cell-type IFN-tone score

Δ is the cell-type-resolved baseline IFN-tone score. It is built from six genes in three equal-weight components: type-I receptor capacity (*IFNAR1* and *IFNAR2*), type-III receptor capacity (*IFNLR1* and *IL10RB*), and baseline IFN-stimulated priming (*ISG15* and *MX1*). The receptor components set the IFN signal a cell can receive; the priming term reports the IFN-stimulated state it sustains at baseline. Each gene enters as normalised counts per million from the Human Protein Atlas single-cell-type reference, transformed as log10(x + 1) and z-scored across the 154 cell types with the population standard deviation. Δ is the unweighted mean of the six z-scores (Methods, Eq. 1), and is therefore expressed in the z units of its inputs, centred at zero by construction. A positive value means that the cell type lies above the atlas mean on average across the six genes; a negative value means that it lies below. Δ is read as a baseline-tone axis whose operational readout is the IFN-stimulated level a cell type attains. Under M1, with the viral defect constant, higher Δ predicts stronger post-entry IFN-mediated suppression of viral RNA accumulation and lower Δ weaker. Δ therefore supplies the ordering of relative cost, and any strictly increasing transformation leaves that ordering unchanged.

The type-III receptor shows the expected epithelial bias: Ocular surface epithelium (113.5 nCPM, normalised counts per million), Urothelium (102.5), Enterocytes (101.8) and Epididymal efferent-duct ciliated cells (91.5) rank highest. Plasmacytoid dendritic cells (110.4), Plasma cells (46.7) and Hepatocytes (38.6) are also high. 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 rather than the type-III receptor. Figure 2a shows the 17-member panel, which spans compartments implicated in filovirus infection, disease signs and persistence rather than the extremes of the ranking. Table 2 lists the top 20, the bottom 20 and ten further cell types named in the text; the complete 154-type ranking is provided in Table S1 (separate Excel file) and the source data. The panel contains 12 types with positive Δ (the high-Δ panel) and 5 with negative Δ (the low-Δ panel).

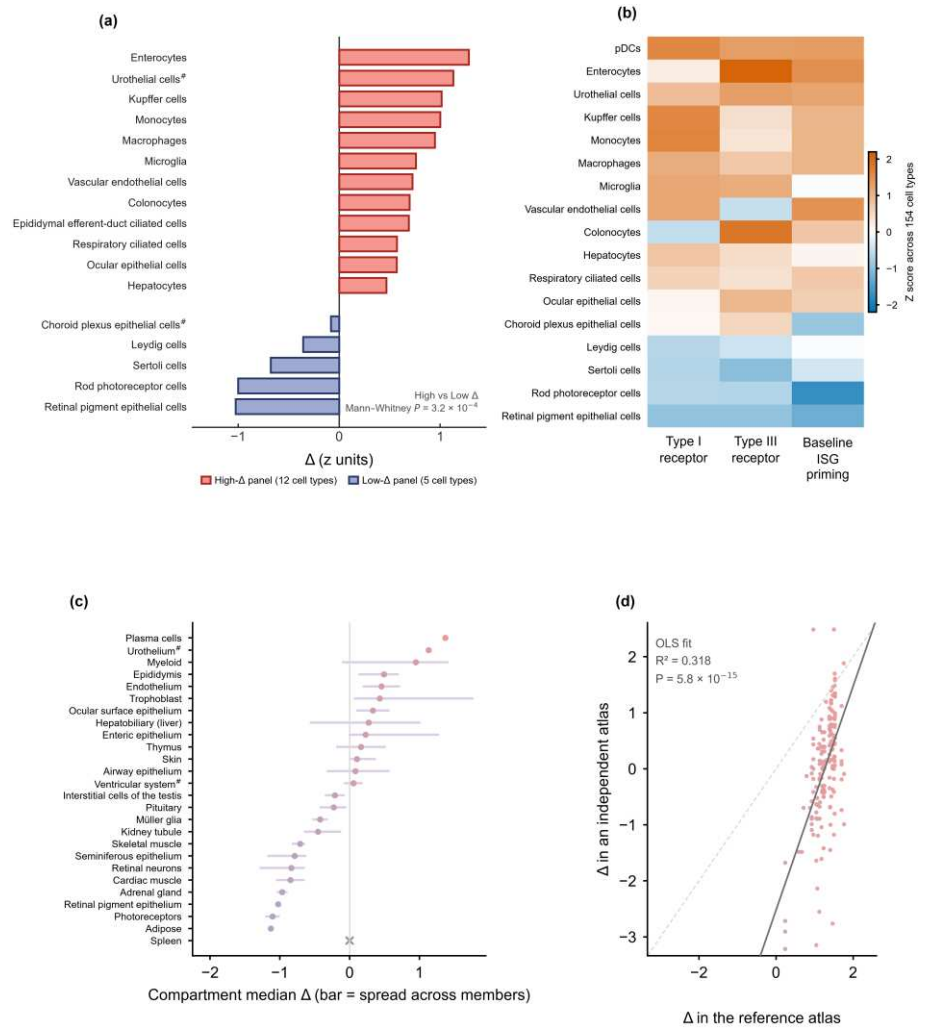


Fig. 2 A cell-type-resolved IFN-tone axis partitions the reference atlas more sharply than organs. (a) The 17 named members of the frozen score Δ, ordered from the highest to the lowest Δ and drawn as horizontal bars from zero. The 12-member High-Δ panel (positive Δ) is salmon and the 5-member Low-Δ panel (negative Δ) is periwinkle. Superscript hash signs (#) mark the two members that are coverage-limited in independent atlases. The high- and low-Δ panels were compared with a two-sided exact Mann–Whitney test ($P = 3.2 \times 10^{-4}$). (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 ×, 0 of 3 members; Supplementary Note 5), and a superscript hash sign 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. The dashed line is the identity line and the solid line is the ordinary least-squares fit ($R^2 = 0.318$, two-sided $P = 5.8 \times 10^{-15}$). (d) also 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

Two same-cell panels identify the quantity the score tracks and establish its use as a multiplier. A cancer-cell-line panel [10] is reported with them in Supplementary Notes 1 and 2 (Figs. S2, S3). In the first panel, every member responds; baseline tone explains essentially none of the variation in the response (-0.08, exact permutation p = 0.84), whereas the induced level tracks it (+0.88, p = 0.003, n = 9). This is the pattern expected if cells converge on a similar post-IFN level, and it supports M1 over M2. In the second panel, primary human immune cells from six healthy donors across five populations [11], the relation reverses for the additional

response, and the reversal reproduces in an independent cohort [12] (Monocytes -1.00 in each; exact p = 0.003 and 0.017; Fig. S3c, d). The pattern is convergence. Two further instruments, a differentiated airway epithelium series [13] and a keratinocyte series, cover the primary epithelium relevant to the prediction. Together, these panels define Δ as a baseline-tone multiplier: higher Δ predicts stronger post-entry IFN-mediated suppression of viral RNA, not entry.

The ordering survives variants of the gene set and score tested here. We recomputed Δ under five variants: dropping any single gene, replacing the two-gene priming term with a seventeen-gene IFN-stimulated module (Supplementary Note 1), ranking instead of z-scoring, taking the first principal component, or omitting the log transformation. Correlations with the frozen score were 0.925 to 0.984, and every variant retained at least 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, moved the correlation to 0.736 to 0.965 and retained 6 to 12 of the 15. Leave-one-out cross-validation supported equal weighting: fitted weights gave out-of-sample rank correlations of 0.00 to -0.07, and equal weighting was retained 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. We mapped the 154 reference cell types onto Cell Ontology terms and rebuilt the same three components from raw counts in eighteen of the nineteen independent single-cell datasets screened. The correlation was 0.519 over 162 matched labels (permutation p = 0.0002), and 0.452 when each dataset was standardised over only its matched labels. Leave-one-atlas-out resampling gave 0.493 to 0.555, so no single dataset drove the result; kidney was the sole negative value (Fig. 2d). 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 (Fig. S4a-d; Supplementary Note 3). Δ therefore behaves as a single operational quantity, the IFN-stimulated level a cell attains, and the ordering is stable across gene sets, score variants and atlases.

An 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 BDBV VP24 competes poorly. A score built only from IFN receptors and baseline IFN-stimulated transcripts omits this host component: a cell with high IFN capacity but little importin-α has few of the sites the competition targets and should pay little cost if competition is the only route (M3). This account is the competing model to M1, 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. The axis is defined at transcript level. Across the 154 cell types the Spearman correlation between Δ and the importin-α axis is -0.27, inside the pre-specified bound of |ρ| below 0.6, and the axis spans an interquartile range of 0.90 z units. The three genes form a heterogeneous set (*KPNA5* and *KPNA6* correlate at +0.27), and the axis is their equal-weight mean. 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 (Fig. 3a); the 47 in each disagreement quadrant discriminate between the two accounts.

Quadrant membership at the two extremes is unchanged across median, tertile, quartile, mixture and k-means cuts of the frozen score. The six extremes also survive a 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. The enteric epithelium is the most IFN-responsive epithelium in the atlas, and on M3 that responsiveness does not translate into a BDBV-specific disadvantage there. Fig. 3b gives the assignment of every compartment under both accounts.

The functional panel tests the two accounts. In the absence of an EBOV/BDBV functional panel, we used the currently available HIV-1 panel as the closest accessible experimental proxy. In the IFN panel that records

which members become restricted to HIV-1 after IFN and which remain permissive [10], the three-component score reached an AUC of 0.87, while the importin-α axis reached 0.47 to 0.50 depending on how the three genes are aggregated (Fig. 3c). The exact paired label-swap comparison over the same eleven units gave 0.37 (two-sided p = 0.23, one-sided p = 0.11), and the nine-unit subset excluding both partially resistant lines gave 0.40 (two-sided p = 0.29). The functional data therefore descriptively support Δ, not the importin-α model, as the discriminator in this panel. The importin-α axis remains the competing model: its reversed penalty for an enumerable set of cell types defines the discriminating predictions, while M1, the IFN-tone axis alone, remains the primary prediction.

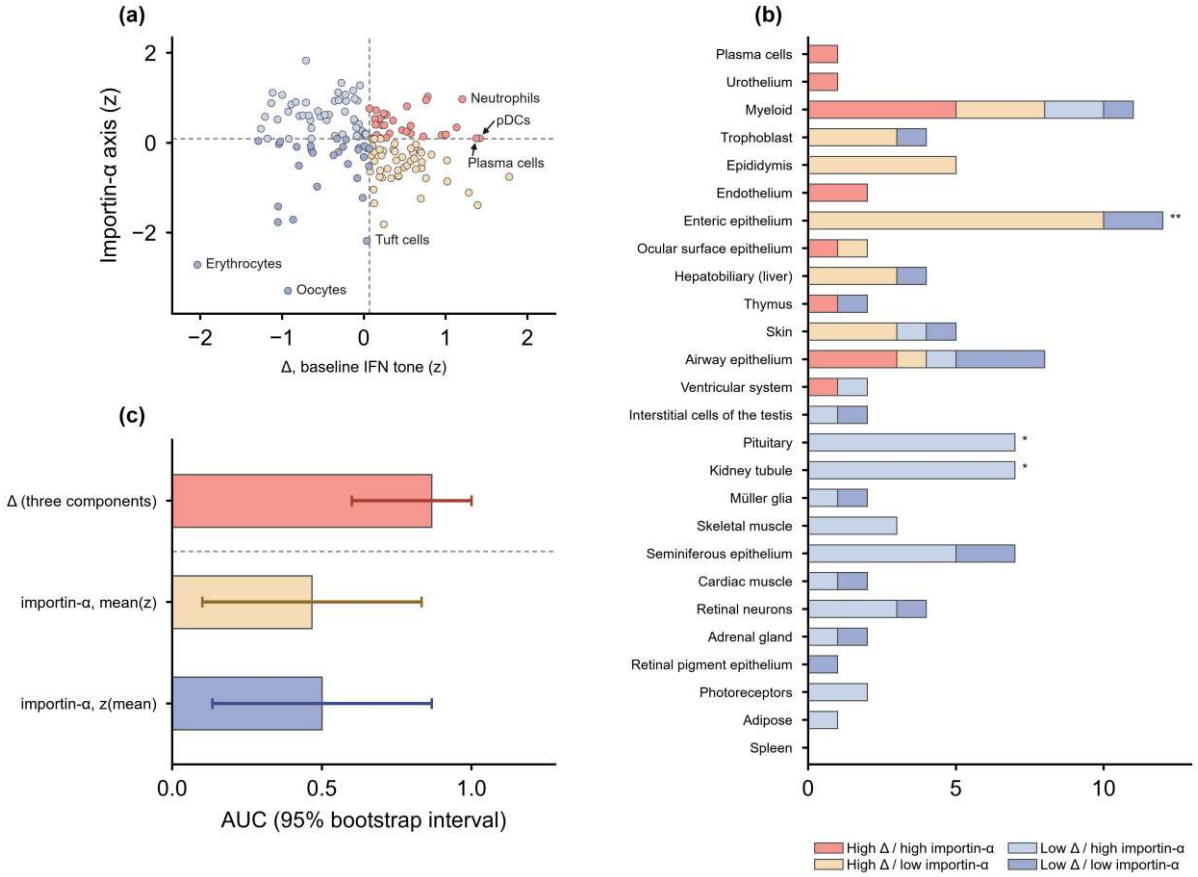


Fig. 3 An importin-α axis reclassifies selected compartments. (a) The 154 reference cell types on the Δ and importin-α axes; dashed lines are the two axis medians; high and low follow a two-axis median split (quadrant sizes 30, 47, 47 and 30); points are coloured by quadrant; the three most extreme members 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). 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

The predicted IFN-restriction gradient is cell-type specific and reverses under the importin-α model

Because the viral defect is the same in every cell type, cost is proportional to Δ (cost = c × Δ), so the ordering of Δ carries the information about relative cost. We therefore report that ordering as relative cost. Within the 17-member panel (Fig. 2a), the high end comprises Enterocytes, Urothelium, Kupffer cells, Monocytes, Macrophages, Microglia, Vascular endothelial cells, Colonocytes, Epididymal efferent-duct epithelium, Respiratory ciliated cells, Ocular surface epithelium and Hepatocytes; the low end comprises Retinal pigment epithelium, rod photoreceptors, Sertoli cells and Leydig cells. Two of the four higher-scoring types outside the panel, extravillous trophoblasts and oesophageal apical cells, are considered in the Discussion; Table 2 lists the ranked extremes and named cell types, and Fig. 2c gives the compartment medians; the complete ranking is provided in Table S1 (separate Excel file). The ordering is stated per cell type because compartments are internally heterogeneous, and it separates cell types within the same organ: surface epithelium scores high while retinal pigment epithelium and photoreceptors are among the lowest; efferent-duct epithelium scores high while Sertoli and Leydig cells are low; microglia score high while astrocytes, neurons and choroid-plexus epithelium score low. Each pair therefore predicts where post-entry IFN-mediated suppression of viral RNA accumulation should be stronger or weaker within the same organ. On M3 the penalty would be confined to the 30 cell types high on both axes, including type 1 alveolar cells; the enteric epithelium and hepatocyte lineage would be exempt. Enterocytes and colonocytes combine high Δ with the lowest importin-α levels of any epithelial set, so M1 and M3 predict opposite penalties for these cells. Choroid-plexus epithelium scores below the median (rank 94 of 154) while ependymal cells score above it (rank 61). In a multiplicity-matched dataset, a hepatocyte-lineage line, BDBV produced less viral mRNA than EBOV, the direction predicted by M1 and consistent with the Δ-based prediction (Fig. S2c, d; Supplementary Note 6).

Together, these results define a predicted gradient of post-entry IFN-mediated suppression of BDBV RNA accumulation. BDBV carries a species-specific VP24 defect, baseline IFN tone orders the cost of that defect across human cell types, and the enteric epithelium and hepatocyte lineage are the cells where an importin-α model reverses the prediction. The functional panel descriptively supports Δ, not that model, as the discriminator.

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 carries the combination. A configuration first characterised in an engineered construct is a natural feature of this species. Three of the four positions contact KPNA5, while position 83 acts through local conformation or stability. The prediction assigns the defect a cell-type-specific price. The tissues where the defect costs least overlap with compartments where filovirus persistence has been documented [3,14]. The model differs from expression-based tropism models in two ways. First, the host quantity multiplies a virus-intrinsic defect, so a cell with high IFN responsiveness pays a larger penalty once infected. Second, the prediction is made per cell type within an organ, capturing differences that organ-level pathology averages away.

In primary human immune cells from healthy donors, the score predicts the additional IFN response with the opposite sign and tracks the level the cells reach. We therefore read it as an axis of baseline IFN tone, with the post-IFN level as its operational phenotype. A score built to track the additional response a cell can mount, correlates with the frozen score at -0.21 across the 154 cell types. It reverses the assignment of six of the seven compartments named above and leaves circulating myeloid cells as the sole named compartment high on both. We adopt the level-priced reading. Reston virus VP24 binds KPNA with EBOV-like affinity [5] yet is non-pathogenic in humans; its attenuation therefore requires a property beyond binding affinity.

The ordering links the lower case-fatality of BDBV relative to EBOV to two classes of tissue: the epithelial and myeloid cells that mediate acute illness, where the VP24 defect is predicted to impose a high cost, and immune-privileged sites behind the blood-testis and blood-retinal barriers, where that cost is low. At the high-Δ end, the enteric epithelium has the highest IFN tone in the gut (enterocytes, Δ = 1.28, rank 5), predicting an IFN-dominated response. Filovirus infection disrupts barrier function and ion transport in human intestinal organoids [15], and gastrointestinal signs are prominent in severe BDBV disease [16]. The oesophageal apical cell, the

first layer contacted by ingested virus, ranks third (Δ = 1.39), whereas suprabasal and basal cells rank 36th and 46th. This superficial response provides a local mechanism for the sore throat or dysphagia reported in 70.6% of treated patients in the 2012 BDBV outbreak [17]; the symptom is more frequent in fatal *Ebola virus* disease [18]. The hepatocyte lineage follows the Δ prediction in the BDBV-versus-EBOV comparison (Fig. S2c, d; Supplementary Note 6), placing the liver outside the sanctuary compartments. The ordering separates low-Δ immune-privileged sites from high-Δ interfaces. At the low-Δ end, Sertoli and Leydig cells (Δ = −0.68 and −0.36), retinal pigment epithelium (Δ = −1.02) and rod photoreceptors (Δ = −1.00) sit behind the blood-testis and blood-retinal barriers. This assignment is consistent with BDBV RNA detected in semen after blood clearance [14] and with the immune-privileged status of the seminiferous epithelium [19]. *Ebola virus* persists in semen [20] and in ocular fluid during convalescence [21], and retinal pigment epithelium raises its IFN-stimulated module once infected [22]. Together, these findings support the predicted prolonged seminal shedding and posterior-segment persistence. Within the eye, this prediction is limited to the posterior segment: the surface lies at the high-Δ end, whereas the aqueous and vitreous compartments correspond to the anterior and low-Δ posterior segments, respectively. The maternal-fetal interface provides a high-Δ counterpoint: the extravillous trophoblast, the highest-scoring cell type in the atlas and outside the 17-member panel, has Δ = 1.77. The trophoblast compartment ranks sixth by median and spans ranks 1 to 79. The interface has substantial clinical significance: 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% [23], and the one detailed BDBV pregnancy ended in maternal and infant death after a live birth, with viral antigen in the syncytiotrophoblast [24]. Trophoblasts are protected by type III interferon [25], an arm the BDBV configuration leaves intact [6]. Together, these features identify the maternal-fetal interface as a high-consequence site where a high IFN-tone score and preserved type III IFN protection converge, making it the most heavily priced compartment and the most direct test of the gradient.

Limitations

The framework has several limits. Δ is a baseline IFN-tone proxy: the functional evidence comes from a small HIV-1 panel rather than filovirus infection, its advantage is borderline after family-wise correction, and the high/low groups are descriptive because Δ defines them. Sequence coverage is sparse for Bombali and Tai Forest viruses; within BDBV, full-length records resolve to only two sequences differing at residue 24, while two outbreak-specific single-position calls lack read-level and within-host validation. The importin-α axis is a transcript proxy with chance discrimination (AUC 0.47–0.50); cross-atlas support is strongest at the low-Δ end, the receptor-protein and keratinocyte comparisons remain incomplete or conditional [26], and the single tissue-level dataset [27] did not consistently reproduce the ordering. The functional phenotype was first defined in an engineered virus, and overexpression, stability differences and an unresolved position-83 simulation do not separate binding from cellular context. No matched BDBV-versus-EBOV infection experiment has yet been performed in a human cell type, and the atlas has no sex/gender dimension. The ordering is therefore a testable cell-type prediction, not evidence of absolute permissiveness, virulence or tissue-level pathology.

Conclusion

BDBV carries natively the four-position VP24 configuration that was engineered to abolish IFN-λ1, IFN-β and ISG15 antagonism. That configuration is fixed across the 2007, 2012 and 2026 outbreaks. 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, separates surface epithelia and circulating myeloid cells from the retinal and testicular compartments, and identifies the enteric epithelium and the hepatocyte lineage as the points where the importin-α account reverses the prediction. The ordering specifies the relative restriction penalty across cell types and the matched comparison that tests it.

Supplementary material

Online Resource 1 contains Figs. S1-S4, Supplementary Notes 1-6 and Supplementary Methods. Online Resource 2 contains Table S1, which gives the baseline IFN-tone score for all 154 reference cell types; Table 2 lists the extremes. All supplementary items are cited in the main text.

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.

References

1. World Health Organization (2026) 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. World Health Organization, Geneva. 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 (2026) Ebola disease caused by Bundibugyo virus, Democratic Republic of the Congo, Kenya and Nigeria. Disease Outbreak News, 6 November 2026 (DON625; data as of 3 November 2026). World Health Organization, Geneva. https://www.who.int/emergencies/disease-outbreak-news/item/2026-DON625. Accessed 8 November 2026

3. Fenton KA, Pigeaud DD, Turcinovic J, Prasad AN, Agans KN et al (2026) Pathogenesis and natural history of the Bundibugyo species of Orthoebolavirus in nonhuman primates. bioRxiv. 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 (2014) 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 16:187–200. https://doi.org/10.1016/j.chom.2014.07.008

5. Schwarz TM, Edwards MR, Diederichs A, Alinger JB, Leung DW et al (2017) VP24-karyopherin α binding affinities differ between ebolavirus species, influencing interferon inhibition and VP24 stability. J Virol 91:e01715-16. https://doi.org/10.1128/jvi.01715-16

6. Ramanathan P, Tigabu B, Santos RI, Ilinykh PA, Kuzmina N et al (2023) Ebolavirus species-specific interferon antagonism mediated by VP24. Viruses 15:1075. https://doi.org/10.3390/v15051075

7. He FB, Khan H, Huttunen M, Kolehmainen P, Melén K et al (2021) Filovirus VP24 proteins differentially regulate RIG-I and MDA5-dependent type I and III interferon promoter activation. Front Immunol 12:694105. https://doi.org/10.3389/fimmu.2021.694105

8. Khan H, Tripathi L, Kolehmainen P, Lundberg R, Altan E et al (2023) VP24 matrix proteins of eight filoviruses downregulate innate immune response by inhibiting the interferon-induced pathway. J Gen Virol 104:001888. https://doi.org/10.1099/jgv.0.001888

9. Park YC (2026) Comparative whole-ORF evolutionary analysis of the seven canonical proteins of Bundibugyo ebolavirus across the 2007, 2012 and 2026 outbreaks. Research Square. https://doi.org/10.21203/rs.3.rs-10801608/v1. Europe PMC PPR1304528. Not peer reviewed

10. Goujon C, Moncorgé O, Bauby H, Doyle T, Ward CC et al (2013) Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature 502:559–562. https://doi.org/10.1038/nature12542. Data: GEO GSE46599, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46599. Accessed 4 October 2026

11. Wood O, Braithwaite AT, Jackson-Wood MA (2026) Sorted bulk RNA-seq profiling of human peripheral blood mononuclear cell populations across diverse in vitro immune stimuli [dataset]. NCBI Gene Expression Omnibus. Accession GSE327707. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE327707. Accessed 4 October 2026

12. Moss N, Sakai C, Kaul SN, Layton ED, Bouvatte P et al (2026) Revealing differential type I and type II IFN responses in different immune cells in various human diseases by a new IFNs transcriptomic atlas [dataset]. NCBI Gene Expression Omnibus. Accession GSE306664. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE306664. Accessed 4 October 2026

13. Tay DJW, Tay HY, Chu JJH, Tan KS (2026) 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 17:1895412. https://doi.org/10.3389/fphar.2026.1895412. Data: GEO GSE342661, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE342661. Accessed 4 October 2026

14. Uhrig A, Pfäfflin F, Tober-Lau P, Corman VM, Stegemann MS et al (2026) A case of Bundibugyo virus disease treated with monoclonal antibodies and remdesivir. Nat Med 32:3600–3605. https://doi.org/10.1038/s41591-026-04663-5

15. Flores EY, Hume AJ, Olejnik J, Mithal A, D'Amico A et al (2025) Filovirus infection disrupts epithelial barrier function and ion transport in human iPSC-derived gut organoids. PLoS Pathog 21:e1013698. https://doi.org/10.1371/journal.ppat.1013698. 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

16. MacNeil A, Farnon EC, Wamala J, Okware S, Cannon DL et al (2010) Proportion of deaths and clinical features in Bundibugyo Ebola virus infection, Uganda. Emerg Infect Dis 16:1969–1972. https://doi.org/10.3201/eid1612.100627

17. Kratz T, Roddy P, Tshomba Oloma A, Jeffs B, Pou Ciruelo D et al (2015) Ebola virus disease outbreak in Isiro, Democratic Republic of the Congo, 2012: signs and symptoms, management and outcomes. PLoS One 10:e0129333. https://doi.org/10.1371/journal.pone.0129333

18. Moole H, Chitta S, Victor D, Kandula M, Moole V et al (2015) 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 5:28406. https://doi.org/10.3402/jchimp.v5.28406

19. Zhao S, Zhu W, Xue S, Han D (2014) Testicular defense systems: immune privilege and innate immunity. Cell Mol Immunol 11:428–437. https://doi.org/10.1038/cmi.2014.38

20. Deen GF, Broutet N, Xu W, Knust B, Sesay FR et al (2017) Ebola RNA persistence in semen of Ebola virus disease survivors — final report. N Engl J Med 377:1428–1437. https://doi.org/10.1056/NEJMoa1511410

21. Varkey JB, Shantha JG, Crozier I, Kraft CS, Lyon GM et al (2015) Persistence of Ebola virus in ocular fluid during convalescence. N Engl J Med 372:2423–2427. https://doi.org/10.1056/NEJMoa1500306

22. Smith JR, Todd S, Ashander LM, Charitou T, Ma Y et al (2017) Retinal pigment epithelial cells are a potential reservoir for Ebola virus in the human eye. Transl Vis Sci Technol 6:12. https://doi.org/10.1167/tvst.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

23. Kayem ND, Benson C, Aye CYL, Barker S, Tome M et al (2022) Ebola virus disease in pregnancy: a systematic review and meta-analysis. Trans R Soc Trop Med Hyg 116:509–522. https://doi.org/10.1093/trstmh/trab180

24. Muehlenbachs A, de la Rosa Vázquez O, Bausch DG, Schafer IJ, Paddock CD et al (2017) Ebola virus disease in pregnancy: clinical, histopathologic, and immunohistochemical findings. J Infect Dis 215:64–69. https://doi.org/10.1093/infdis/jiw206

25. Bayer A, Lennemann NJ, Ouyang Y, Bramley JC, Morosky S et al (2016) Type III interferons produced by human placental trophoblasts confer protection against Zika virus infection. Cell Host Microbe 19:705–712. https://doi.org/10.1016/j.chom.2016.03.008

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

27. Normandin E, Triana S, Raju SS, Lan TCT, Lagerborg K et al (2023) Natural history of Ebola virus disease in rhesus monkeys shows viral variant emergence dynamics and tissue-specific host responses. Cell Genom 3:100440. https://doi.org/10.1016/j.xgen.2023.100440. Data: GEO GSE226106, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE226106. Accessed 4 October 2026

Statements and Declarations

Funding

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

Competing Interests

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

Ethics Approval

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 NCBI GEO 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.

Author Contributions

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

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
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

Segment	Rank	Cell type	Δ
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.

- OnlineResource1Supplementary.docx