Figure S1

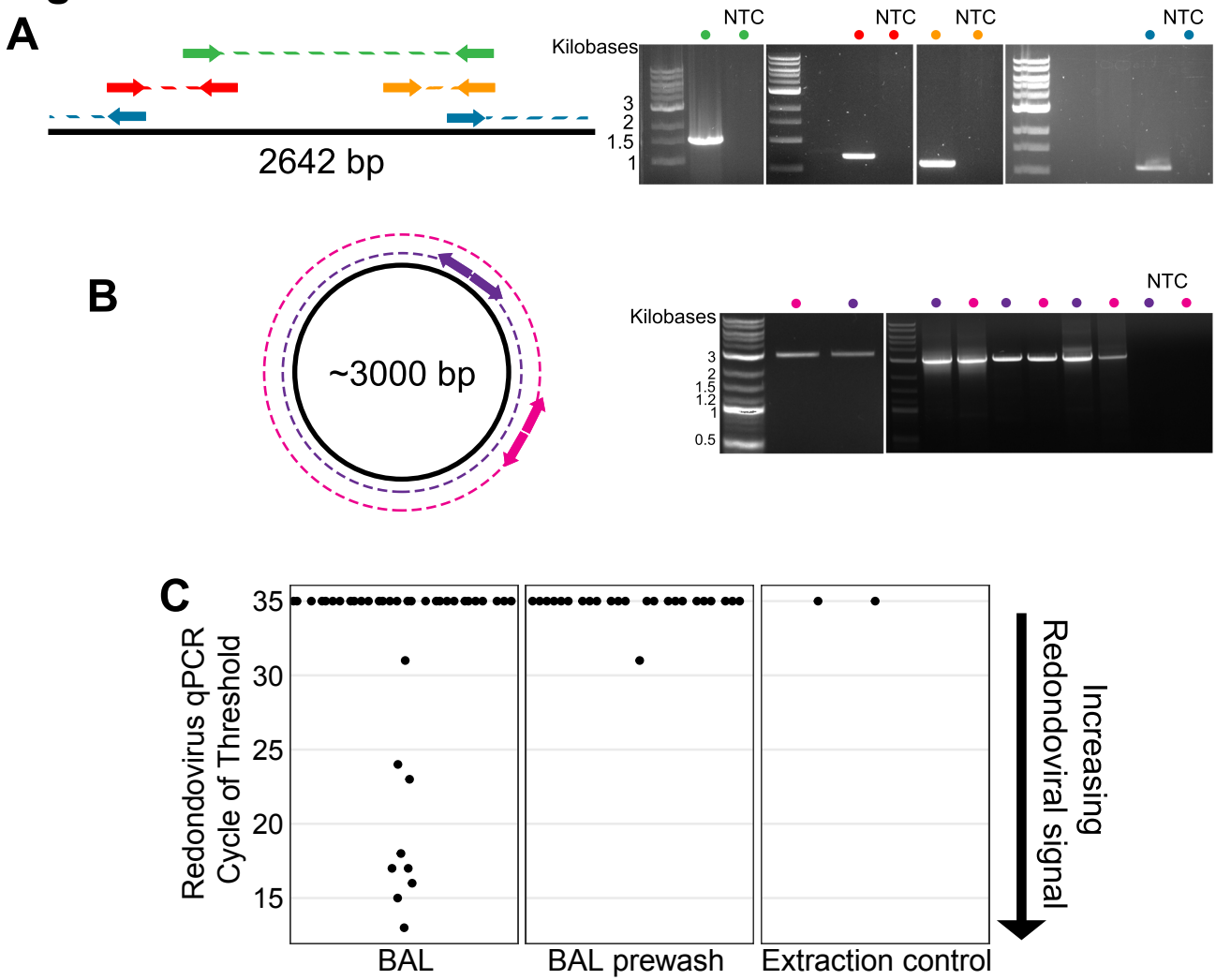

Figure S1: PCR Amplification and Detection of Redondovirus Genomes, Related to Figure 1

A-PCR amplicons of the expected size were observed from whole-genome amplified DNA from the BAL sample where Human lung-associated brisavirus RC was discovered. The outward facing primer set (blue) yielded a 600 bp product which was sequenced by the Sanger method and used to complete genome assembly. Irrelevant lanes from the gel were digitally eliminated and splicing of multiple gel images is indicated by white lines.

B-Examples of approximately 3000 base pair products of two different outward facing sets of primers are shown. These represent, from left to right, the complete genomes of Human lung-associated vientovirus FB and Human lung-associated brisavirus MD, AA and II. DNA species visualized with ethidium bromide on a 1% agarose gel are shown.

C-qPCR, performed in triplicate, was used to detect redondovirus sequences in acellular human bronchoalveolar lavage (BAL) samples after multiple displacement amplification. The average Ct value of replicates with any detection of redondoviruses is plotted on the y-axis. Samples with undetermined (i.e: no amplification signal) values in all 3 replicates are plotted at an arbitrarily high value of 35. The cycle of threshold value of the limit of quantification of the assay was 31, corresponding to 75 target copies per reaction. Samples falling below this value were counted as authentic detections. Sample types surveyed included BAL from organ donors, lung transplant recipients, and individuals with various lung diseases. BAL prewash samples represent the saline solution passed through the broncho-scope before insertion into patient. The BAL prewash point near the limit of detection represents a single replicate, with the other two replicates yielding values below the limit of detection. Extraction controls represent sterile water processed through DNA extraction kits. NTC; no template control

Figure S2

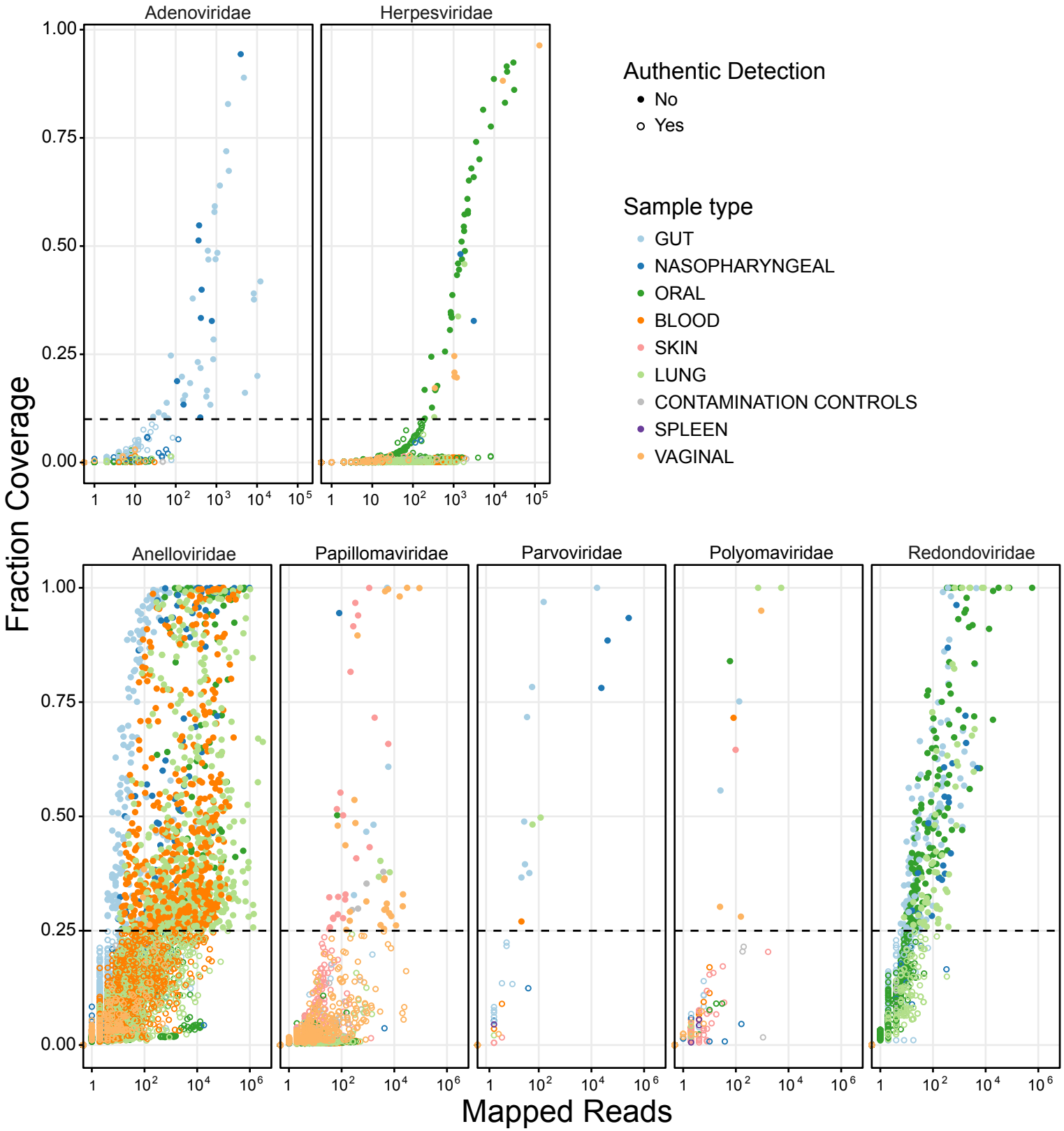

Figure S2: Summary of Read Alignments to Novel and Known Human DNA Viruses, Related to Figure 4
Reads from 2,675 human samples were aligned to 20 Redondovirus genomes and 133 human DNA viral genomes from 6 families from the NCBI RefSeq database. The log10 number of mapped reads (x-axis) and the fraction of target genome covered (y-axis) for each sample are plotted. Sample types are colored based on human body site or control. Each panel represents detection of a separate viral family. The dotted line represents the empirically determined threshold, based on examining depth and breadth of aligning reads across viral genomes, for calling a positive detection of a given viral family.

Figure S3

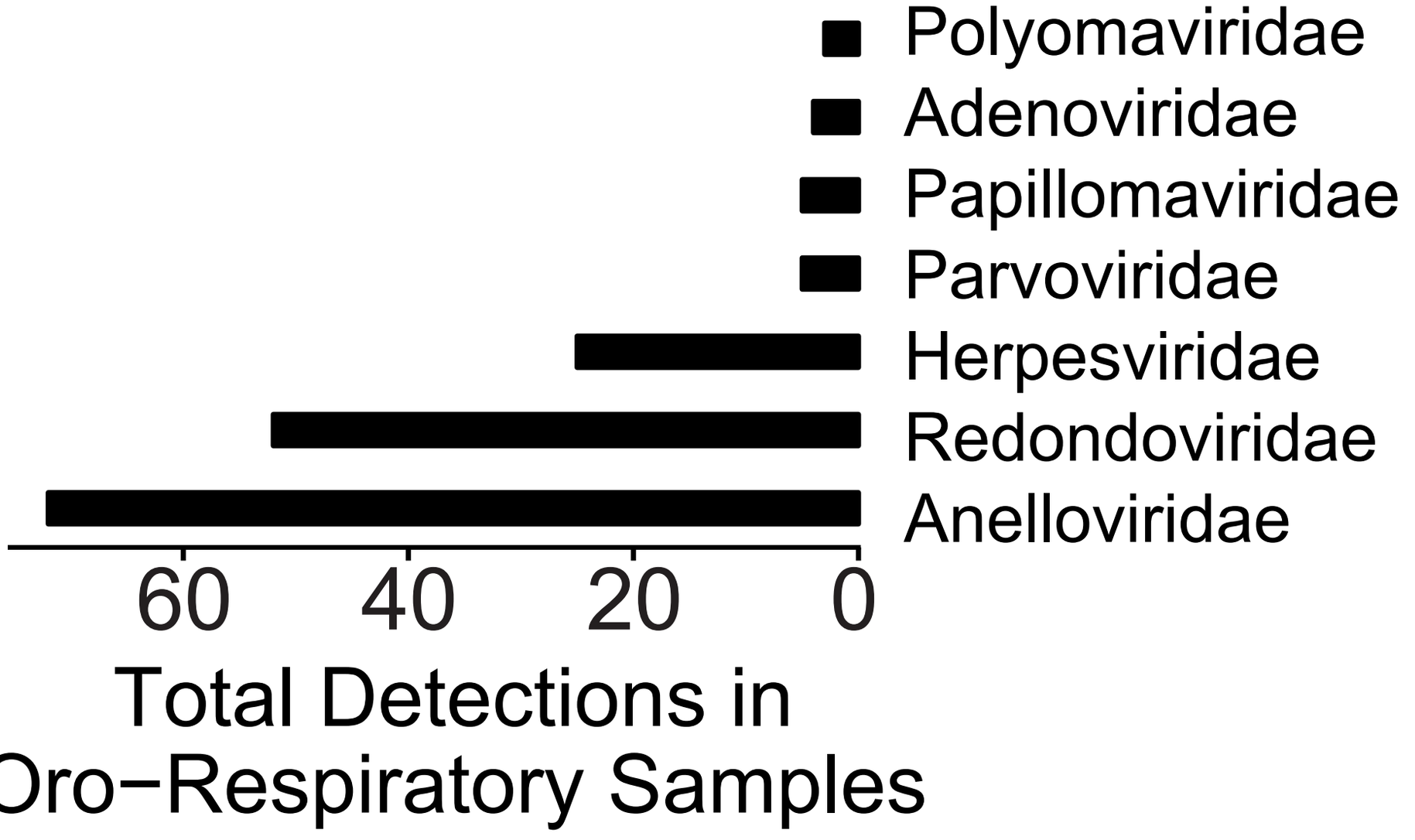

Figure S3: Prevalence of Human DNA Virus Families in Oro-Respiratory Samples, Related to Figure 4

Reads from 13 metagenomic datasets encompassing 916 oro-respiratory sample types (lung, oral, or nasopharyngeal) were aligned to 20 redondovirus genomes and 133 animal-cell DNA viruses from six viral families. A positive hit was determined as described previously. The length of each bar represents the total number of samples in which members of each viral family were detected.