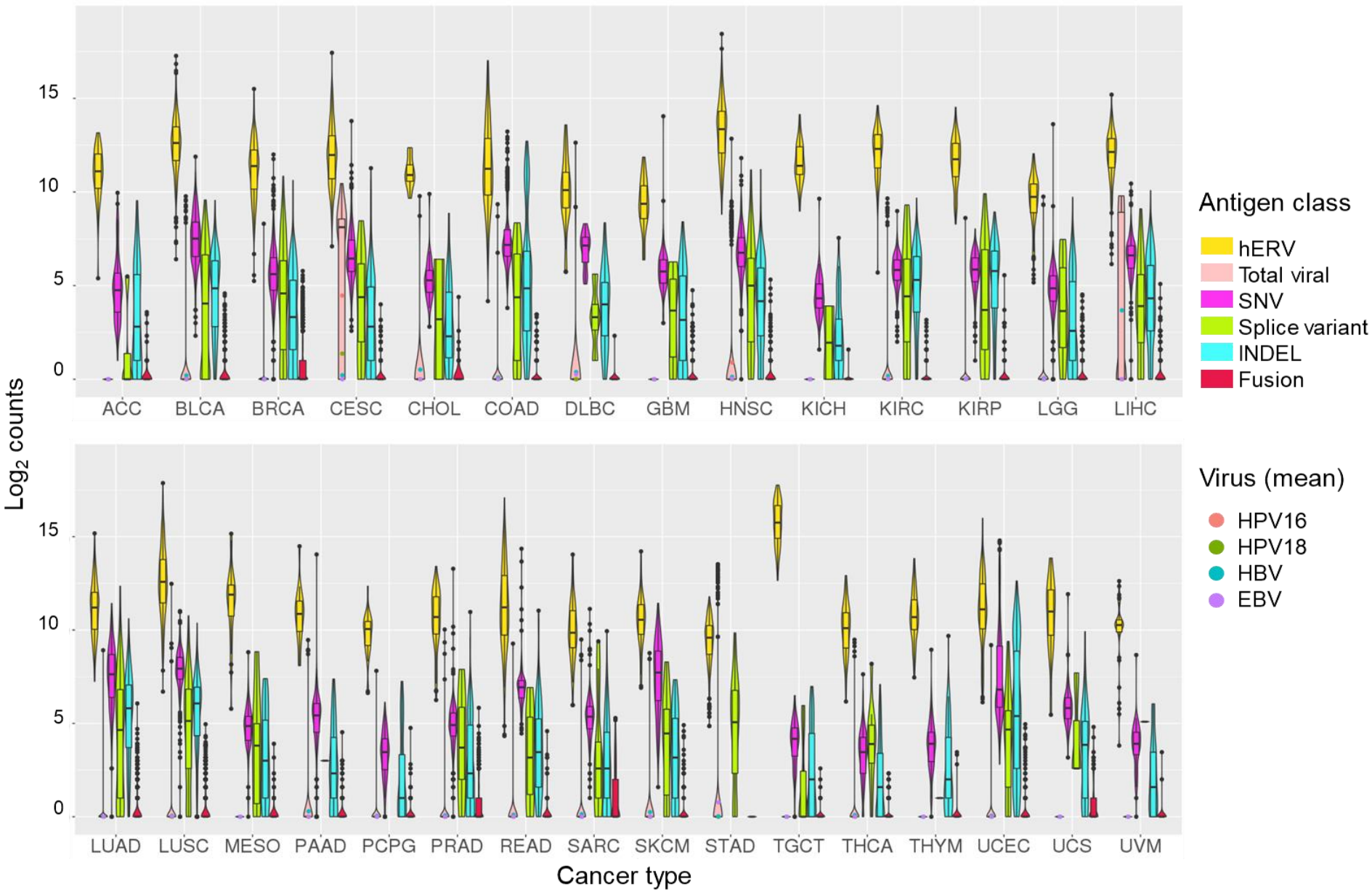
Log2 counts
15
10
5
0
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
GBM
HNSC
KICH
KIRC
KIRP
LGG
LIHC
LUAD
LUSC
MESO
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM
Cancer type
Antigen class
hERV
Total viral
SNV
Splice variant
INDEL
Fusion
Virus (mean)
HPV16
HPV18
HBV
EBV

Supplementary Figure 1: Number of unique tumor-specific antigens by cancer type. Plots represent number of unique identified epitopes by The Cancer Genome Atlas (TCGA) cancer type. Insertion or deletion (INDEL)-neoantigen counts demonstrated significant correlation with single nucleotide variant (SNV)-neoantigens among all cancer types (coefficient: 0.81, $p < 0.0001$). Notable outliers in this correlation were kidney renal clear cell carcinoma (KIRC; commonly known as clear cell renal cell carcinoma (ccRCC)) and kidney renal papillary cell carcinoma (KIRP; commonly known as papillary RCC), where the INDEL-to-SNV ratio was significantly higher than other cancer types (ccRCC: 0.85 and papillary RCC: 0.90; all others: 0.43 – 0.72). Analysis of splice variant antigens demonstrated similar burden to INDEL-neoantigens, with significant correlation with INDEL- and SNV-neoantigen burden. A notable outlier is thyroid cancer (thyroid carcinoma (THCA)), where the average number of splice variant antigens per sample is higher than SNV-neoantigens. Mean burden of fusion-derived neoantigens was highest in sarcomas (sarcoma (SARC): 1.1, uterine carcinosarcoma (UCS): 0.78), with carcinoma fusion burden highest in breast (breast invasive carcinoma (BRCA); 0.70) and prostate (prostate adenocarcinoma (PRAD); 0.58) cancer. Testicular cancer (testicular germ cell tumors (TGCT)) contained substantially greater burden of human endogenous retrovirus (hERV)-derived tumor-specific antigens (TSAs) than any other TCGA cancer type. SNV and INDEL epitopes are derived from Thorsson et al. (Immunity, 2018)[12]. Fusion epitopes are derived from Gao et al. (Cell Reports, 2018)[198]. Splice variant epitopes are derived from Jayasinghe et al. (Cell Reports, 2018)[52]. Viral epitopes are derived from Selitsky et al. (mSystems, 2018)[199]. hERV epitopes are derived from differentially expressed hERVs (>10-fold tumor-vs-mean normal expression by DESeq2) in Smith et al. (JCI, 2018)[11]. All TSA classes represent the average number of predicted class I human leukocyte antigen (HLA) binders (8-11mers, < 500 nM) predicted from NetMHCPan. Stomach adenocarcinoma (STAD) INDEL and SNV calls were absent from Thorsson et al. and esophageal carcinoma (ESCA), acute myeloid leukaemia (LAML), and ovarian serous cystadenocarcinoma (OV) were not included in all original reports. Data shown represents reanalysis of the above reports, with modification of data in order to derive values comparable across TSA groups. ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; CESC, cervical and endocervical cancers; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, lymphoid neoplasm diffuse large B-cell lymphoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; READ, rectum adenocarcinoma; SKCM, skin cutaneous melanoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UVM, uveal melanoma.