Supplemental Material

Multi-site comparison of anti-HIV microbicide activity in explant assays using a novel endpoint analysis

Nicola Richardson-Harman[1∞], Carol Lackman-Smith[1]*, Patricia S. Fletcher[2], Peter A. Anton[3], James W. Bremer[4], Charlene S. Dezzutti[5], Julie Elliott[3], Jean-Charles Grivel[6], Patricia Guenthner[7], Phalguni Gupta[8], Maureen Jones[1], Nell S. Lurain[4], Leonid B. Margolis[6], Swarna Mohan[5], Deena Ratner[8], Patricia Reichelderfer[6], Paula Roberts[1], Robin J. Shattock[2], and James E. Cummins Jr.[1]

Microbicide Quality Assurance Program, Southern Research Institute, Frederick, MD[1], Centre for Infection, Department of Cellular and Molecular Medicine, St. Georges University of London, London, UK[2], Center for Prevention Research, University of California, Los Angeles, CA[3], Virology Quality Assurance Laboratory, Department of Immunology and Microbiology, Rush University Medical Center, Rush University, Chicago, IL[4], Magee-Women's Research Institute, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA[5], Program in Physical Biology, Eunice Kennedy-Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD[6], National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control, Atlanta, GA[7], and Department of Pathology, School of Medicine, University of Pittsburgh, Pittsburgh, PA[8].

[∞]Current affiliation: Alpha StatConsult LLC, Damascus, MD. nicola@alphastatconsult.com.

*Corresponding author: Carol Lackman-Smith, Southern Research Institute, 431 Aviation Way, Frederick, MD 21701; 301-694-3232 (Tel.); 301-694-7223 (Fax); lackmansmith@southernresearch.org

Contents

Supplement 1 3
Supplement 2 6
Supplement 3 8

Supplement 1

To explore the possibility of using both increases and decreases in virus concentration to determine the soft endpoint, a two-way soft endpoint (SOFT-2) metric was tested, where both significant *increases* and *decreases* in virus growth were used to determine the assay soft endpoint. SOFT-2 was calculated to determine whether a two-way endpoint would be useful to the field in capturing the virus growth of an assay. SOFT-2 was defined as the last time point where:

$$\Delta p24\ ABS(Time_k - Time_{k-1}) > \sqrt{\Delta p24 \sum ABS(Time_k - Time_{k-1})}$$

The SOFT and SOFT-2 endpoints were compared to determine which metric would be most representative of viral growth. Even though the SOFT-2 was an effective predictor of virus growth across the explant models tested ($P<0.0001$, Table S1), the application and implications of each calculation were evaluated for individual donors. The applications of the SOFT and SOFT-2 formulas were predicted to have the greatest impact on p24 measurements from tonsil explants due to the high virus growth found in this system. The SOFT and SOFT-2 endpoints were calculated and compared for two tonsil donor explant experiments. For the first donor, virus concentration continued to increase throughout the duration of the assay, resulting in both SOFT and SOFT-2 at day 12 (see Figure S1). In contrast, virus growth for the 2nd donor declined after day 9 leading to a different result being obtained by SOFT (last significant increase in virus growth at day 9) and SOFT-2 (last significant increase or decrease in virus growth at day 12) endpoint determinations. This example illustrates how an assay summary measure using SOFT-2

could result in a p24 measurement being made at a time point when the tissue is no longer able to support virus growth. As it is not always possible to characterize decreases in virus growth, especially in virus control assays, a single measure at a time when the tissue is no longer supporting growth is unlikely to be the ideal endpoint for the determination of drug efficacy. Due to the inherent biological implications of the SOFT-2 endpoint, all analyses were conducted using the SOFT endpoint.

TABLE S1. Linear discriminant analysis of the various endpoint measures. Results of the LDA where the virus endpoint measures of viral growth (pg/ml p24) in explant tissues differentiated between low (5 assays), medium (3 assays) and high (4 assays).

Endpoint measure of virus growth	R^2	F Value	Probability (F)
SOFT	0.93	57.24	<0.0001
SOFT-2	0.87	30.89	<0.0001

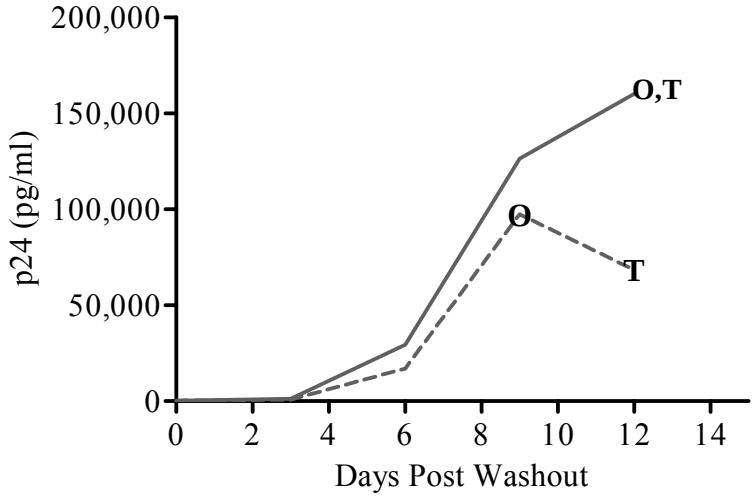


Figure S1. Comparison of SOFT and SOFT-2 analyses using "high" viral growth. SOFT (one-ways, letter "O") and SOFT-2 (two-way, letter "T") results are compared for replicate tonsil tissue donors. Two individual donors are shown where each donor is distinguished by the line pattern.

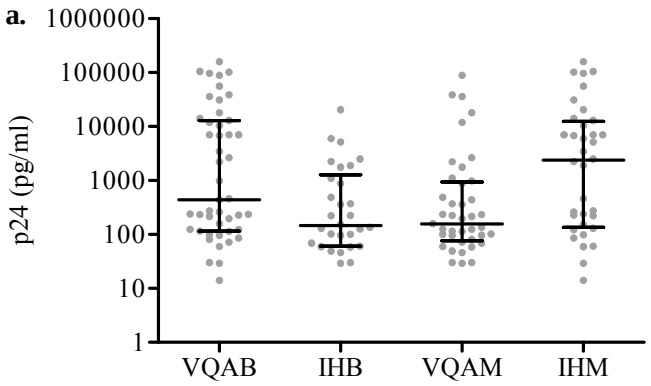
Supplement 2

All laboratories were asked to test a common medium (VQA LAB medium or VQAM) and a common virus (VQA LAB HIV-1_{Ba-L} or VQAB) alongside their usual in-house medium (IHM) and virus (IHB). Irrespective of tissue type, virus growth was found to be higher across laboratories using VQAB compared to IHB (Mann Whitney Test, P=0.0165; Figure S2a). In contrast, virus growth was found to be higher when IHM was used compared to VQAM (P=0.0081). Combining VQAB with IHM gave the highest viral growth, which was significantly greater than the IHB/VQAM condition (Kruskal Wallis, P=0.005, Dunn's multiple comparison test, P<0.01 Figure S2b).

Figure S2. Effect of source of HIV-1_{Ba-L} and media on explant virus growth. Data shown (median and inter-quartile range) represent the combined infection results for cervical, rectal and tonsil tissue explant systems infected with VQA LAB (VQAB) or in-house (IHB) HIV-1_{Ba-L}, using VQA LAB (VQAM) or in-house (IHM) media. **(a)** Comparison of p24 levels obtained for HIV-1_{Ba-L} and media sources. Table shows the probability values for the comparison between the two sources of media and virus (Mann Whitney test). **(b)** Comparison of p24 (pg/ml) levels across the four experimental conditions (Kruskal Wallis analysis of variance; P=0.005). Table shows probability values for the comparison of p24 across experimental conditions (Dunn's multiple comparison test).

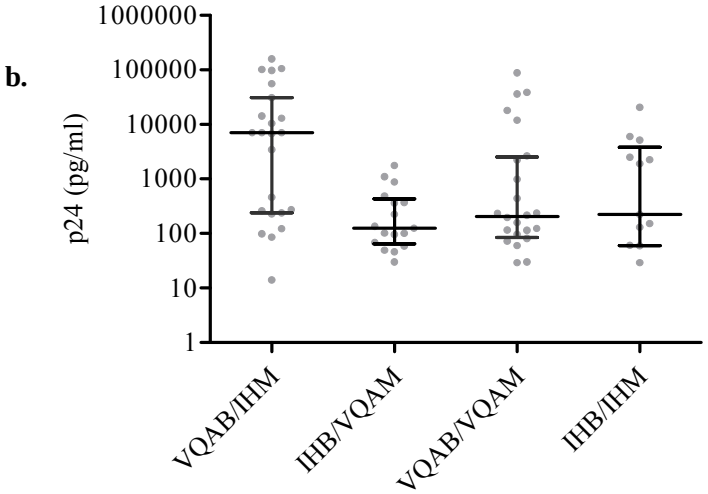
n.s. = not significant.

Figure S2.



Source of Reagents

Reagent	VQA LAB vs. In-house
HIV-$1_{Ba\text{-}L}$ (VQAB vs. IHB)	$P<0.05$
Media (VQAM vs. IHM)	$P<0.01$



Condition

Condition	VQAB/ VQAM	IHB/ IHM	IHB/ VQAM
VQAB/IHM	n.s.	n.s.	$P<0.01$
VQAB/VQAM	~	n.s.	n.s.
IHB/IHM	~	~	n.s.

Supplement 3

The effects of p24 calculation method were tested by comparing laboratory reported p24 values to values calculated using a universal curve (2nd order polynomial, non-linear curve) by Bland Altman analysis (1). The effects of p24 assay measurement method were tested by comparing p24 results for matched aliquot samples measured in-house and by the VQA LAB using the Spearman rank correlation. Intra-assay reproducibility for replicate p24 measurements was determined using the % coefficient of variation [%CV = (standard deviation/mean)*100], where a higher %CV indicates greater intra-assay variability. In-house p24 measurements corresponded to measurements made by the VQA LAB for conditions varying in sources of virus (VQAB and IHB) and source of media (VQAM and IHM; Pearson, $r = 0.82$, $P<0.0001$; Mann Whitney Test, P=n.s.; Figure S3a). Laboratory and universal standard curves provided a good fit for ELISA plate results as indicated by the proportion of variance: laboratory standard curve r^2 = 0.98-0.99 and universal standard curve r^2 = 0.95-0.99 (data not shown). A Bland Altman plot was used to illustrate the impact of the type of standard curve formula (laboratory vs. universal curve) on p24 results (Figure S3b). The percentage difference (x axis) between the p24 result reported by the laboratory and the p24 result calculated using the universal standard curve were plotted for each sample at the average of the two p24 measurements (y axis; Figure S4b). For example, if the p24 from a sample calculated using the universal curve was 50 pg/ml and the p24 from the same sample but, submitted by the laboratory was 100 pg/ml then this would be a 100% difference on the y axis at a mean p24 of 75 pg/ml on the x axis. Figure S4b illustrates the effect of standard curve formula at different p24 levels, where a higher percentage difference (y axis) indicates a greater effect of standard curve formula. The effects of standard curve formula were

minimal (y axis; majority less than a 50% difference) for p24 results above an average of 100 pg/ml (x axis; shaded area of Figure S3b). Given this minimal difference between in-house and VQA LAB p24 measurement, laboratory calculated p24 values, at the assay soft endpoint, were used in the analysis.

1. **Bland, J. M., and D. G. Altman.** 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **1:**307-10.

FIGURE S3 (a) Correlation between p24 measurements for matched supernatant aliquots measured in-house and by VQA LAB. **(b)** Bland Altman plot comparing the difference between p24 results calculated with the universal curve (UC) and those calculated in-house, with the laboratory preferred curve methodology at each sample concentration.