

# Revealing HIV viral load patterns using unsupervised machine learning and cluster summarization

Samir A. Farooq<sup>1,2</sup>, Samuel J. Weisenthal<sup>1,4</sup>, Melissa Trayhan<sup>1,2,4</sup>, Robert J. White<sup>1,2,4</sup>, Kristen Bush<sup>1,4</sup>, Peter R. Mariuz<sup>3</sup>, and Martin S. Zand<sup>1,2,4,\*</sup>

<sup>1</sup>Rochester Center for Health Informatics, 265 Crittenden Boulevard, Rochester, NY 14642-0708, USA

<sup>2</sup>Department of Medicine, Division of Nephrology, 601 Elmwood Ave - Box 675, Rochester, NY 14642, USA

<sup>3</sup>Department of Medicine, Division of Infectious Diseases, University of Rochester Medical Center/Strong Memorial Hospital AIDS Center, 601 Elmwood Ave, Rochester, NY 14642, USA

<sup>4</sup>Clinical and Translational Science Institute, University of Rochester Medical Center, 265 Crittenden Boulevard, Rochester, NY 14642-0708, USA

\*martin.zand@urmc.rochester.edu

## Supplementary Tables

**Supplementary Table S1.** Correlation coefficient matrix of features.

	Area	wRR	Adj MD	IQR
Area	1	-0.5639	-0.0155	0.5243
wRR	-0.5639	1	0.0325	-0.717
Adj MD	-0.0155	0.0325	1	-0.2379
IQR	0.5243	-0.717	-0.2379	1

**Supplementary Table S2.** Centroids and radii from polyhedral CM.

		Area	wRR	Adj MD	IQR	Radius
LT	$m_F$	1.18	1.248	0.23	0.234	
	$b_F$	0	-0.248	1	0	
LT (Centers)	DSVL	0.09	0.905	0.78	0.098	0.295
	SLVL	0.185	0.638	0.848	0.157	0.412
	SHVL	0.67	0.729	0.93	0.095	0.352
	HVLS	0.248	0.65	0.429	0.376	0.635
	RVL	0.495	0.351	0.811	0.491	0.732

$LT(\vec{F}) = m_F \circ \vec{F} + b_F$ , where  $m_F$  and  $b_F$  represent the slope and intercept, respectively, in the linear transformation of Eq ?? in terms of  $F$ .