Evaluation of the Use of Saliva Metabolome as a Surrogate of Blood Metabolome in Assessing Internal Exposures to Traffic-Related Air Pollution

Zhenjiang Li[1], Jeremy A. Sarnat[1], Ken H. Liu[2], Robert B. Hood[3], Che-Jung Chang[1], Xin Hu[2], ViLinh Tran[2], Roby Greenwald[4], Howard H. Chang[1,5], Armistead Russell[6], Tianwei Yu[7], Dean P. Jones[1,2], Donghai Liang[1,#]

[1] Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[2] Clinical Biomarkers Laboratory, Department of Medicine, School of Medicine, Emory University, Atlanta, Georgia, USA
[3] Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[4] Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, Georgia, USA
[5] Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[6] School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA
[7] School of Data Science, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China

This supporting information contains: 9 pages, 3 tables and 3 figures.

Table S1. Spearman correlation coefficients among the log-transformed intensities of the overlapping features detected in both plasma and saliva samples.

Technical Column	Type of correlation	Spearman correlation coefficient (ρ)				
		Min	Q1	Median	Q3	Max
HILIC	Grand *			0.73		
	Subject-specific	0.47	0.54	0.58	0.64	0.82
	Feature-specific	-0.60	-0.06	0.01	0.08	0.90
C18	Grand *			0.98		
	Subject-specific	0.76	0.87	0.89	0.90	0.92
	Feature-specific	-0.64	-0.06	0.01	0.07	0.68

Q1, quartile 1; Q3, quartile 3.

* Overall correlation had a single value calculated by first averaging the intensities of each overlapping feature across all subjects and calculated the correlation of averaged intensities of all overlapping features.

Table S2. Mean indoor and outdoor levels of Traffic-Related Air Pollution at Near Dorm and Far Dorm during the study period.

Pollutant, mean (SD)	Indoor		Outdoor	
	Near Dorm	Far Dorm	Near Dorm	Far Dorm
BC ($\mu g/m^3$)	1.06 (0.32)	0.76 (0.62)	0.91 (0.92)	1.04 (0.63)
CO (ppb)	366 (103)	245 (109)	391 (101)	246 (110)
NO (ppb)	10.1 (3.1)	8.7 (7.2)	22.8 (8.5)	17.6 (8.0)
NO_2 (ppb)	31.8 (12.2)	28.9 (0.4)	23.2 (7.8)	23.3 (1.8)
NO_x (ppb)	41.8 (15.3)	37.6 (7.3)	46.0 (16.1)	40.9 (7.8)
$PM_{2.5}$ ($\mu g/m^3$)	10.4 (3.1)	12.7 (1.1)	13.0 (0.9)	14.2 (1.7)

SD, standard deviation; BC, black carbon; CO, carbon monoxide; NO, nitric oxide; NO_2, nitrogen dioxide; NO_x, nitrogen oxide; $PM_{2.5}$, fine particulate matter.

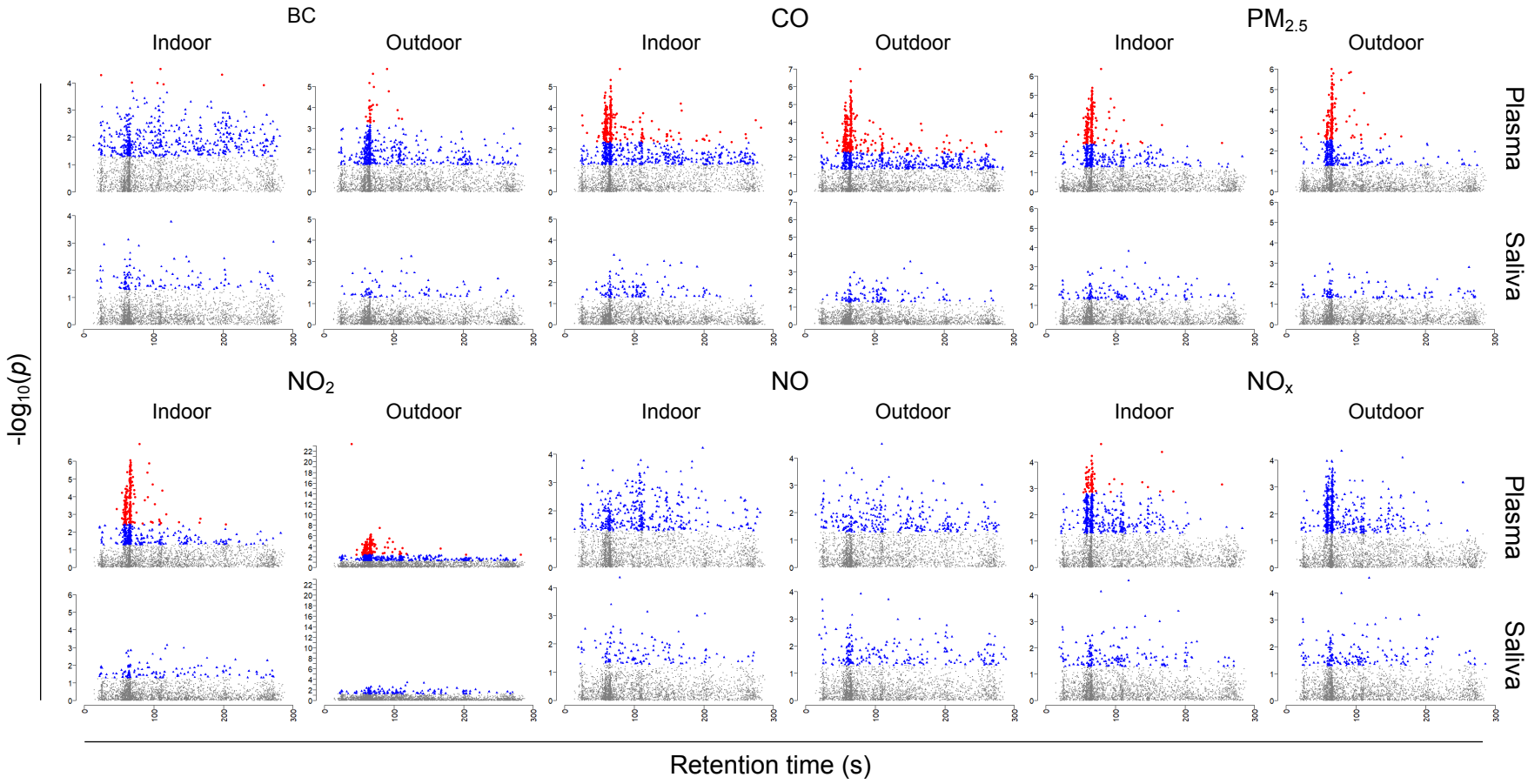

Fig S1a. Manhattan plots of associations between changes in log-transformed intensities of the overlapping features with traffic-related air pollutants using the HILIC column with positive ion mode. X-axis denotes the retention time of the metabolic features (in seconds). Y-axis denotes the negative log-10 value of the *p*-value from the association between intensity of each metabolic feature and level of traffic related air pollutant using the mixed effect model. Blue triangle denotes metabolic features significant at the alpha level of 0.05 based on raw *p*-value, and red circle denotes significant metabolic features after multiple comparison correction ($FDR_{B\text{-}H} < 0.05$).

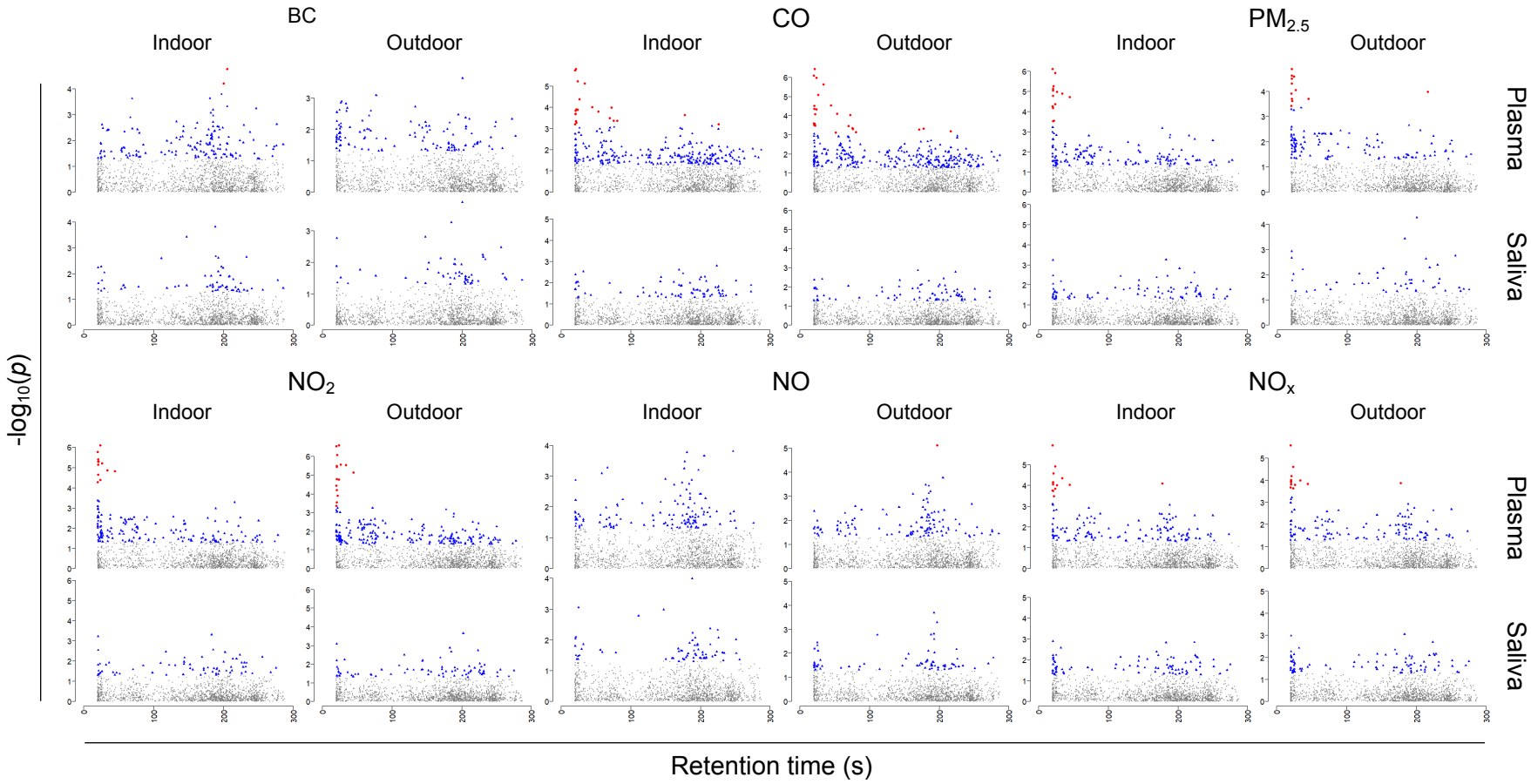

Fig S1b. Manhattan plots of associations between changes in log-transformed intensities of the overlapping features with traffic-related air pollutants using the C18 column with negative ion mode. X-axis denotes the retention time of the metabolic features (in seconds). Y-axis denotes the negative log-10 value of the *p*-value from the association between intensity of each metabolic feature and level of traffic related air pollutant using the mixed effect model. Blue triangle denotes metabolic features significant at the alpha level of 0.05 based on raw *p*-value, and red circle denotes significant metabolic features after multiple comparison correction ($FDR_{B\text{-}H} < 0.05$).

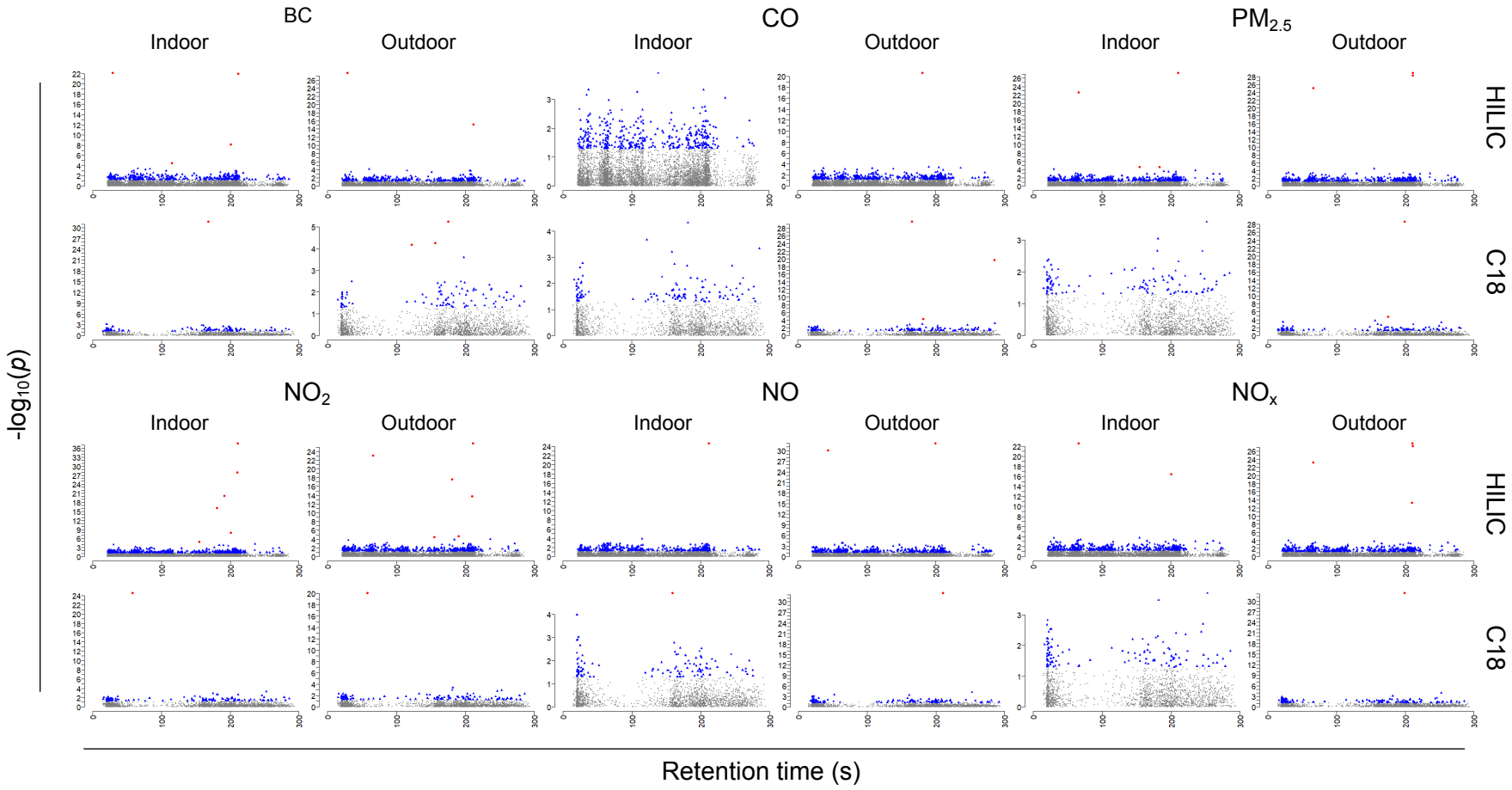

Fig S1c. Manhattan plots of associations between changes in log-transformed intensities of the unique features detected in saliva samples with traffic-related air pollutants. X-axis denotes the retention time of the metabolic features (in seconds). Y-axis denotes the negative log-10 value of the *p*-value from the association between intensity of each metabolic feature and level of traffic related air pollutant using the mixed effect model. Blue triangle denotes metabolic features significant at the alpha level of 0.05 based on raw *p*-value, and red circle denotes significant metabolic features after multiple comparison correction ($FDR_{B\text{-}H} < 0.05$).

Table S3. Number of overlapping metabolic features that were both significantly (raw $p < 0.05$) associated with traffic-related air pollutants in plasma and saliva.

Column	BC		CO		NO_2		NO		NO_x		$PM_{2.5}$	
	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
HILIC	15 (8) *	16 (9)	31 (11)	30 (9)	26 (14)	30 (12)	22 (7)	23 (8)	26 (13)	23 (10)	24 (12)	15 (9)
C18	7 (5)	4 (1)	12 (8)	13 (9)	3 (2)	6 (3)	5 (3)	8 (4)	6 (2)	6 (2)	7 (2)	1 (0)

BC, black carbon; CO, carbon monoxide; NO_2, nitrogen dioxide; NO, nitric oxide; NO_x, nitrogen oxide; $PM_{2.5}$, fine particulate matter; In, indoor; Out, outdoor.

* The number of significant overlapping features that had coefficients with the same sign was summarized in parentheses.

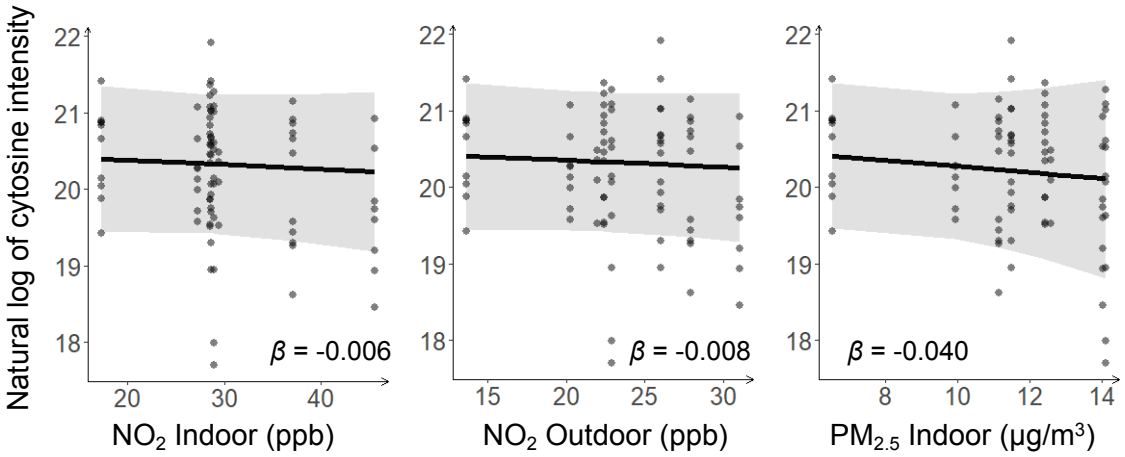

Fig S2. Chemical identity of the metabolic feature that was unique to saliva and significantly associated with TRAPs (FDR_{B-H} < 0.05) in the DRIVE study. Among the metabolic features that were unique to saliva and significantly associated with TRAPs, only one metabolic feature (*m/z* = 112.0507, retention time = 154.9s) was matched to a known chemical identity – cytosine. Acronym: NO_2, nitrogen dioxide; $PM_{2.5}$, fine particulate matter; TRAPs, traffic-related air pollutants; *m/z*, mass to charge ratio.

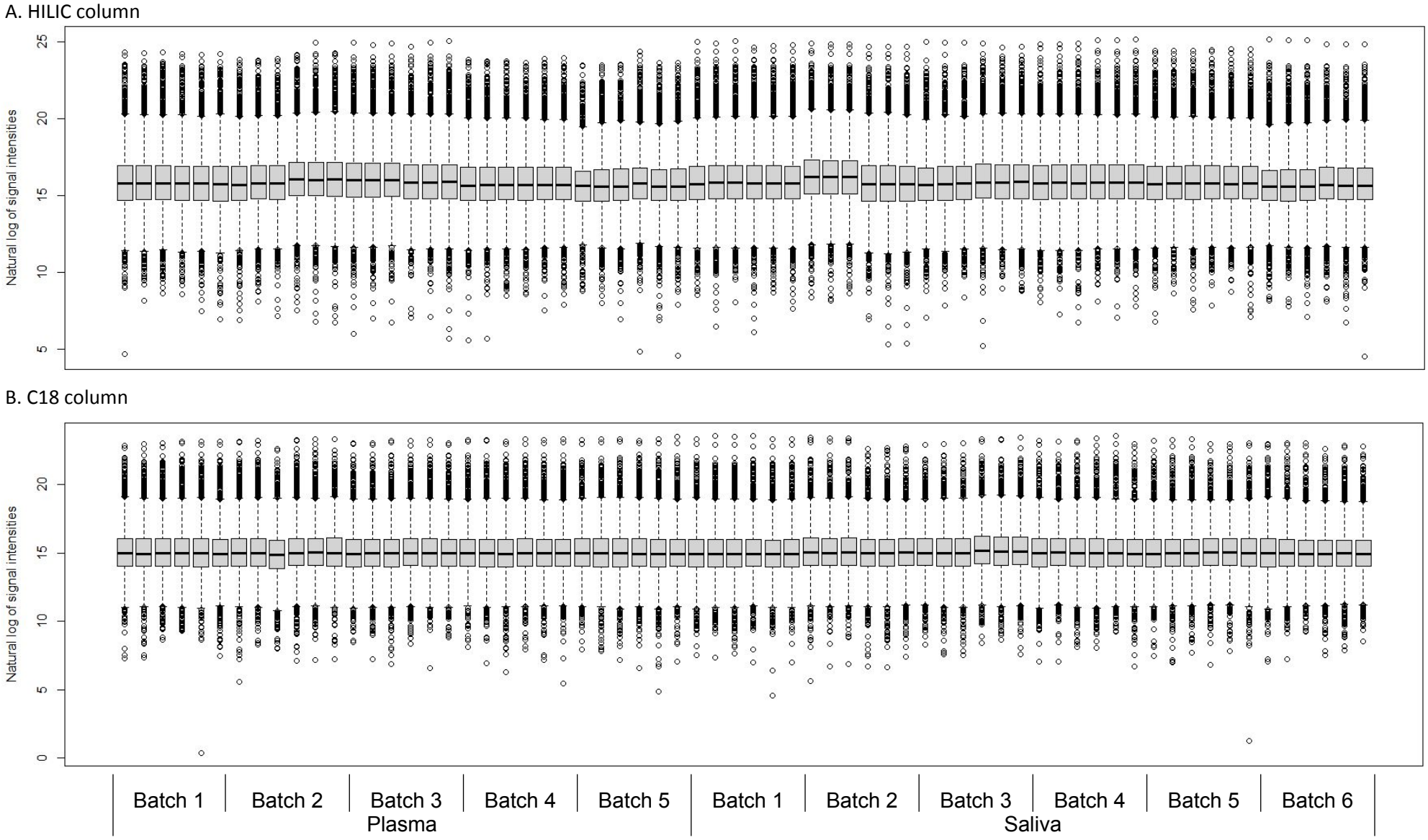

Fig S3. The distribution of feature intensities in each pooled quality control samples for all batches running for plasma and saliva samples. The feature intensities were log-transformed. Each batch contained six replicates of the pooled quality control samples. No substantial batch effect was observed after batch effect correction.