Reference Free Cell Correction Comparison in Cord Blood

Meg Jones and Rachel Edgar

22/11/2017

Table 1: Cohort Description

|	UBC data (n=24)	GenR (n=196)
Female	54.17%	
Mean Gestational Age	37.8	
FACS Cell Composition	**Mean % (sd)**	
Monocytes	6.68% (2.39)	
Granulocytes	49.35% (10.66)	
nRBCs	3.36% (2.73)	
NK Cells	4.15% (2.76)	
B Cells	6.51% (2.19)	
CD4 T Cells	25.92% (8.81)	
CD8 T Cells	4.03% (1.5)	
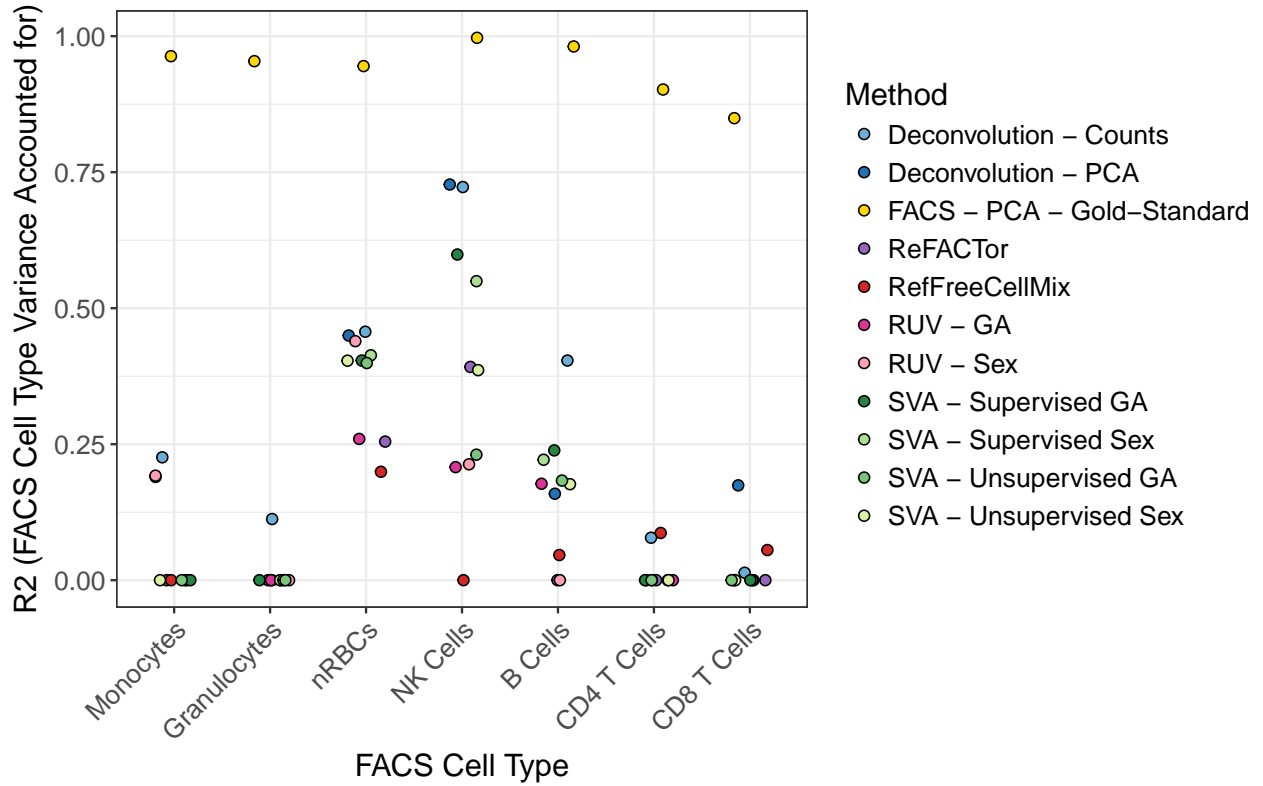
Table 2: Method name used in comparison, explanation of method, whether it is reference free, reference based or semi reference free, whether the method models with a given phenotype for EWAS, and the relevant citation for the method.

Method	Explanation	Type	Phenotype Required	Citation
FACS – PCA	Top 5 PCs from PCA on FACS counts	Reference	N	Paper in prep
FACS - Drop One Cell Type	FACS cell types as covariates in correction model with one dropped	Reference	N	Paper in prep
Deconvolution - PCA	Top 5 PCs from PCA on deconvolution predicted cell counts	Reference	N	Paper in prep
Deconvolution - Drop One Cell Type	Decouvolution predicted cell counts as covariates in correction model with one dropped	Reference	N	Paper in prep
ReFACTor	Components as model covariates	Reference free	N	PMID:27018579
RefFreeCellMix	Components as model covariates	Reference free	N	PMID: 27358049
SVA - Unsupervised GA	SVA on all CpGs with gestational age (GA) in model	Reference free	Y	PMID: 17907809
SVA - Unsupervised Sex	SVA on all CpGs with sex in model	Reference free	Y	PMID: 17907809
SVA - Supervised GA	SVA on 700 cord blood cell type differentially methylated CpGs with gestational age (GA) in model	Semi Reference Free	Y	PMID: 17907809
SVA - Supervised Sex	SVA on 700 cord blood cell type differentially methylated CpGs with sex in model	Semi Reference Free	Y	PMID: 17907809
RUV - GA	RUV with 700 cord blood cell type differentially methylated CpGs as control probes and with gestational age (GA) in model	Semi Reference Free	Y	PMID: 25990733
RUV - Sex	RUV with 700 cord blood cell type differentially methylated CpGs as control probes and with sex in model	Semi Reference Free	Y	PMID: 25990733
Uncorrected	Not corrected for cell composition	No correction	N	NA
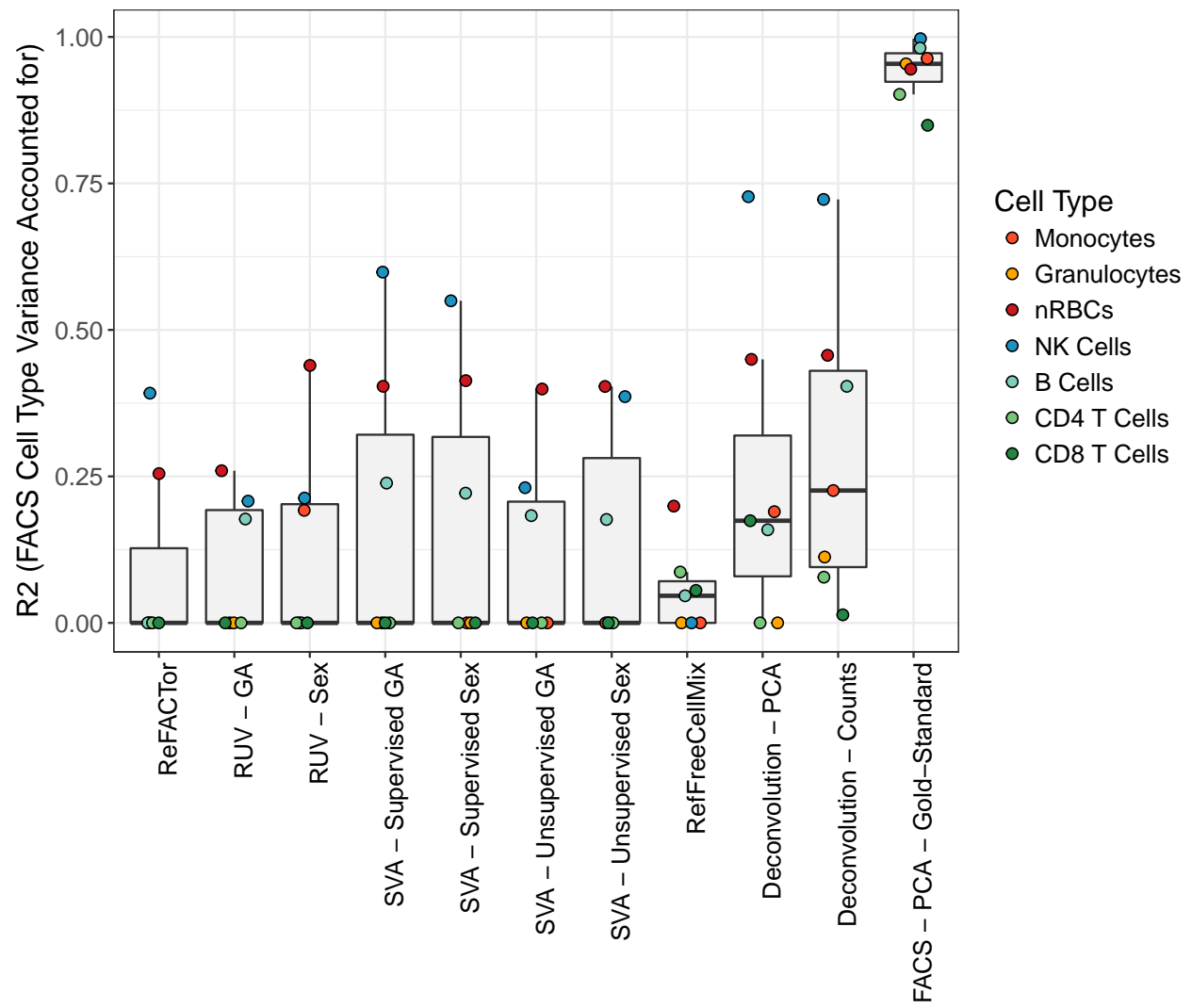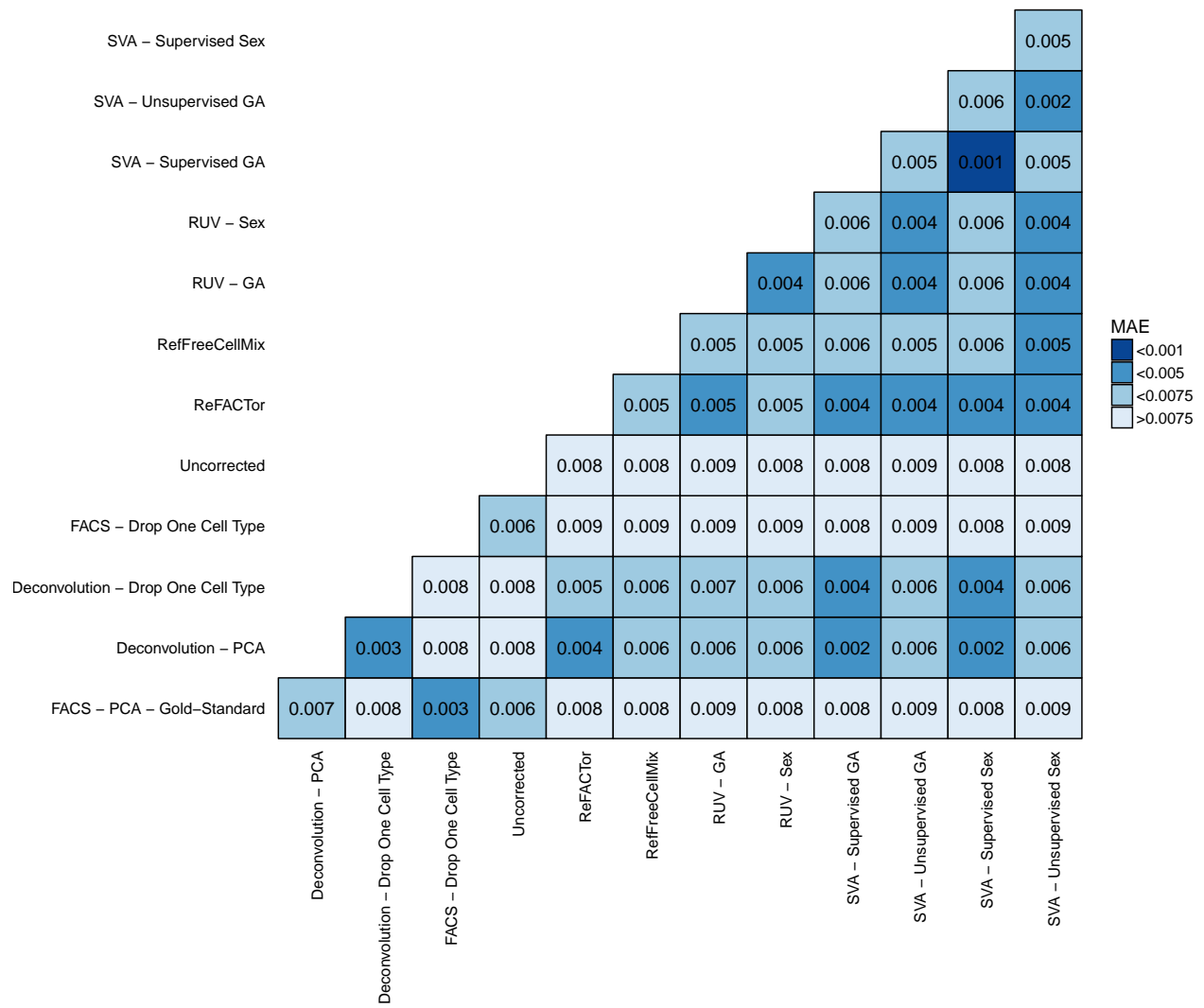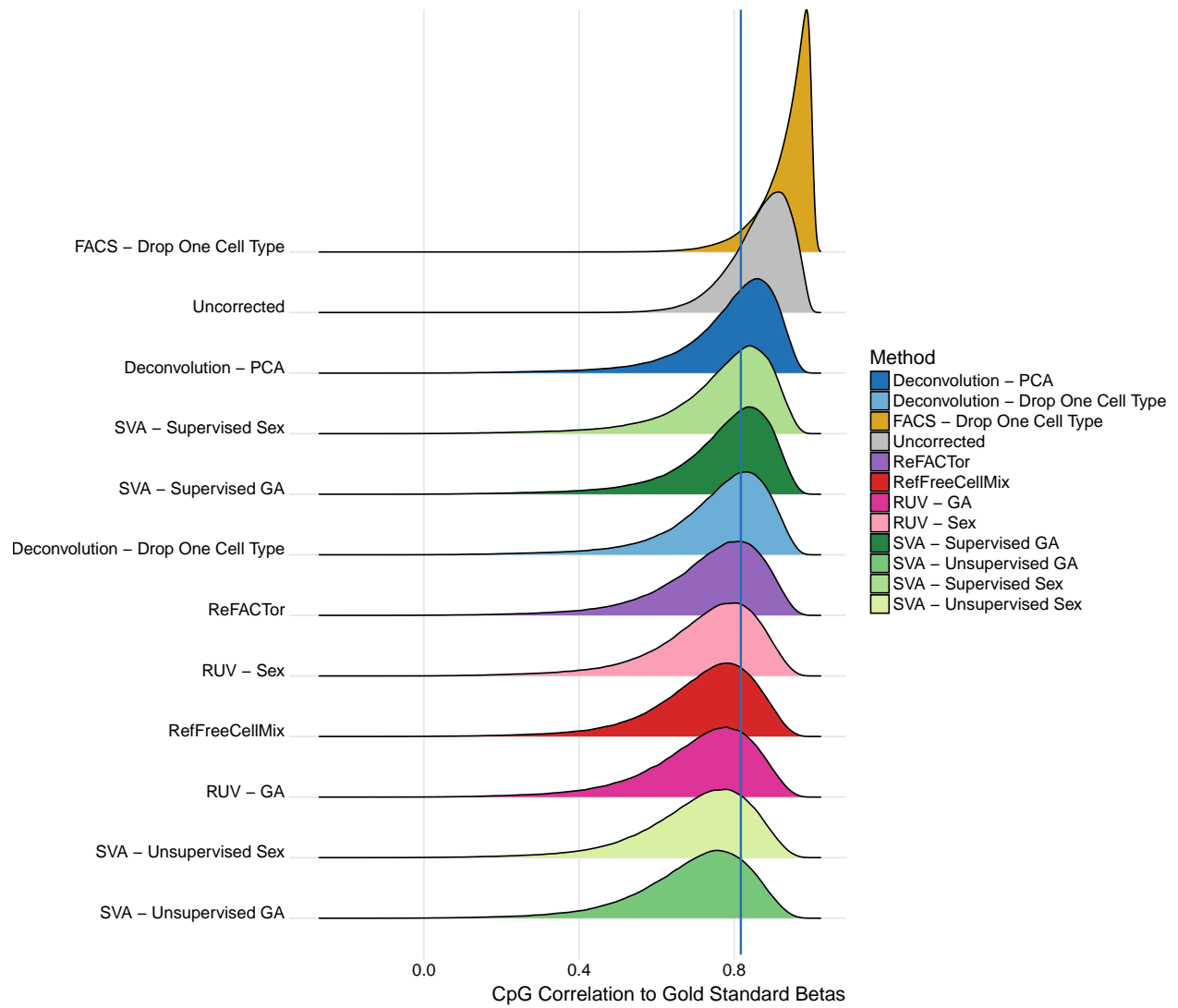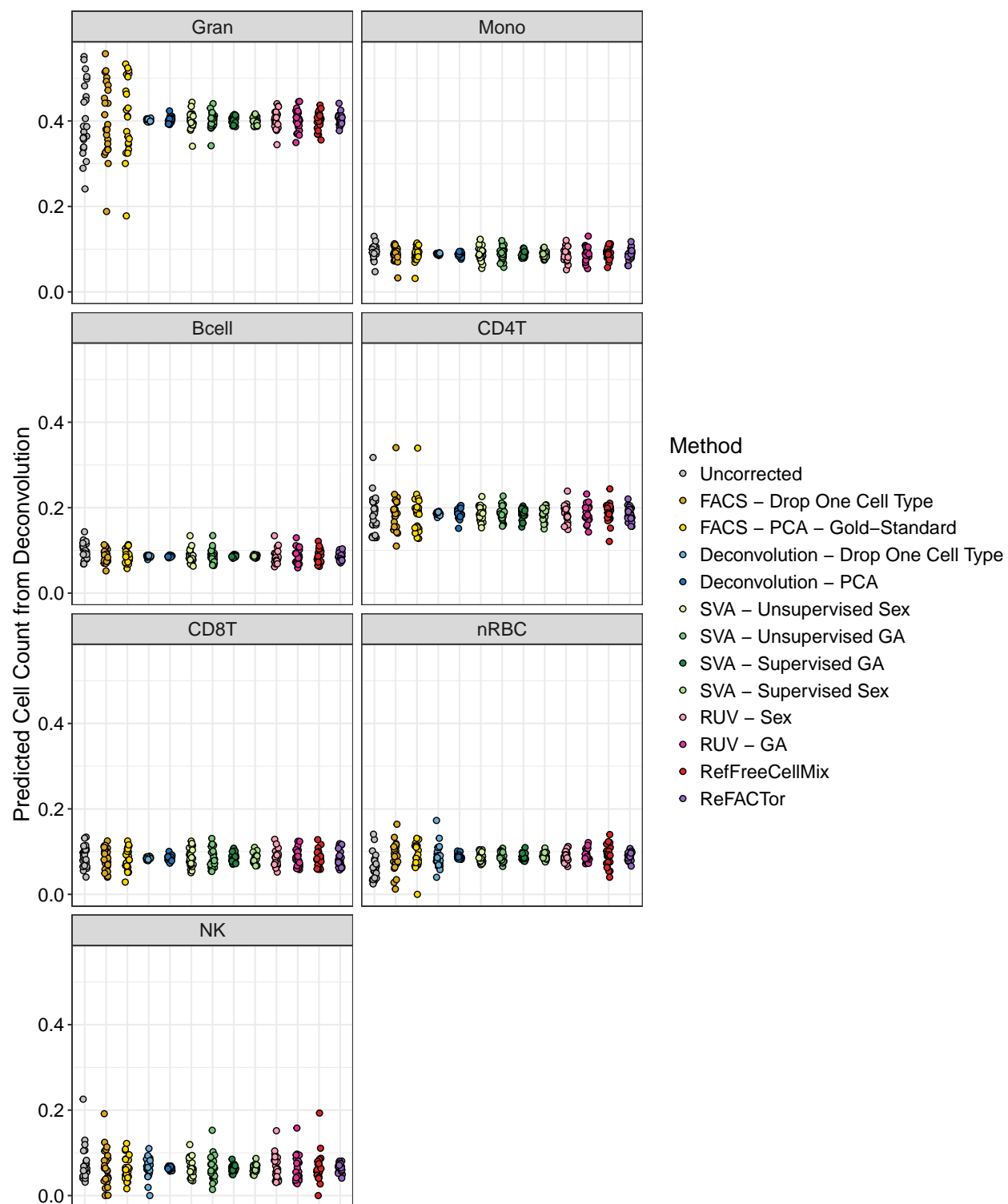
	Bcell	CD4T	CD8T	Gran	Mono	NK	nRBC
Uncorrected	0.019	0.044	0.026	0.089	0.012	0.025	0.028
ReFACTor	0.011	0.016	0.017	0.01	0.009	0.012	0.011
RefFreeCellMix	0.016	0.018	0.012	0.017	0.011	0.024	0.024
RUV – GA	0.015	0.019	0.021	0.023	0.013	0.03	0.013
RUV – Sex	0.014	0.017	0.017	0.02	0.016	0.031	0.011
SVA – Supervised GA	0.002	0.01	0.008	0.008	0.006	0.007	0.008
SVA – Supervised Sex	0.002	0.011	0.011	0.008	0.007	0.008	0.008
SVA – Unsupervised GA	0.01	0.015	0.023	0.016	0.012	0.023	0.012
SVA – Unsupervised Sex	0.011	0.019	0.019	0.015	0.013	0.021	0.008
Deconvolution – PCA	0.001	0.007	0.006	0.005	0.003	0.003	0.002
Deconvolution – Drop One Cell Type	0.002	0.002	0.002	0.003	0.002	0.021	0.021
FACS – PCA – Gold–Standard	0.013	0.045	0.028	0.098	0.013	0.026	0.027
FACS – Drop One Cell Type	0.016	0.042	0.026	0.085	0.016	0.039	0.029

Cell Count MAD
- <0.005
- <0.01
- <0.05
- >0.05

Table 3: Comparison of cell type correction methods through EWAS results. EWAS was performed for sex on each corrected dataset. True and false positives and negatives were cvalculated in comparison to the gold-standard FACS - PCA correction. Spearman is the correlation coefficient for all CpG p values, and Kendall is the rank correlation coefficient for the top 1000 CpG p values.

Method	Hits	True Positives	False Positives	False Negatives	Spearman	Kendall
FACS - PCA - Gold-Standard	5	-	-	-	-	-
FACS - Drop One Cell Type	0	0	0	5	0.636	0.42
Deconvolution - Drop One Cell Type	0	0	0	5	0.429	0.241
Deconvolution - PCA	0	0	0	5	0.458	0.264
ReFACTor	0	0	0	5	0.298	0.255
RefFreeCellMix	0	0	0	5	0.148	0.236
RUV	0	0	0	5	0.426	0.204
SVA - Supervised	4	0	4	5	0.56	0.279
SVA - Unsupervised	0	0	0	5	0.273	0.247
Uncorrected	631	4	627	1	0.595	0.283