Reference Free Cell Correction Comparison in Cord Blood

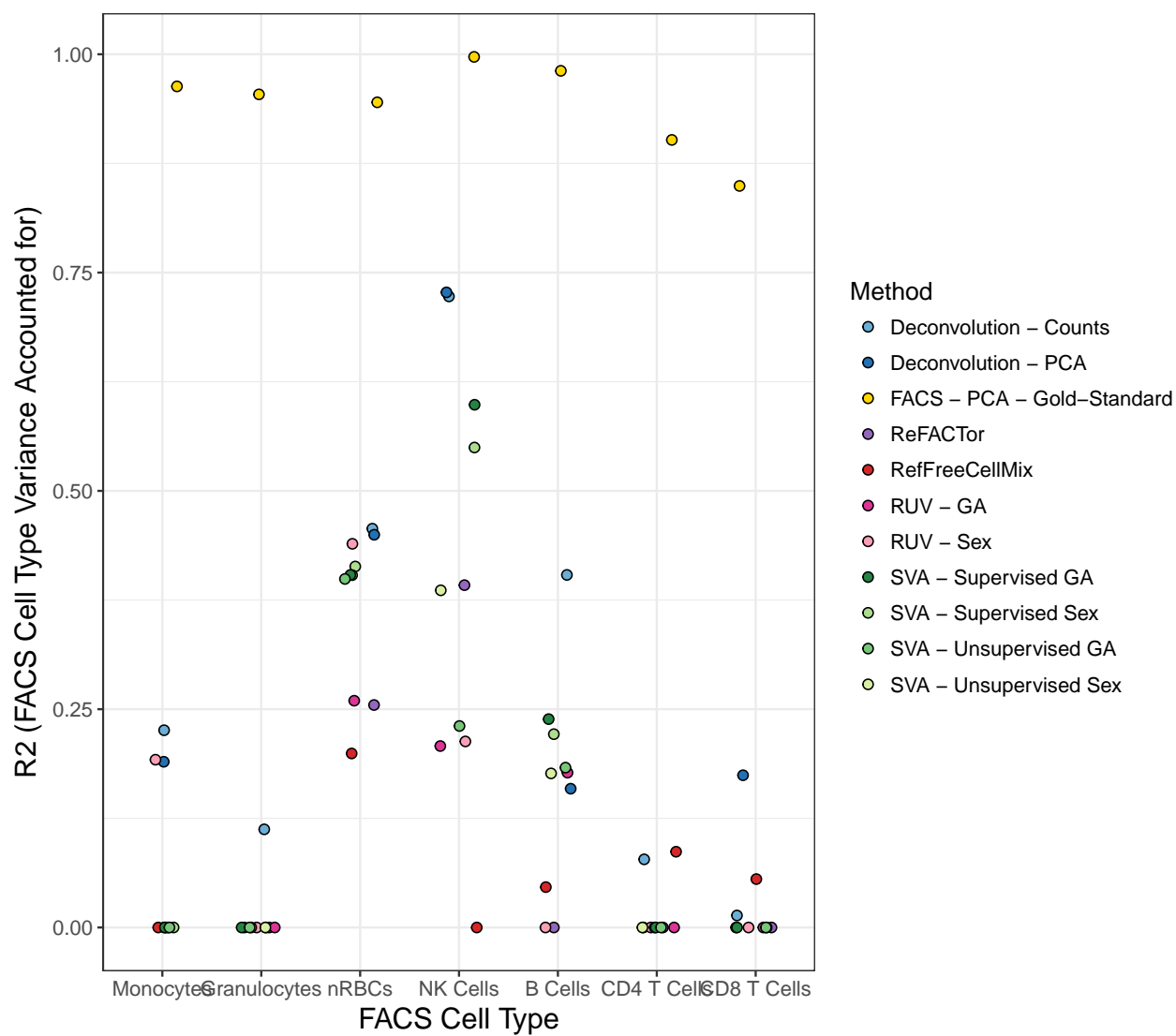*Meg Jones and Rachel Edgar*

22/11/2017

Figure 1: Variance explained in each FACS cell count by models using components from several cell type estimation methods. Models were fit used a nested-models likelihood ratio test. Components from each method were included if they significantly (p<0.05) improved model fit.
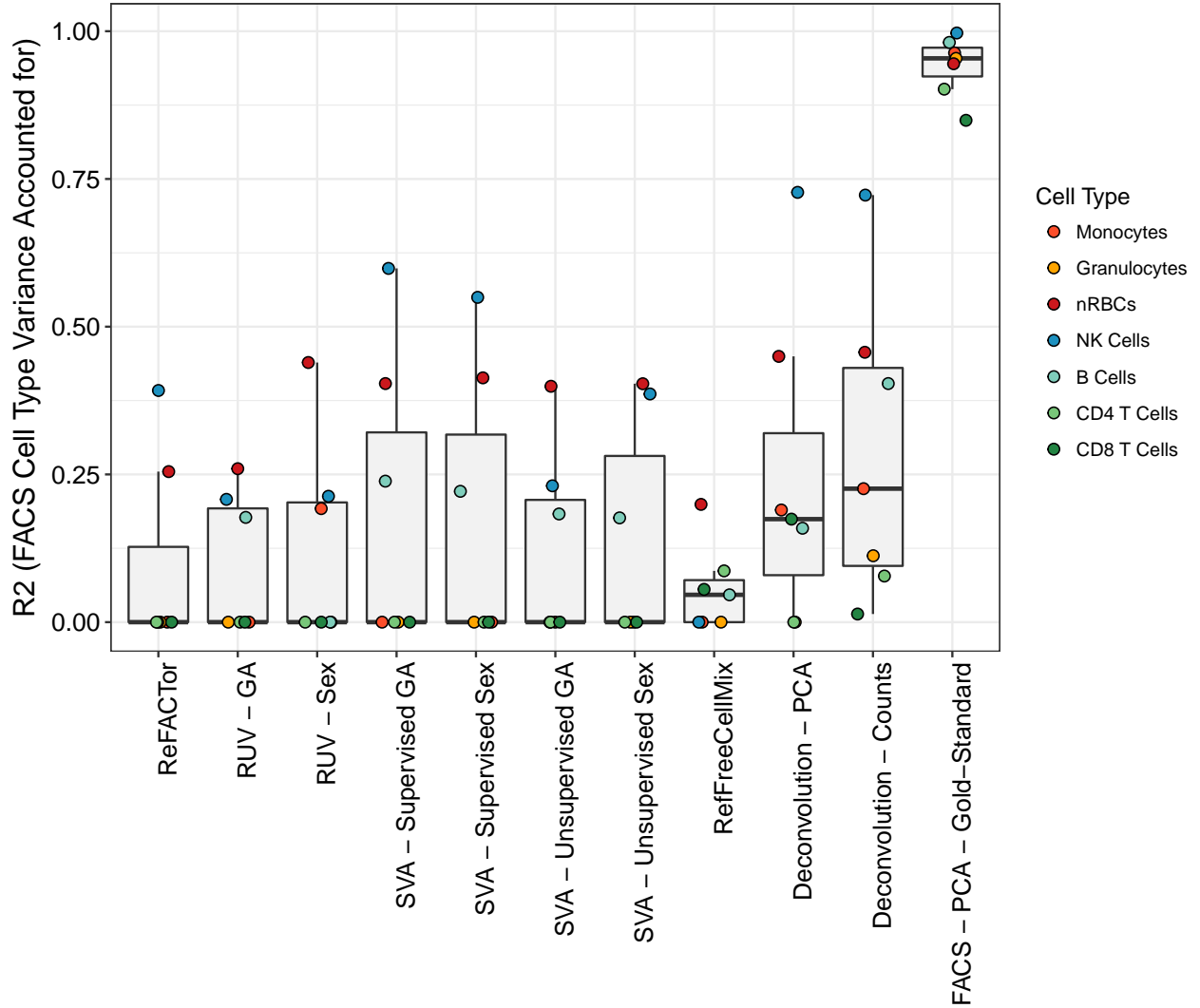
Figure 2: Variance explained in each FACS cell count by models using components from several cell type estimation methods. Models were fit used a nested-models likelihood ratio test. Models include components from each method if the component had significant likelihood ratio test p value (p<0.05). Expect for deconvolution - counts which gives measure of each cell type and R2 values are only for models with the deconvolution predicted cell type. Points represent each of 7 cell types, split by the method.

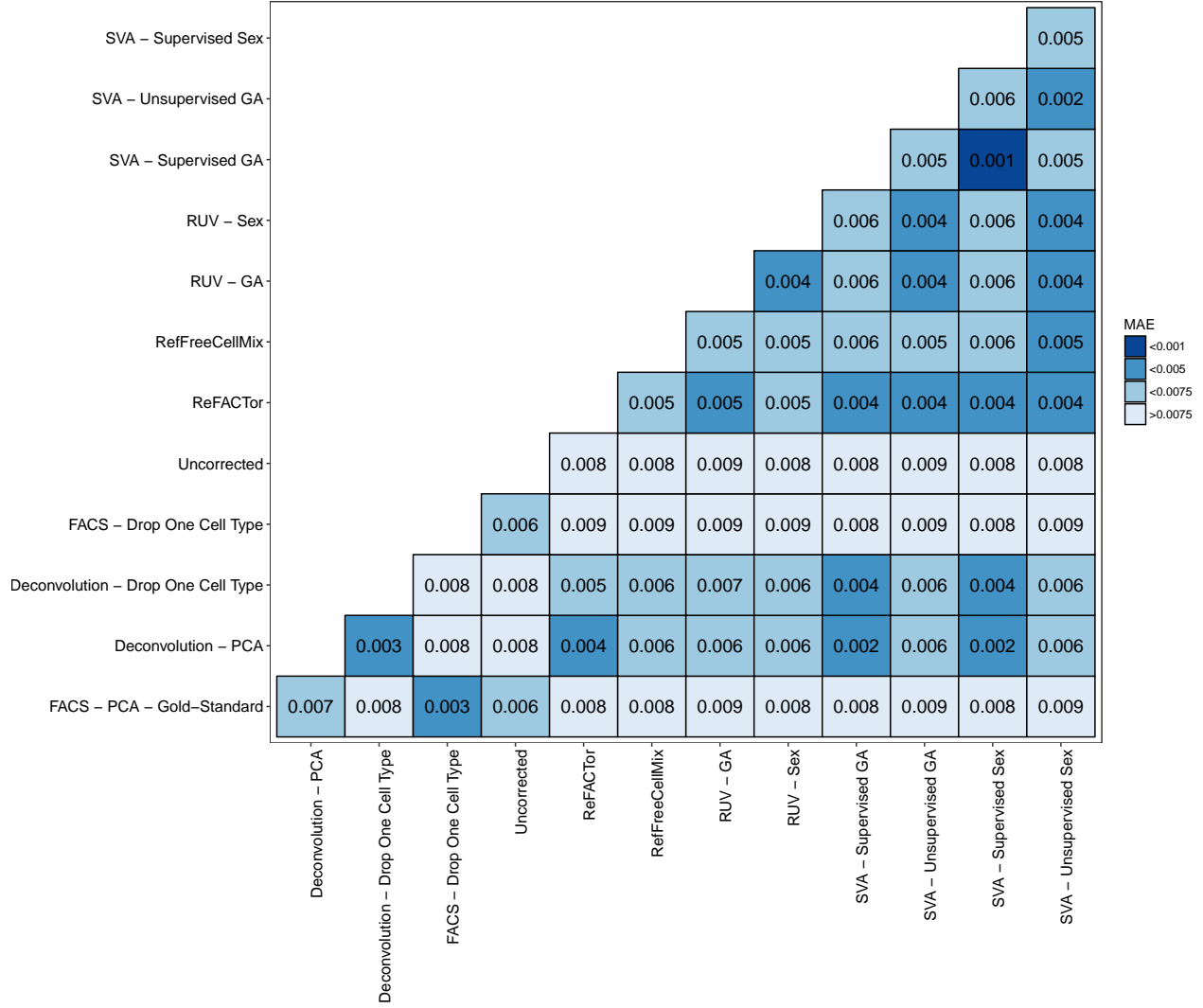|	Deconvolution – PCA	Deconvolution – Drop One Cell Type	FACS – Drop One Cell Type	Uncorrected	ReFACTor	RefFreeCellMix	RUV – GA	RUV – Sex	SVA – Supervised GA	SVA – Unsupervised GA	SVA – Supervised Sex	SVA – Unsupervised Sex
SVA – Supervised Sex												0.005
SVA – Unsupervised GA											0.006	0.002
SVA – Supervised GA										0.005	0.001	0.005
RUV – Sex									0.006	0.004	0.006	0.004
RUV – GA								0.004	0.006	0.004	0.006	0.004
RefFreeCellMix							0.005	0.005	0.006	0.005	0.006	0.005
ReFACTor						0.005	0.005	0.005	0.004	0.004	0.004	0.004
Uncorrected					0.008	0.008	0.009	0.008	0.008	0.009	0.008	0.008
FACS – Drop One Cell Type				0.006	0.009	0.009	0.009	0.009	0.008	0.009	0.008	0.009
Deconvolution – Drop One Cell Type			0.008	0.008	0.005	0.006	0.007	0.006	0.004	0.006	0.004	0.006
Deconvolution – PCA		0.003	0.008	0.008	0.004	0.006	0.006	0.006	0.002	0.006	0.002	0.006
FACS – PCA – Gold–Standard	0.007	0.008	0.003	0.006	0.008	0.008	0.009	0.008	0.008	0.009	0.008	0.009

MAE
<0.001
<0.005
<0.0075
>0.0075

Figure 3: Mean absolute error(MAE) of CpGs compared to the gold-standard FACS-PCA corrected beta values. Errors shown are an average of all CpGs MAE. Boxes are coloured by the discretized MAE value to highlight method performance.

Table 1: 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	5	0	0	1.000	1.000
FACS - Drop One Cell Type	0	0	0	5	0.636	0.420
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	0.560	0.279
SVA - Unsupervised	0	0	0	5	0.273	0.247
Uncorrected	631	4	627	1	0.595	0.283
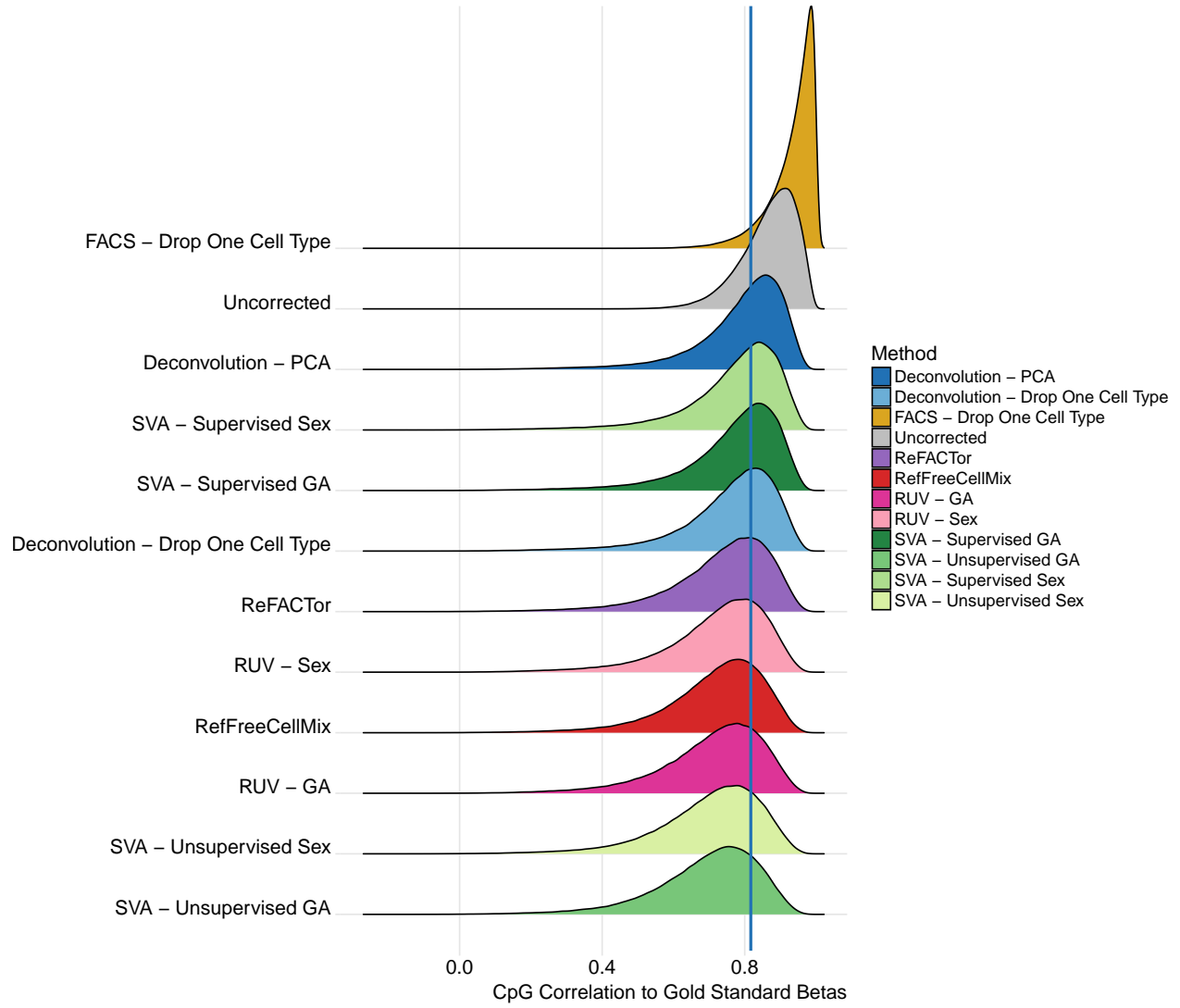
Figure 4: Distributions of correlation of beta values across samples at each CpG. Betas across samples were correlated between each method and the gold-standard FACS-PCA corrected betas. Vertical blue line shows the median correlation across all CpGs for the Deconvolution PCA corrected betas.