Supplemental Information for

funtooNorm: An improvement of the funNorm normalization method for methylation data from multiple cell or tissue types.

Kathleen Oros Klein et al.

Supplementary Methods

funtooNorm normalization method:

Let X (NxC) represent a matrix of summary control-probe data for N samples and C control probe signals, where there is a column for the average log signal from each control probe type and each colour (red, green). We create a larger matrix, X^*, by adding additional columns representing the interactions between the control probe summaries and cell type indicators. For example, if there are 3 cell types, then the matrix X^* will have $4C$ columns: the original matrix, as well as all interactions with 3 cell-type indicator variables. That is,

$$X^* = \left[X \quad X^{(1)} \quad X^{(2)} \quad \ldots \quad X^{(T)} \right]$$

where $X^{(t)}$ represents the matrix X multiplied by an indicator for cell type t, so that all that rows from samples that are *not* cell type t are zeros.

The user can then choose whether to fit principal component regressions (PCR; as in the funNorm algorithm (Fortin, et al., 2014)), or partial least squares regressions (PLS)(Tenenhaus, 1998) predicting a series of quantiles of the A and B signals from the Illumina 450 BeadChip for each sample using this augmented X^* as the covariates. As in funNorm, these models are fit separately for probe type I red, type I green, and type II. We fit models at 529 quantiles: every 0.002^{nd} percentile plus a slightly finer grid in the tails of the distributions.

The augmented covariate matrix containing interactions with cell-type or tissue-type indicators allows the relationship between quantiles and control probes to be cell- (or tissue-) type specific, hence implementing additional flexibility.

As in *funNorm*, predictions for signals A and B are obtained for all quantiles by linear interpolation between the quantile fits.

An important element of any PLS or PCR model is the number of components needed. *funtooNorm* includes a graphical display of cross-validated errors so that an appropriate number of components can be chosen (see Figure 1 and Supplemental Figure 5). All results except for Supplemental Figure 5 are based on 4 components and PCR; Supplemental Figure 5 demonstrates cross-validation results for PLS (with 4 components). The data for the 10-fold cross-validation is separately partitioned at each quantile, hence the plots are quite noisy.

Measures of agreement between replicates:

Performance was assessed by agreement between repeated measures of methylation on the same individual and in the same tissue. Let b_{ijk} be a measure of methylation for individual i, probe j and replicate k. We define M, a function of the intra-replicate squared differences, as $M = \sum_{j=1}^{J} \sum_{i=1}^{N} \left(\sum_{k_1 > k_2} \left(b_{ijk_1} - b_{ijk_2} \right)^2 / J \right)$.

In addition, we have measured performance with three metrics defined for this purpose in the *wateRmelon* Bioconductor package (Schalkwyk, et al., 2013). These metrics measure performance in the absence of replicates by making assumptions about methylation levels at known imprinted loci, at the SNP probes on the Illumina 450K array, or on the X chromosome. Performance (Supplemental Table 3) tends to be better in *funtooNorm* than *funNorm* for most situations. The method called *dasen*, described in Schalkwyk et al. 2013, is also shown in Supplemental Tables 2 and 3. Although *dasen* shows better agreement between replicates, this does not necessarily mean better ability to detect effects of interest, as discussed by (Fortin, et al., 2014). Supplemental Table 2 for the SARDs data shows that between pre-post treatment replicates there is little difference between *funNorm* and f*untooNorm*; we note that the replicate samples were taken approximately 6 months apart after immunosuppressive treatment. Hence agreement between replicates should be interpreted cautiously and may suggest generalized differences in methylation profiles after treatment.

Data sets used to evaluate performance:

Replication Data: Methylation was measured in ten healthy individuals who contributed 2-3 samples of each of whole blood, buccal swab and dried blood spots, including a mixture of technical and biological replicates. One blood spot sample was removed since the participant was pregnant. Results in Figure 1 are reported separately by tissue type, and for technical versus biological replicates taken several weeks apart. Results for these data are also reported separately by probe type, for autosomes versus the X chromosome, by positioning around CpG islands.

Systemic Auto-Immune Rheumatic Diseases (SARDs) data: In 8 controls and 44 individuals with SARDs (myositis N=4; systemic lupus erythematosus N=10; rheumatoid arthritis N=13; and scleroderma N=17), whole blood samples were fractionated and methylation profiles measured in three cell types: CD4+ T-cells, B-cells and monocytes(Hudson, et al., 2015). Repeated measurements at two times (before and after treatment with methotrexate) were available for six patients with rheumatoid arthritis and five patients with scleroderma, in CD4+ T-cells and monocytes. Results in Figure 1 compare methylation levels before and after treatment in the same individuals since similarity in serial samples is expected to be greater than between cell types (Jiang, et al., 2015). Results in Supplemental Table 3 use all samples from CD4+ T-cells and Monocytes pre-treatment.

Gestational diabetes (GD) data: A case-control study of 47 mothers was undertaken to examine the influence of gestational diabetes on methylation in fetal placenta and cord blood (Ruchat, et al., 2013). Technical replicates were available for one fetal placenta sample and for

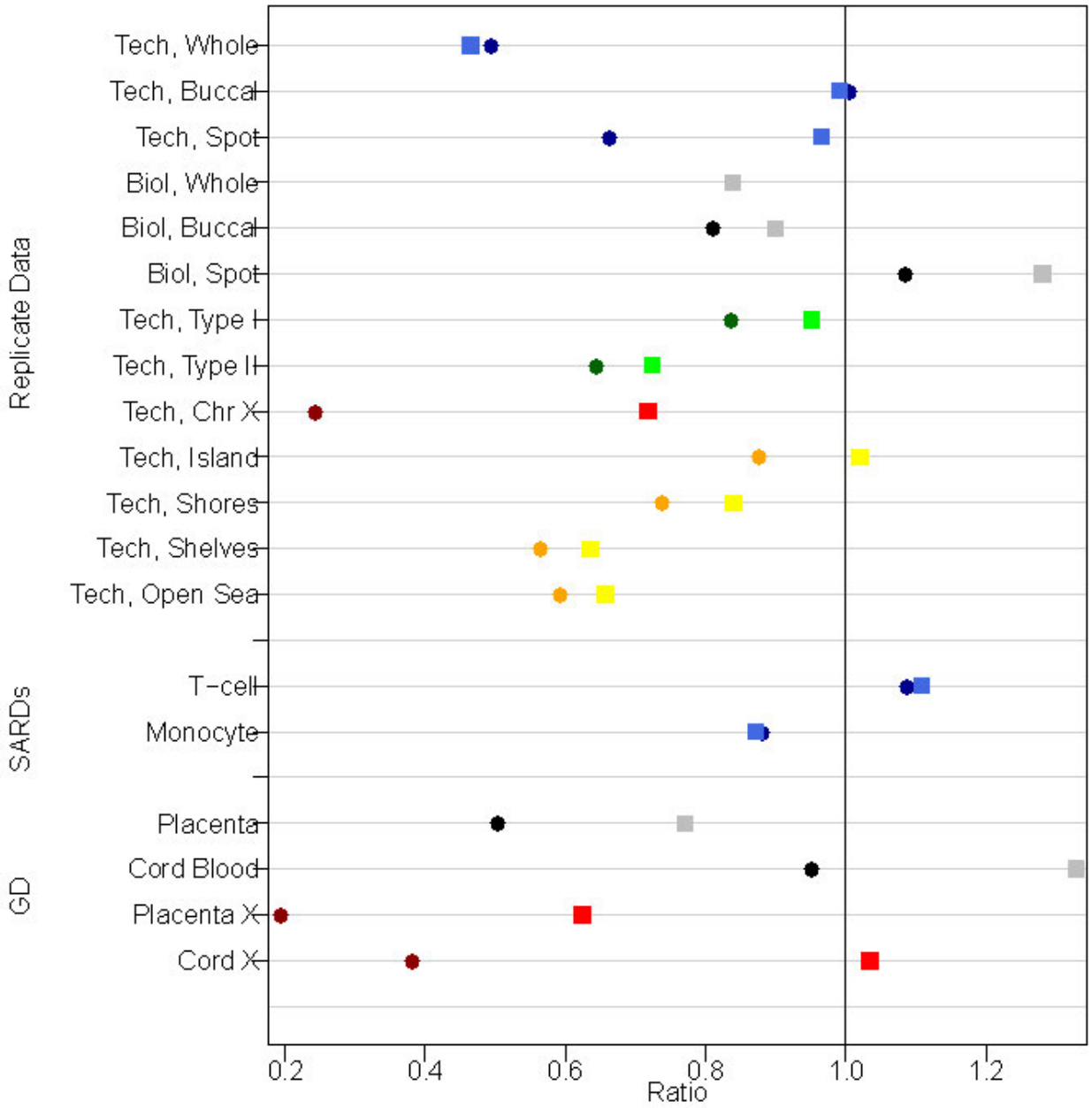
one cord blood sample and Figure 1 shows agreement for these two technical replicates. For the placenta replicate, the improvement in M was similar in magnitude to that seen in the Replication Data (Supplemental Table 2). In contrast, little improvement was observed for cord blood. However, we suggest caution in interpretation of these results due to the extremely small sample size for technical replicates. Results in Supplemental Table 3 use all available samples.

Illumina 450K (Sandoval, et al., 2011) methylation data were assessed for quality according to manufacturer recommendations and with our in-house pipeline.

REFERENCES

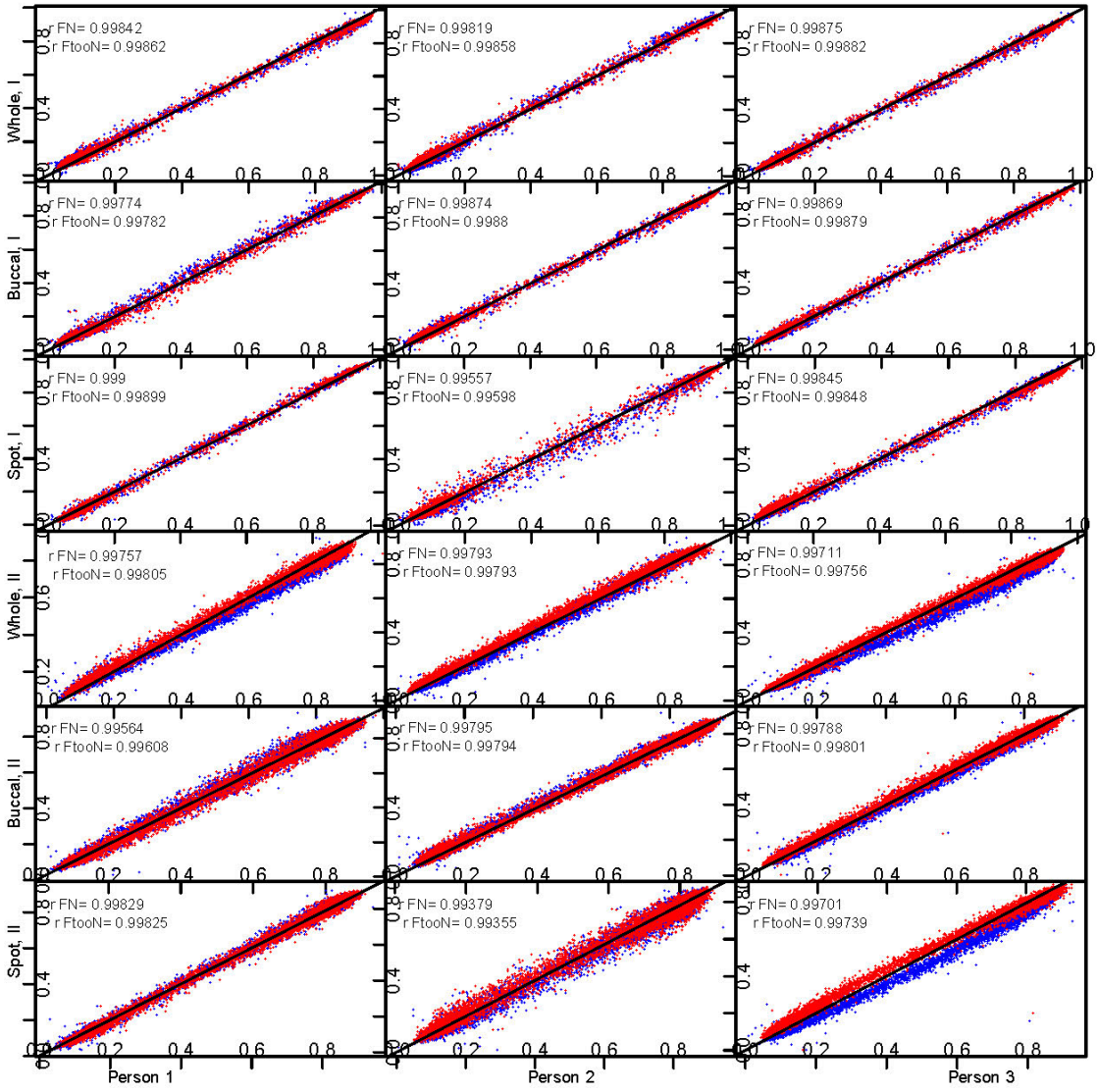
Fortin, J.-P*., et al.* (2014) Functional normalization of 450K methylation array data improves replication in large cancer studies, *Genome Biology*, **15**, 503.
Hudson, M*., et al.* (2015) Novel approaches to discovery of biomarkers in rheumatoid arthritis., *Arthritis and Rheumatology*, **accepted**.
Jiang, R*., et al.* (2015) Discordance of DNA methylation variance between two accessible human tissues, *Sci Rep*, **5**, 8257.
Ruchat, S.M*., et al.* (2013) Gestational diabetes mellitus epigenetically affects genes predominantly involved in metabolic diseases, *Epigenetics*, **8**, 935-943.
Sandoval, J*., et al.* (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome, *Epigenetics*, **6**, 692-702.
Schalkwyk, L.C*., et al.* (2013) wateRmelon: Illumina 450 methylation array normalization and metrics. *R package version 1.5.1.*
Tenenhaus, M. (1998) *La régression PLS: Théorie et Pratique*. Éditions Technip, Paris.

Supplemental Figure 1. Ratios of agreement between replicates for different cell or tissue types and for various subsets of probes in the 3 data sets. Values less than 1.0 indicate better agreement for *funtooNorm*. In each row, two ratios of agreement measures (see Supplemental Methods) are shown. $M_{funtooNorm}/M_{funNorm}$ and $M_{funtooNorm}/M_{sep}$ are shown in darker and lighter colours, respectively, where 'sep' indicates that samples from each cell/tissue type were separately normalized with *funNorm*. In addition to results for all probes combined, for the Replication Data Set, performance is also shown separately for for type I and type II probes, for chromosome X probes, and for different probe positions relative to CpG islands. "Tech" implies a technical replicate, "Biol" implies samples from the same individual were taken at two different times.

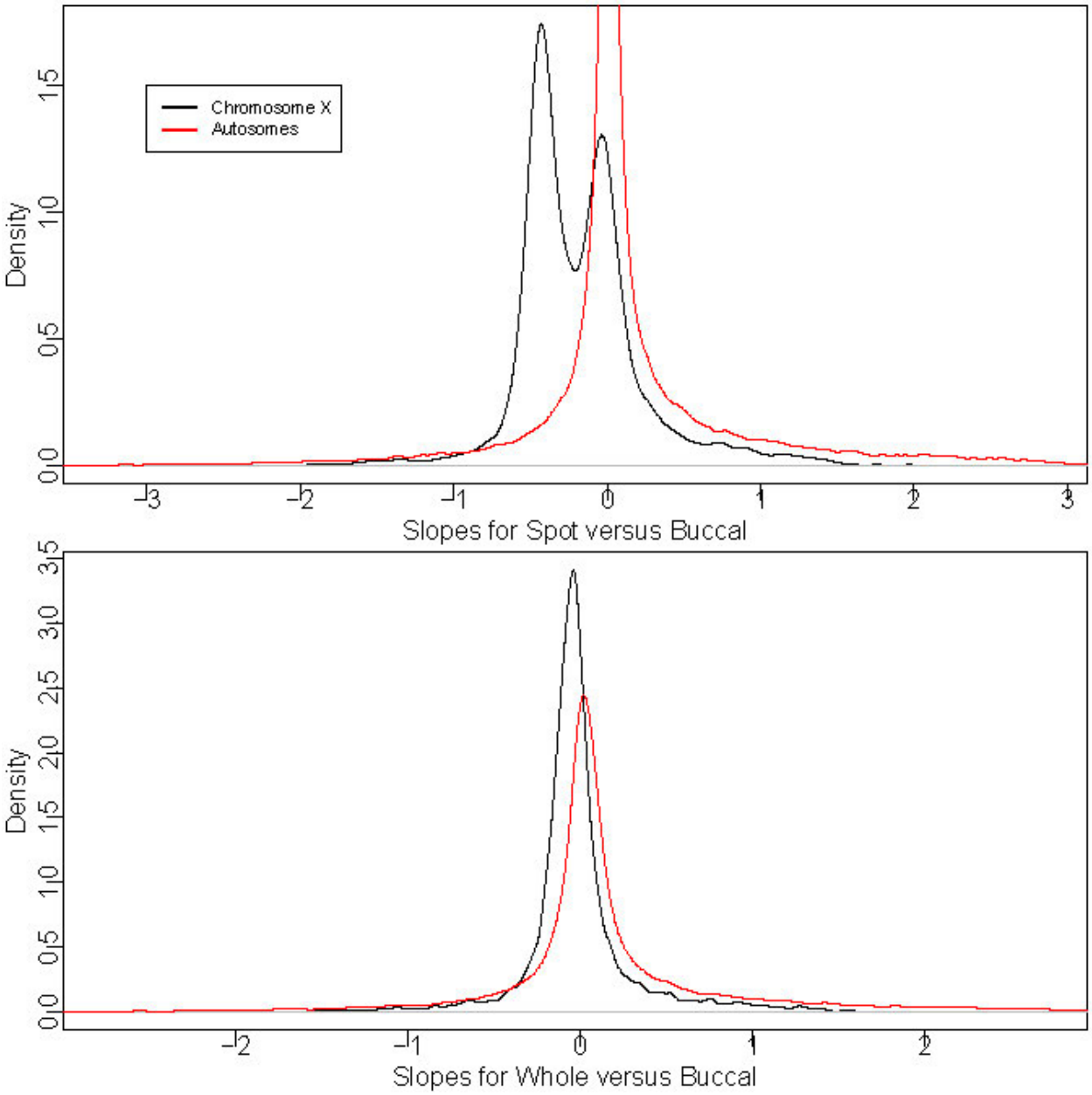


Supplemental Figure 2. Correlations between technical replicates from the Replication Data. Scatter plots of methylation measures are shown for technical replicates from the same person, together with the correlation between the two measures for each method (FN: FunNorm [blue], FtooN: FuntooNorm [red]). The three columns correspond to 3 different individuals for whom technical replicates were available for a particular tissue type. The rows show results separately by tissue type and for probe types I and II: e.g. the row label "Whole I" implies whole blood, probe type I.

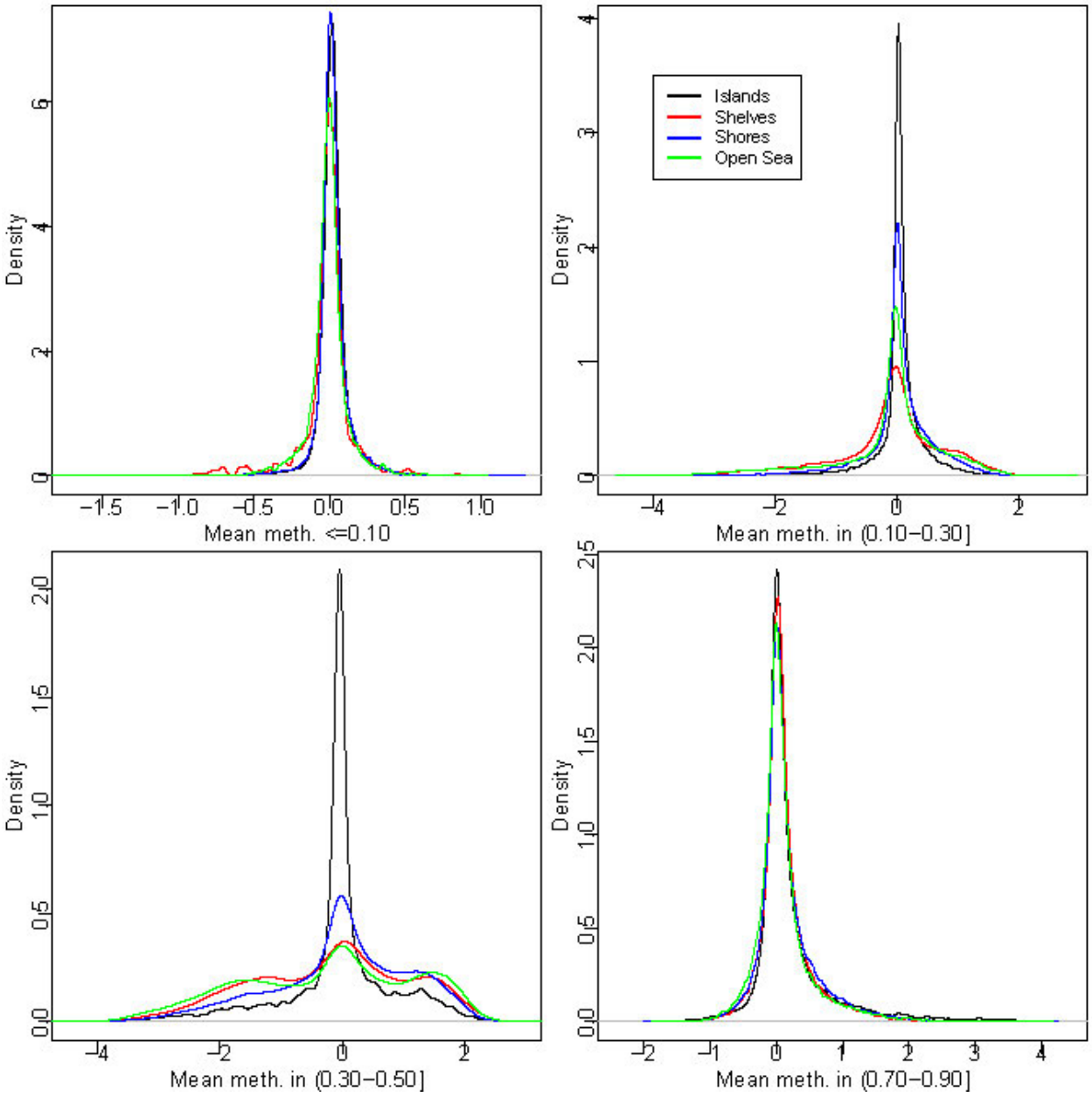
In many panels, the blue scatter is slightly wider than the red (even though red is superimposed); the benefit of *funtooNorm* is more apparent for Whole Blood and Blood Spot in probe type II.



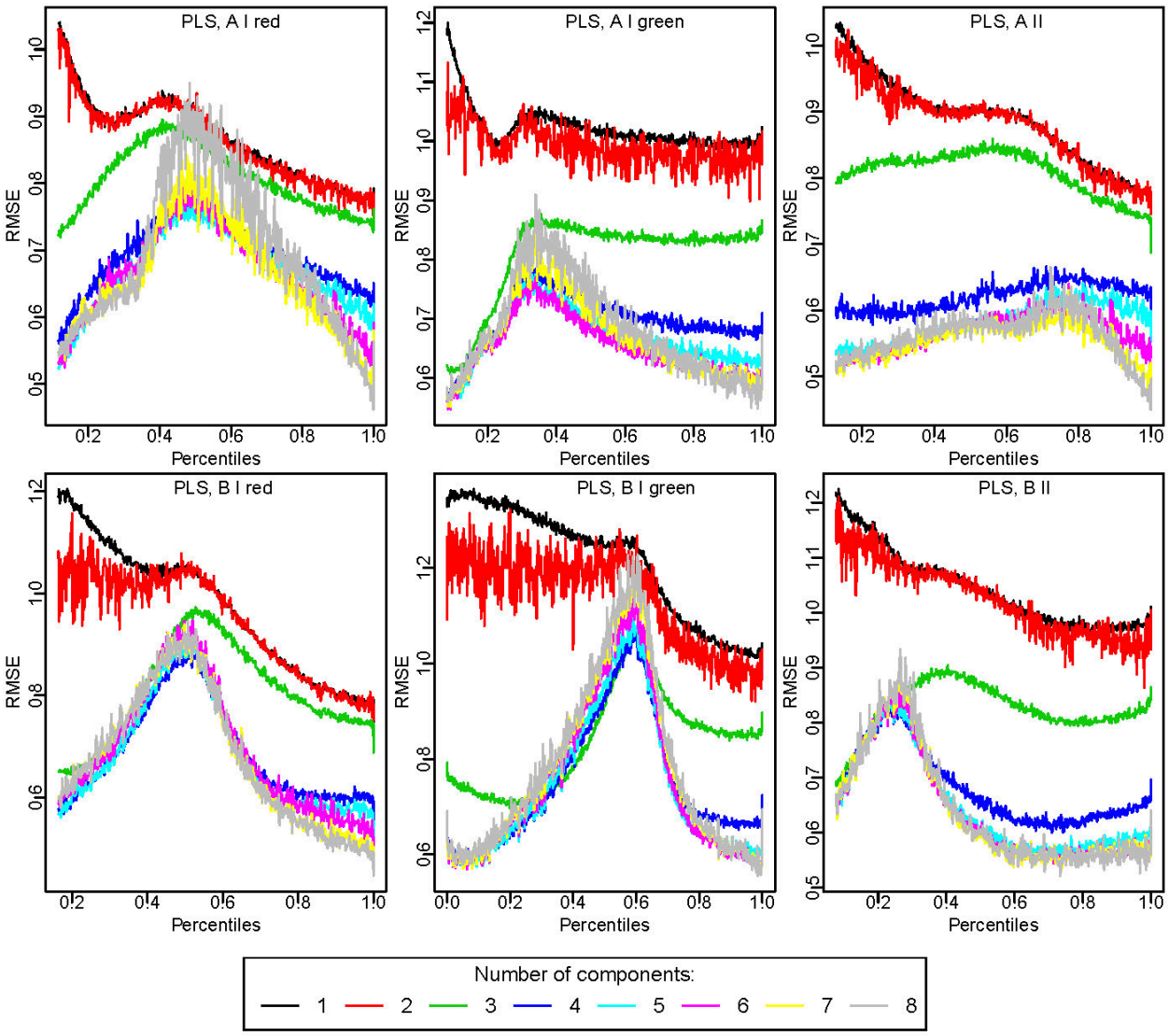
Supplemental Figure 3. Smoothed distributions of methylation differences between tissues, for probes on the X chromosome versus probes on autosomes, from the Replication Data. Slope estimates were obtained from linear models predicting methylation values as a function of tissue type, comparing Blood Spot to Buccal (top) or Whole Blood versus Buccal (bottom). The distributions of the slope estimates were separately smoothed for X chromosome and autosomal probes, demonstrating that the distribution of inter-tissue differences in methylation is quite distinct on the X chromosome. Data were normalized with funNorm.



Supplemental Figure 4. Smoothed distributions of methylation differences between tissues across probes, separately by probe annotation and mean methylation levels, from the Replication Data. Slope estimates were obtained from linear models predicting methylation values as a function of tissue type, comparing Blood Spot to Buccal. The distributions of the slope estimates were separately smoothed for probes in islands, shores, shelves or open sea, and by mean methylation level (Top left: ≤0.1; top right: (0.1-0.3]; bottom left: (0.3- 0.5]; bottom right: (0.7-0.9]. The plots demonstrate that the distributions of differences between tissues varies by annotation and mean methylation levels.



Supplemental Figure 5. Root mean square error (RMSE) from cross-validation from partial least squares regression comparing different number of components in *funtooNorm* on the Replication Data Set2. Separate model fits are implemented for the A and B signals, and for probe type I (red, green) versus II.



Supplemental Table 1. Characteristics of the three data sets used to evaluate performance

	Replication Data	**SARDs Data**	**Gestational Diabetes Data (Offspring)**
Number of samples	69	129	96
Number of individuals	10	55	47
Males/Females	1/9	18/37	25/22
Tissues or cell types (# samples)	Whole blood (23) Buccal swab (23) Blood spot (23)	T-cells (59) B-cells (7) Monocytes (59)	Cord blood (48) Fetal placenta (48)
Technical replicates	Whole blood (3 pairs) Buccal (3 pairs) Blood spot (3 pairs)	None	Fetal placenta (1 pair) Cord blood (1 pair)
Biological replicates	Whole blood (7 pairs, 3 triplets) Buccal swab (7 pairs, 3 triplets) Blood spot (7 pairs, 3 triplets)	# patients repeated pre and post treatment in same cell type Monocytes : 8 Tcells: 10	None
WateRmelon metrics used to assess performance	Seabird DMSRE GENKII	Seabird DMSRE GENKII	Seabird DMSRE GENKII

Supplemental Table 2. Agreement (M) between replicates for *funNorm* and *funtooNorm* and including also the method *dasen* (Schalkwyk et al. 2013). Smaller values indicate more agreement. *funtooNorm* results are based on PCR with 4 components.

	Cell or tissue type	***funNorm***	***funtooNorm***	***dasen***	**Ratio funtooNorm/ funNorm**
Replicate data					
Technical replicates	Whole blood	6.92e-4	3.43e-4	2.48e-4	0.495
	Buccal swab	4.12e-4	4.14e-4	3.59e-4	1.005
	Blood spot	8.74e-4	5.78e-4	4.06e-4	0.662
Biological replicates	Whole blood	0.0068	0.0057	0.0046	0.838
	Buccal swab	0.0101	0.0082	0.0069	0.812
	Blood spot	0.0078	0.0084	0.0051	1.085
SARDS data					
Biological replicates	T-cells Pre-post	0.0021	0.0023	0.0017	1.088
	Monocytes Pre-post	0.0019	0.0017	0.0014	0.881
Gestational Diabetes data					
Technical replicate	Placenta	2.17e-4	1.09e-4	0.947e-4	0.504
	Cord Blood	1.38e-4	1.31e-4	1.12e-4	0.952

Supplemental Table 3. Results of measures of performance from the *WateRmelon* Bioconductor package. For each statistic, a smaller value is better; the normalization with the best performance for each measure is given in bold font. The *dasen* method is described in Schalkwyk et al. 2013.

	Measure based on imprinted loci (DMRSE)					**X-chromosome-based measure (Seabird)**					**SNP-based measure (GENKI)***			
	funnorm	*funtoonorm*		*dasen*	Raw	*funnorm*	*funtoonorm*		*dasen*	Raw	*funtoonorm*		*dasen*	Raw
		PLS	PCR				PLS	PCR			PLS	PCR		
Replication Data (N=69)														
Whole blood	0.0033	0.0032	0.0032	0.0025	0.0038	0.3217	0.1694	0.1718	0.1418	0.1859	4.64e-5	4.13e-5	3.02e-5	4.25e-5
Buccal swab	0.0036	0.0021	0.0022	0.0022	0.0039	0.2654	0.1802	0.1867	0.1932	0.1741	5.88e-5	5.34e-5	5.01e-5	6.63e-5
Blood spot	0.0038	0.0022	0.0022	0.0019	0.0042	0.2141	0.1312	0.1312	0.1207	0.1323	7.85e-5	7.65e-5	4.30e-5	7.40e-5
All together	0.0025	0.0020	0.0020	0.0013	0.0026	0.1358	0.0919	0.0947	0.1002	0.1077	4.14e-5	3.89e-5	6.60e-5	6.88e-5
SARDs Data (N=127)														
T-cell	0.0036	0.0040	0.0040	0.0025	0.0047	0.0569	0.0741	0.0730	0.0612	0.0568	8.89e-5	9.79e-5	7.22e-5	10.19e-5
B-cell	0.0072	0.0112	0.0127	0.0073	0.0140	0.2139	0.1390	0.1323	0.1543	0.1858	9.35e-5	10.07e-5	13.95e-5	11.61e-5
Mono-cyte	0.0038	0.0042	0.0039	0.0023	0.0048	0.0542	0.0710	0.0659	0.0579	0.0550	7.50e-5	7.65e-5	5.20e-5	8.62e-5
All together	0.0028	0.0031	0.0030	0.0019	0.0035	0.0528	0.0734	0.0708	0.0578	0.0516	5.60e-5	5.95e-5	5.95-5	6.30e-5
Gestational Diabetes Data (N=93)														
Placenta	0.0023	0.0018	0.0012	0.0012	0.0020	0.0808	0.0906	0.0882	0.0987	0.0873	6.09e-5	4.81e-5	5.09e-5	5.71e-05
Cord blood	0.0024	0.0014	0.0014	0.0010	0.0024	0.0384	0.0463	0.0456	0.0473	0.0477	4.38e-5	4.59e-5	3.37e-5	2.89e-05
All together	0.0055	0.0052	0.0051	0.0061	0.0052	0.0666	0.0608	0.0592	0.0620	0.0602	9.77e-5	9.38e-5	14.14e-5	16.53e-5

* GENKI cannot be calculated in the output from *funNorm* inside the Bioconductor package minfi, since the SNP probes are removed. Therefore, for comparison, here we also provide results using the raw data.