SUPPLEMENTARY FIGURES

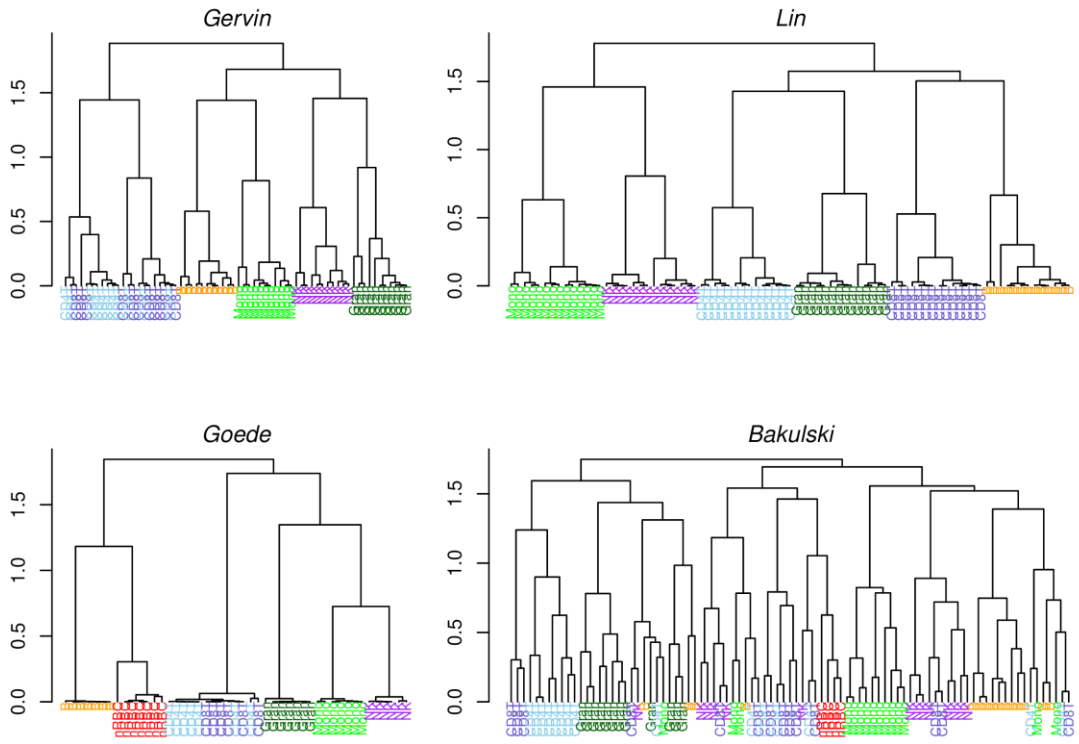

Fig S1: Hierarchical clustering of the 4 cord-blood DNAm datasets prior to sample removal. For each of the 4 studies we display the clustering dendrogram, with samples annotated by sorted cell-type. Color labels: skyblue=CD4T, darkblue=CD8T, orange=B, lightgreen=Monocytes, purple=NK-cells, darkgreen=Granulocytes.

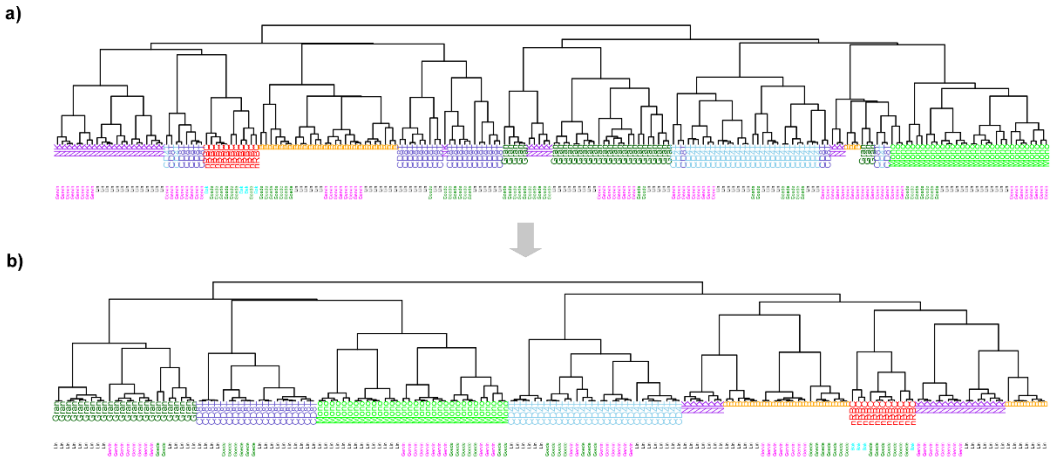

Fig S2: Hierarchical clustering of the merged cord blood DNAm dataset. a) Hierarchical clustering dendrogram of the merged samples from the 4 cord blood DNAm datasets, after removal of previous low purity samples. Sorted immune cell-types and study sorted sample is derived from is indicated below the dendrogram. Color labels for cell-types as in Fig S1. Color labels for studies: Magenta=Gervin, Black=Lin, Darkgreen=Goede, Cyan=Bakulski. **b)** As a) but after removing additional low purity samples, as inferred from a). The residual segregation of NK and B-cells is not observed in PCA-plots, and not observed in a), hence we conclude that this is not likely to cause confusion in downstream supervised analyses.

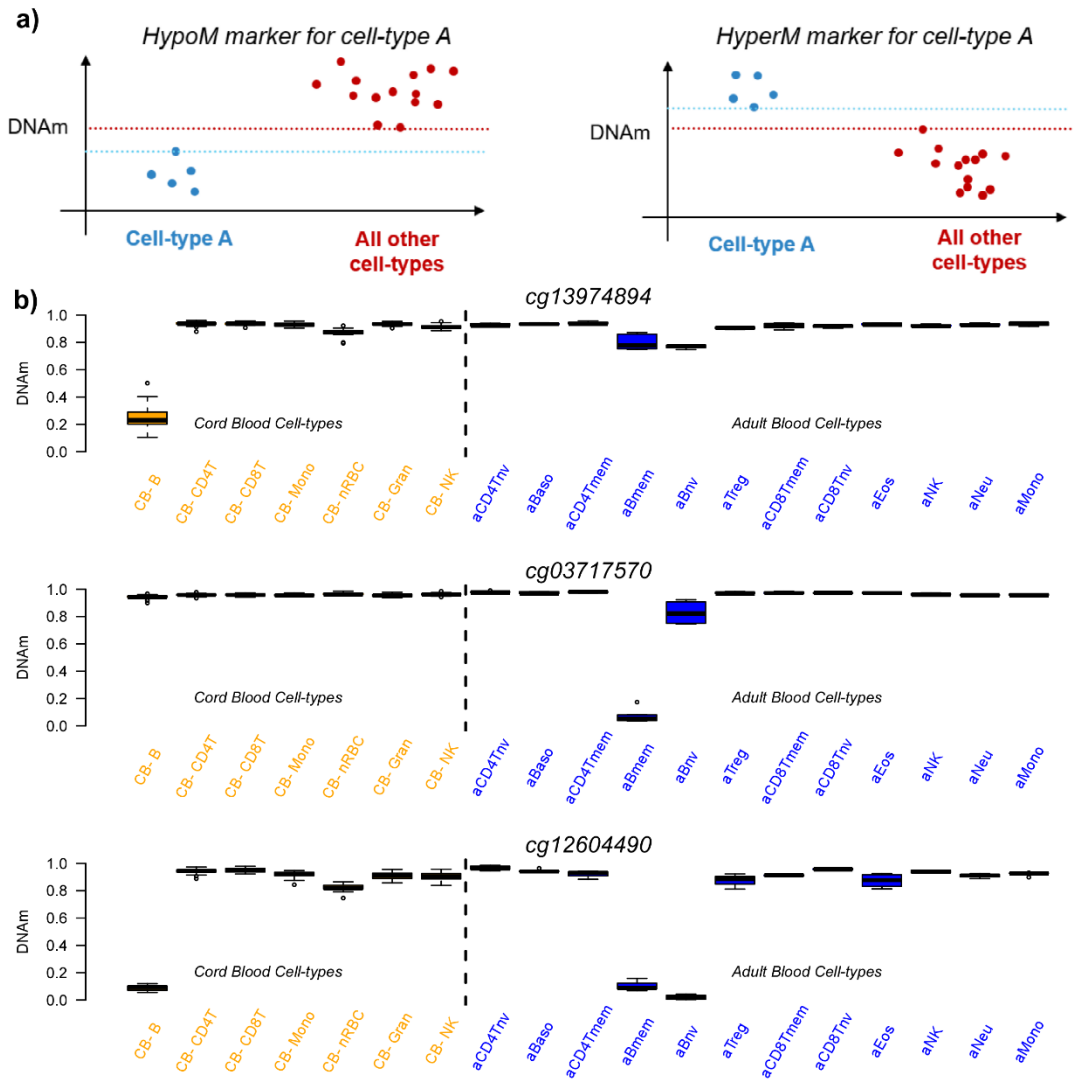

Fig S3: Definition of Gap specificity score and real examples of B-cell markers. a) Diagrams illustrate hypothetical examples of cell-type specific hypomethylated (left) and hypermethylated (right) markers for a given cell-type "A" compared to all other cell-types, for which the gap specificity score is indicated as the difference between the blue and red dashed lines. For hypomethylated markers, we want the blue line to be lower than the red one, and vice-versa for hypermethylated markers. Features can be ranked according to this gap-specificity score, to ensure that markers are truly cell-type specific. **b)** Boxplots depicting the DNAm distributions of 3 selected markers across the 19 immune cell-types, the first six being from cord-blood (CB, orange labels) and the rest (12) being from adult blood (blue labels). The top marker is specific to cord-blood B-cells, the middle marker is specific for adult memory B-cells and the bottom marker is specific for all 3 B-cell types, although it is lowest for adult naïve B-cells. Whilst not strongly specific to naïve B-cells, to confidently distinguish these 3 B- cell types from each other it suffices to have the top 2 markers, with the bottom one distinguishing B-cells from the rest.

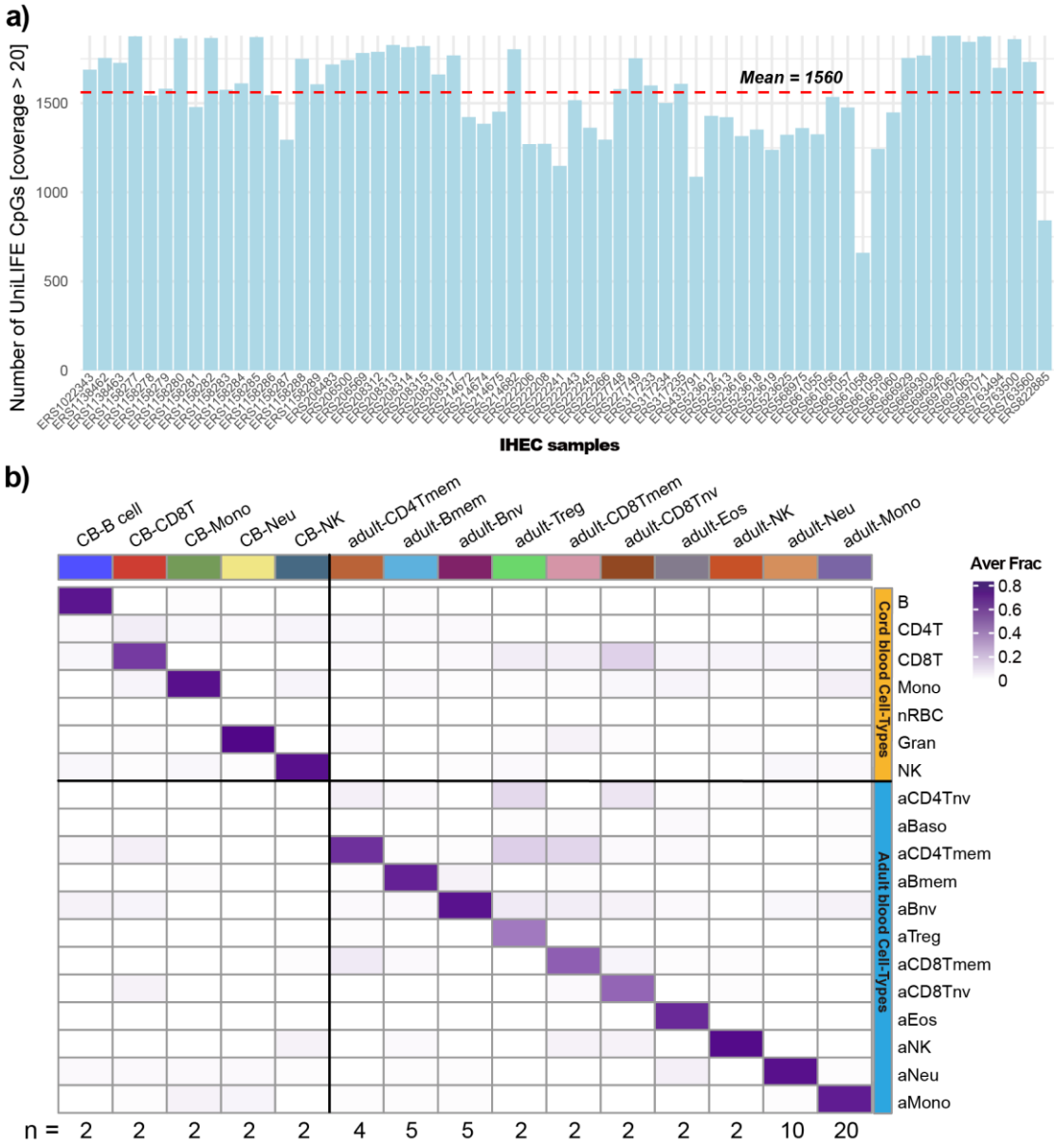

Fig S4: Validation of UniLIFE DNAm reference in sorted WGBS from IHEC. a) Barplot displays the number of UniLIFE CpGs (total=1906) that are covered by at least 20 reads in each of the sorted WGBS IHEC samples**. b)** Heatmap displaying the estimated UniLIFE fractions (rows) of sorted cord blood (CB) and adult immune cell-type WGBS samples from IHEC (columns) using a 20-read coverage threshold. The number of independent samples for each sorted cell-type is given below the heatmap. The estimated fractions were averaged over these independent samples of each cell-type.

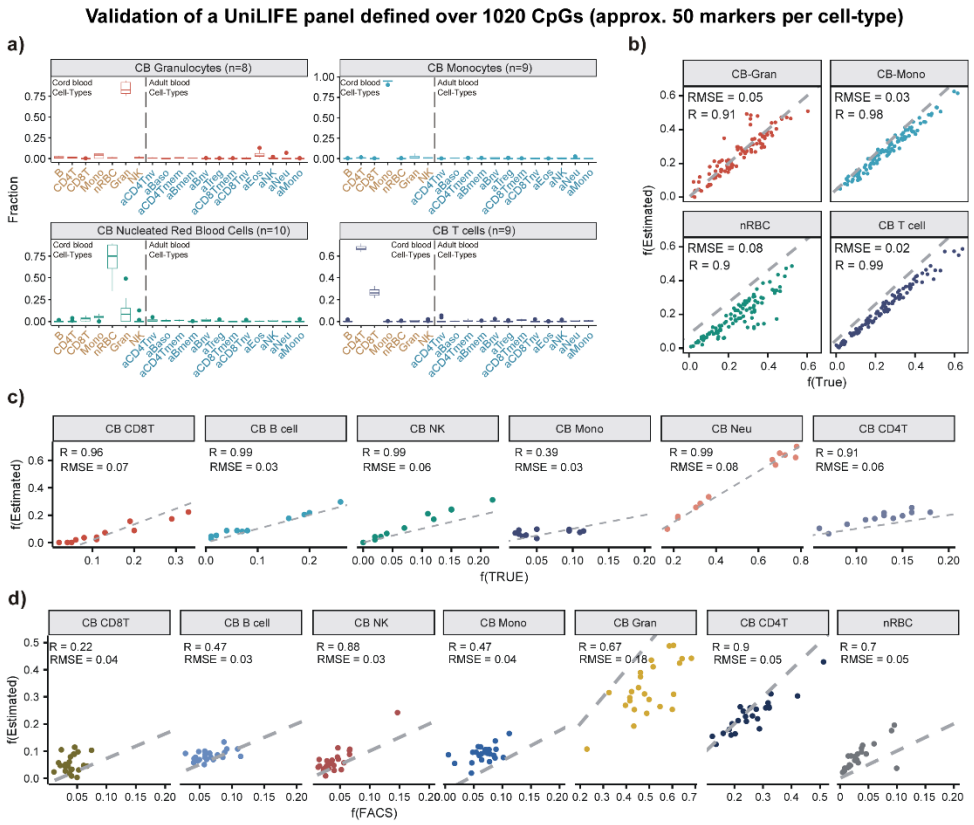

Fig S5: Validation of the smaller UniLIFE panel of 1020 CpGs in cord blood. a) Boxplots display the inferred cell type proportions for four sorted cord blood cell types using an independent Illumina 450k dataset of deGoede. The number of sorted samples corresponding to each cell type is provided. **b)** Scatterplots of true mixture weights (x-axis) vs estimated fractions (y-axis) for 100 in-silico mixtures. Each mixture is a linear combination of 4 sorted cord blood samples from a) with weights drawn randomly from a uniform distribution. Pearson R-value and RMSE are given. **c)** Scatterplots of true fractions vs estimated fractions for 6 cord blood cell subtypes using an independent EPIC DNAm dataset of 12 umbilical cord blood experimental mixtures where the underlying mixing proportions were known. For each estimated cell type, we display the Pearson R-value and RMSE. **d)** Scatterplots of FACS-derived fractions vs estimated fractions for 24 independent cord blood 450k DNAm profiles with matched FACS cell-counts for 7 cell-types, as shown. Pearson R-value and RMSE are given.

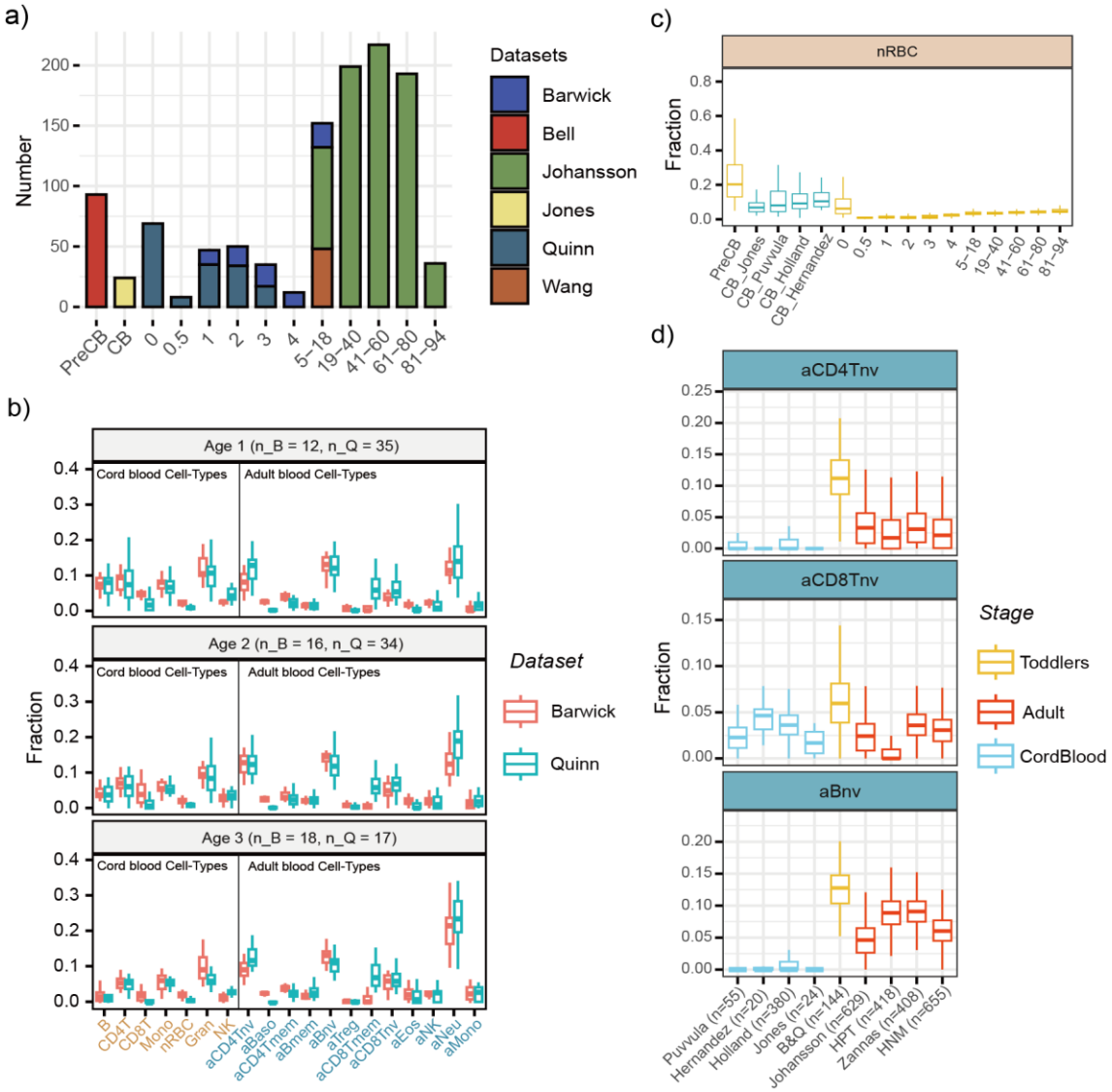

Fig S6: Comparison of batch effects and age-related changes in cell type fractions.

a) The distribution of all datasets used in Main Figure 5 suggests potential batch effects between different datasets from various sources and age groups (PreCB: cord blood samples from premature infants, CB: corb blood samples from normal infants). **b)** A comparison of cell type fractions for each age from 1 to 3 years between the Barwick and Quinn datasets reveals that the fractions between datasets are similar. The number of samples included at each age for each dataset is provided (n_B: number in Barwick, n_Q: number in Quinn). Cord-blood cell-types labeled in orange. Adult-blood cell-types labeled in skyblue. **c)** Boxplot of nucleated red blood cell (nRBC) fractions across different age groups. In addition to the Jones CB dataset, three additional CB datasets (Puvvula, Holland, and Hernandez) were included to compare batch effects with true biological signals. **d)** Boxplots show that toddlers have the highest proportions of adult naive cell types, and this biological effect is significantly greater than the batch effect (three additional datasets of adults over 20 years old were included to demonstrate the robustness of the maximum for toddler group).

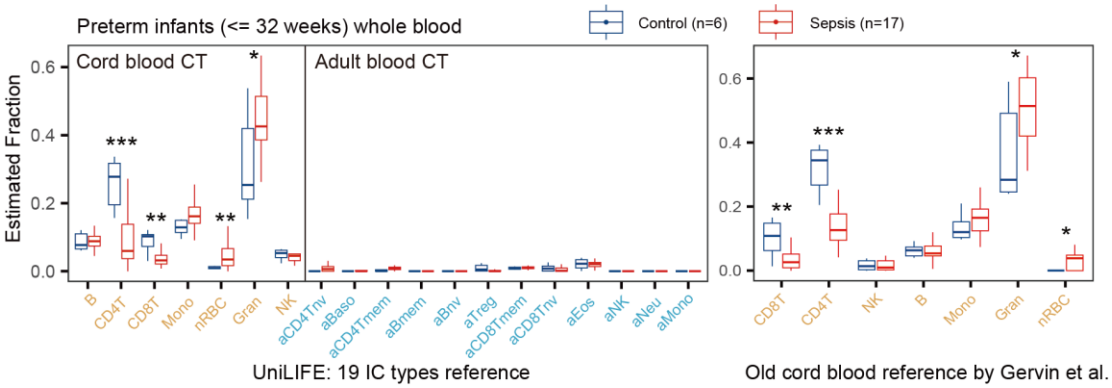

Fig S7: Immune cell fraction variation in preterm sepsis cases and controls. Left panel: Boxplots display the estimated 19 immune cell-type fractions in a cord blood DNAm dataset of 17 preterm sepsis cases and 6 preterm controls. Cord-blood cell-types labeled in orange. Adult-blood cell-types labeled in skyblue. Significance levels are derived from the Wilcoxon rank-sum test, with the symbols *** representing $P < 0.001$, ** representing $P < 0.01$, and * representing $P < 0.05$. **Right panel:** As right panel, but for 7 cord blood cell-type fractions, as estimated using the cord-blood reference DNAm panel from Gervin et al.

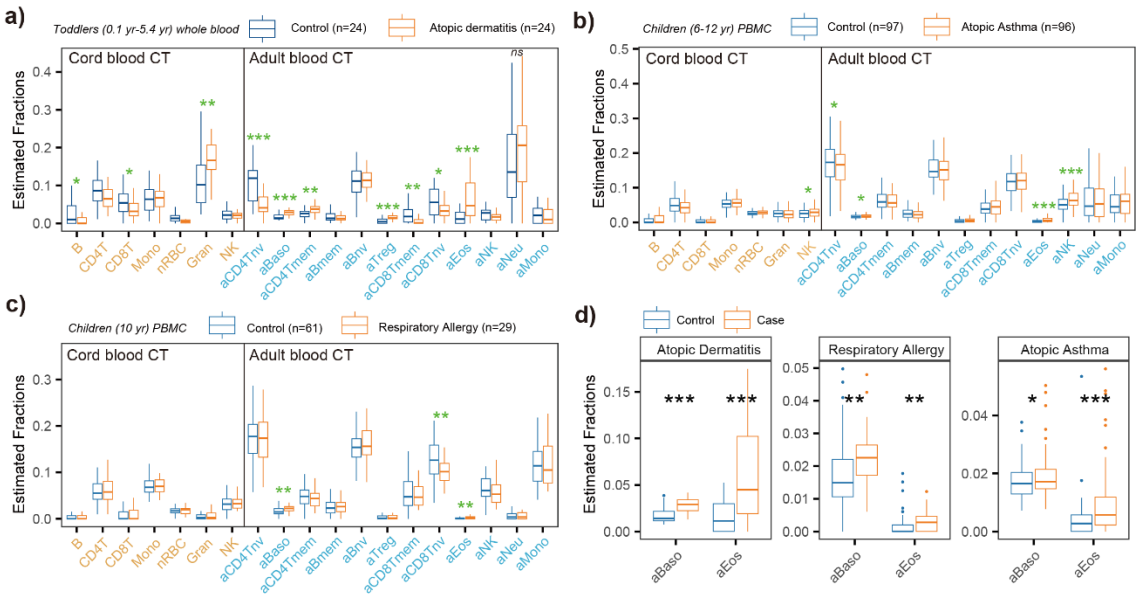

Fig S8: Increased eosinophil and basophil fractions in dermatitis, allergy and asthma. a) Boxplot displaying the estimated immune cell type fractions for a DNAm dataset profiling whole blood from toddlers with and without atopic dermatitis. Significance levels are derived from the Wilcoxon rank-sum test, with the symbols *** representing $P < 0.001$, ** representing $P < 0.01$, and * representing $P < 0.05$. **b)** As a) but for a DNAm dataset profiling peripheral blood mononuclear cells of children with and without asthma. **c)** As a) but for a DNAm dataset profiling peripheral blood mononuclear cells of children with and without respiratory allergy. **d)** Boxplots display zoomed in versions of the estimated adult eosinophil and basophil fractions between cases and controls for each of the 3 studies displayed in a-c). Significance levels are derived from the Wilcoxon rank-sum test, with the symbols *** representing $P < 0.001$, ** representing $P < 0.01$, and * representing $P < 0.05$.

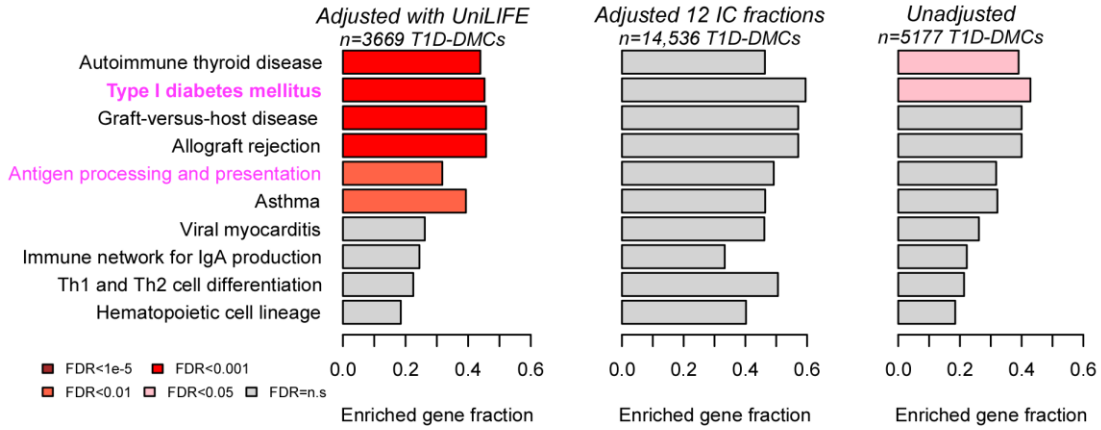

Fig S9: Comparative enrichment analysis using common FDR threshold. Left: The enriched fraction of genes (x-axis) of the top-10 enriched KEGG pathways among the 3669 T1D-DMCs (FDR<0.05) derived with UniLIFE in the DAISY prospective T1D study. Color of bars indicate statistical significance of enrichment (FDR) level, as shown. Middle & Right panels: As left, but now for the 14,536 T1D-DMCs (FDR<0.05) and 5,177 T1D-DMCs (FDR<0.05) derived using adjustment with 12 adult immune cell fractions or with no adjustment for cell-type heterogeneity.

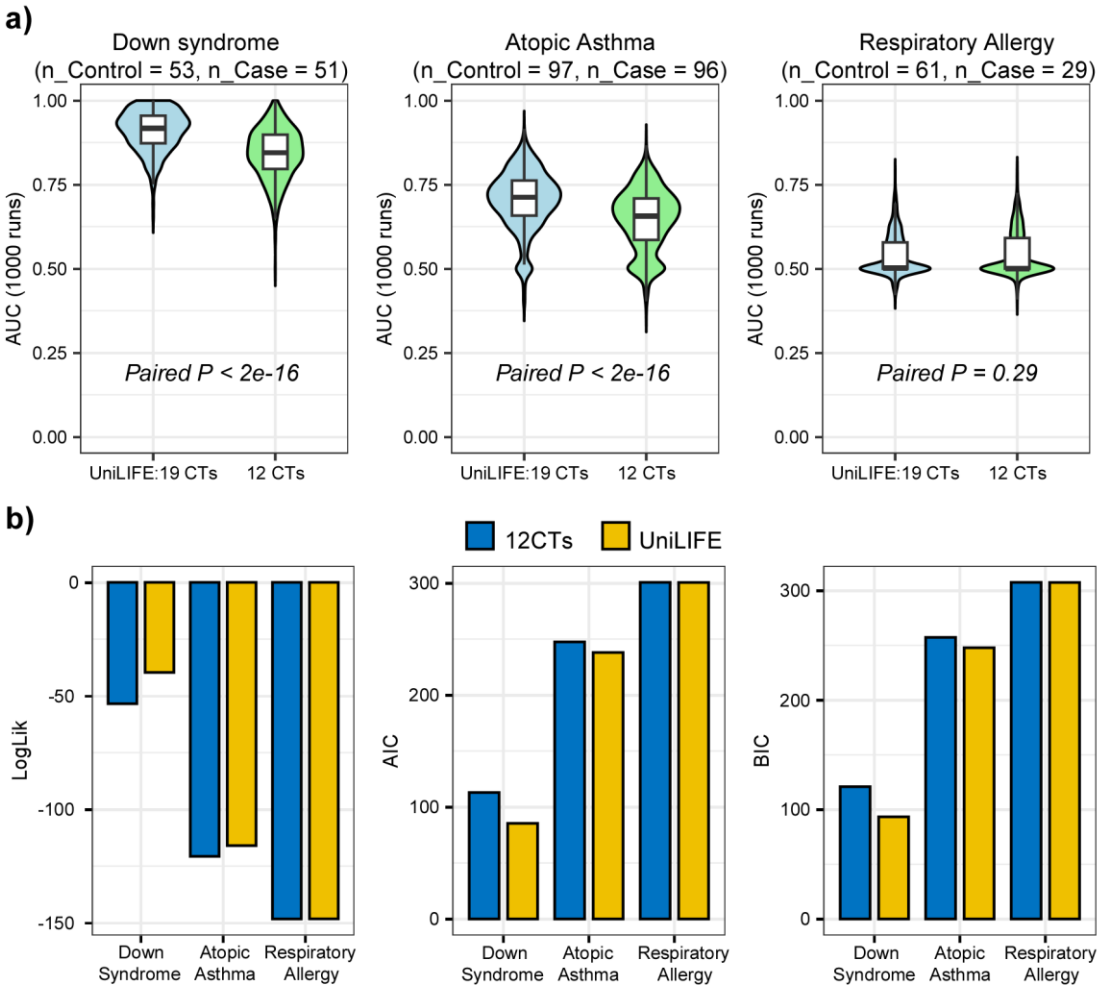

Fig S10: Comparison of UniLIFE and 12 cell-type predictive models for pediatric diseases. a) For three separate diseases, we display violin plots of AUC values obtained by elastic net logistic models trained on 19 immune cell-type fractions (UniLIFE) and 12 adult immune cell fractions (12 CTs), on respective test sets. Models were trained to predict case-control status. Number of cases and controls given above each plot. Each violin plot contains 1000 AUC values for 1000 different random training-test set partitions. P-value is from a paired two-tailed Wilcoxon rank sum test. **b)** Barplots representing the log-Likelihood (left), Aikake Information Criterion (AIC) and Bayesian Information Criterion (BIC) for optimal logistic multivariate predictors inferred by a nested forward-selection strategy. Lower AIC and BIC values indicate better models taking into account model complexity (number of cell-type fractions).

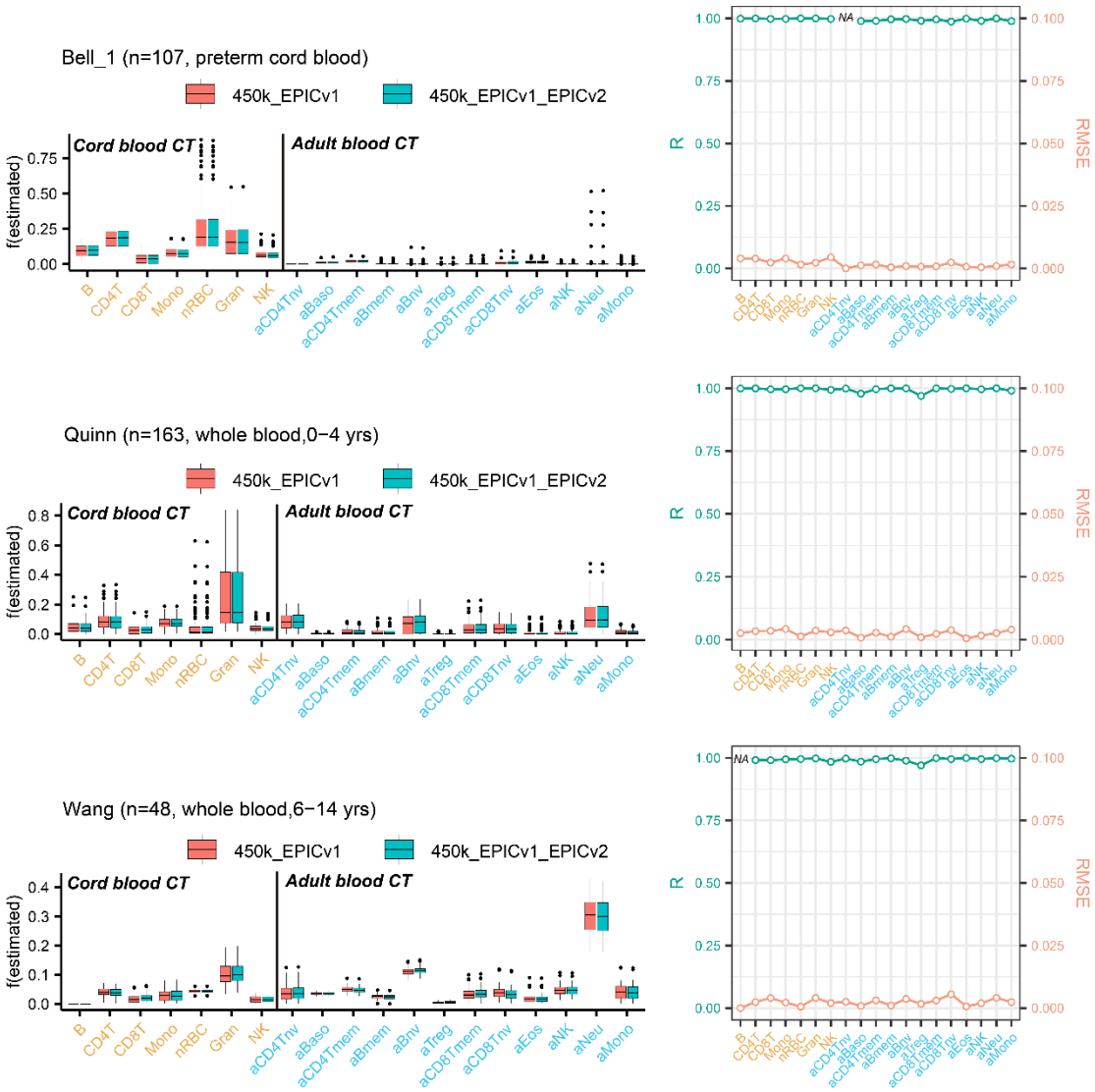

Fig S11: Robustness of cell-type fraction estimation to missing probes on EPICv2. Left panels: For each of 3 Illumina DNAm datasets we contrast boxplots of estimated cell-type fractions, as obtained with UniLIFE using all joint EPICv1-450k CpGs in the reference (red) and by removing those not present on EPICv2, i.e. using only joint EPICv1-EPICv2-450k probes (blue). Right panels: For each dataset, R-values and RMSEs for all cell-types. Cord-blood cell-types are labelled in orange, adult blood cell-types are labelled in skyblue. NA means that no R-value (Pearson Correlation Coefficient) is computable because all fractions were zero.

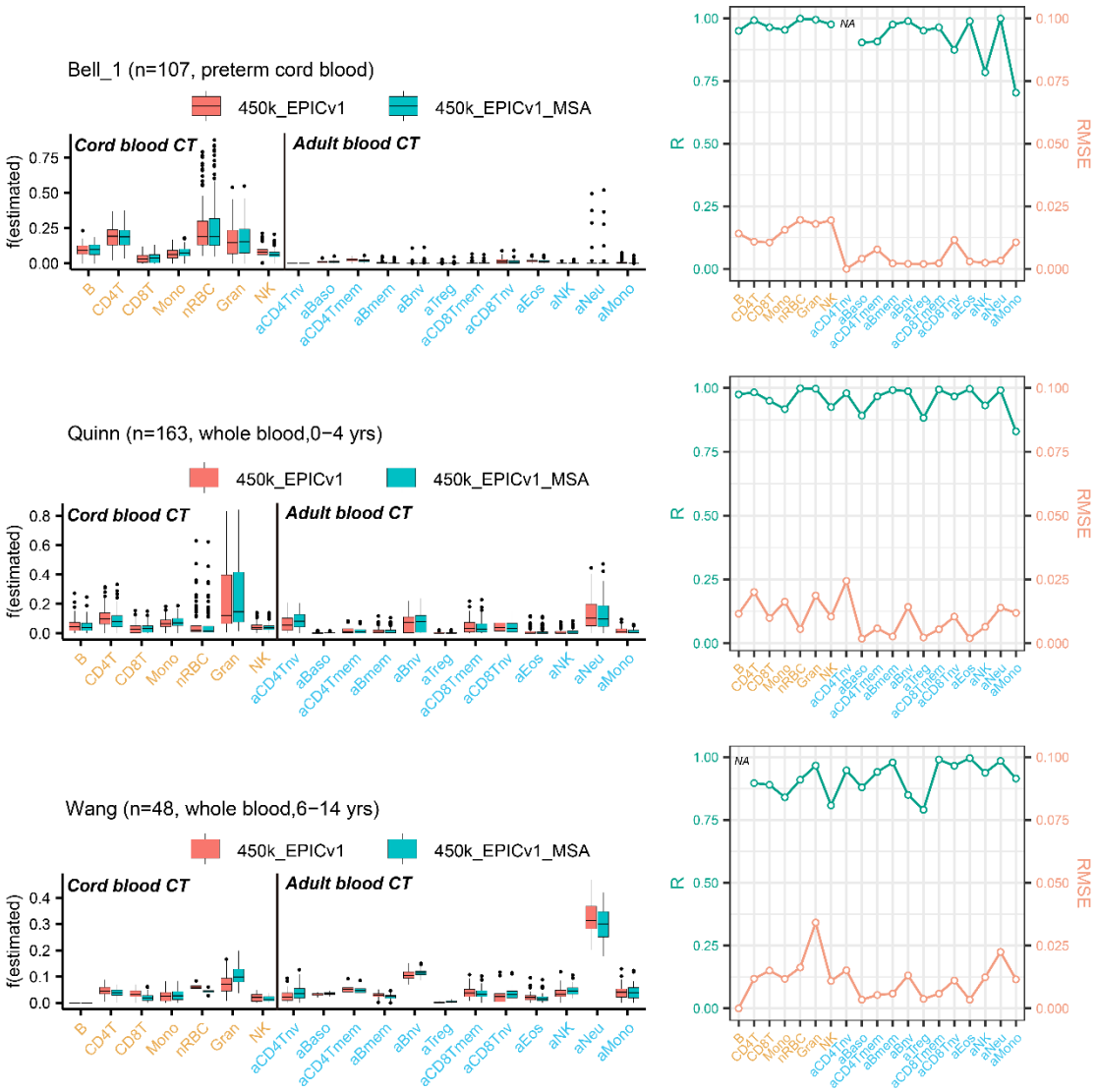

Fig S12: Robustness of cell-type fraction estimation to missing probes on MSA. Left panels: For each of 3 Illumina DNAm datasets we contrast boxplots of estimated cell-type fractions, as obtained with UniLIFE using all joint EPICv1-450k CpGs in the reference (red) and by removing those not present on the Methylation Screening Array (MSA), i.e. using only joint EPICv1-450k-MSA probes (blue). Right panels: For each dataset, R-values and RMSEs for all cell-types. Cord-blood cell-types are labelled in orange, adult blood cell-types are labelled in skyblue. NA means that no R-value (Pearson Correlation Coefficient) is computable because all fractions were zero.

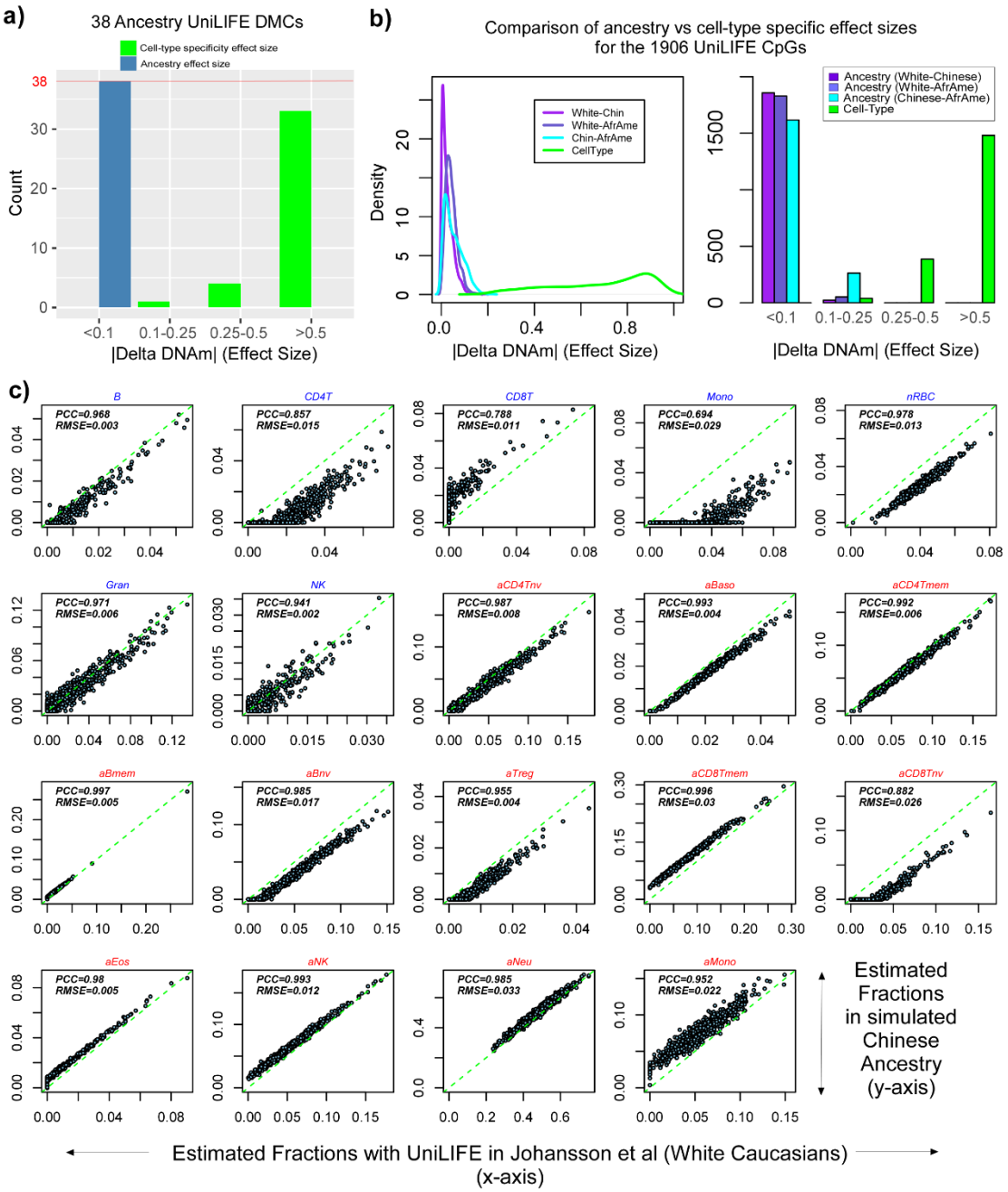

Fig S13: UniLIFE estimates are robust to ancestry. a) Barplot comparing the DNAm effect size of 38 UniLIFE CpGs that are also ancestry-DMCs, as derived from Zannas et al who profiled DNAm in 393 African Americans and 24 White Caucasians. The barplot compares the absolute DNAm difference between ancestries to the DNAm difference it displays between cell-types. **b)** Using 3 independent large DNAm datasets profiling White Caucasians (n=732, Johansson et al), Han Chinese (n=710, You et al) and African Americans (n=393, Zannas et al) we estimated ancestry effect sizes (absolute DNAm differences) between each pair of ancestries for all 1906 UniLIFE CpGs, displaying their probability density distributions and comparing them to the corresponding cell-type specific effect sizes. Right panel displays the same data as a barplot for different bins of effect size. **c)** A simulation of the effect of ancestry on the estimation of immune cell fractions in an adult blood cohort. Shifts in the DNAm values

of UniLIFE probes were induced by simulating Han Chinese specific effect sizes, as estimated from panel b), in the Johansson et al DNAm dataset. For each cell-type, x-axis labels the estimated fractions obtained with UniLIFE on the unperturbed Johansson (White Caucasian) dataset, whilst the y-axis labels the corresponding fractions when the same panel is applied on a simulated Han-Chinese cohort. Pearson correlation coefficients and root mean square errors (RMSE) are given.

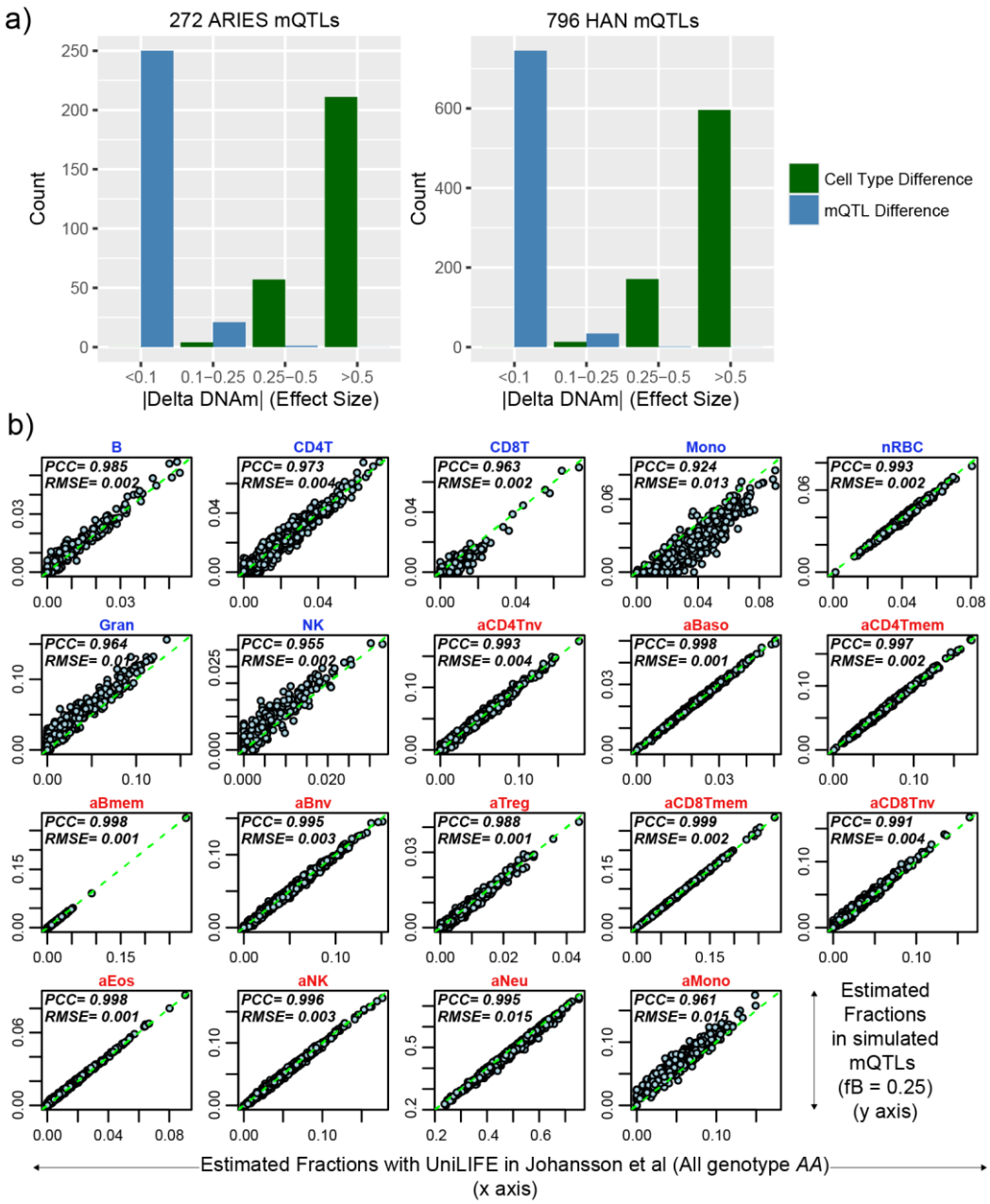

Fig S14: UniLIFE estimates are robust to variations in genotype. a) Barplot comparing the DNAm effect size of UniLIFE CpGs that are also mQTLs, as defined in the ARIES mQTL database or as derived in a large Han Chinese (TZH) cohort. The barplot compares the absolute DNAm difference between the AA and BB genotypes of the most significantly linked SNP, to the DNAm difference it displays between cell-types. **b)** A simulation of the effect of genotype on the estimation of immune cell fractions in an adult blood cohort. DNAm values of UniLIFE CpGs were perturbed by simulating mQTL effects in the Johansson et al DNAm dataset. For each cell-type, x-axis labels the estimated fractions obtained with UniLIFE on the unperturbed Johansson (White Caucasian) dataset, whilst the y-axis labels the corresponding fractions when the same panel is applied on the simulated mQTL-perturbed cohort. Pearson correlation coefficients and root mean square errors (RMSE) are given.