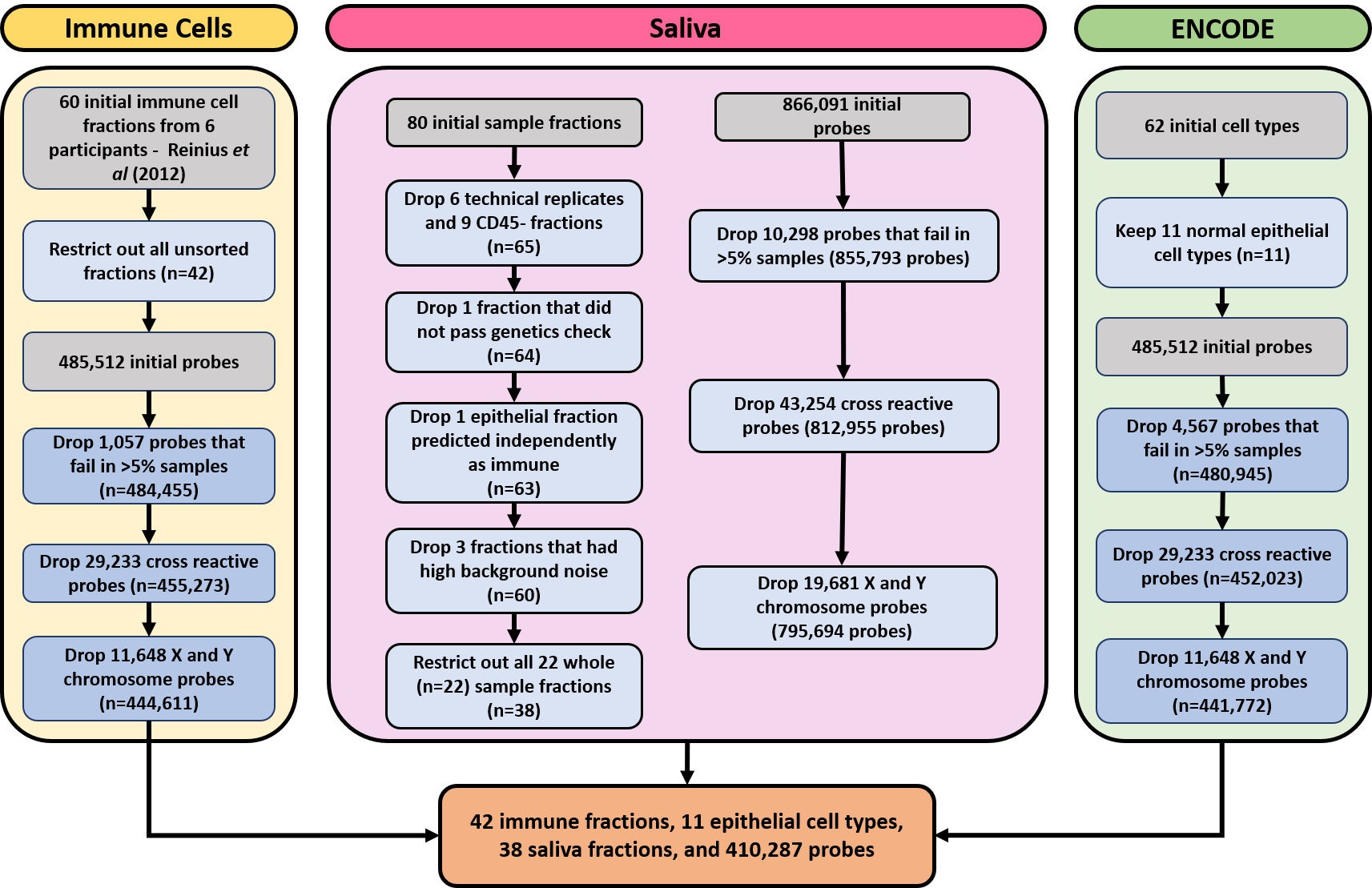
**Title**: Saliva cell type DNA methylation reference panel for epidemiology studies in children

**Authors:** Lauren Y. M. Middleton, John Dou, Jonah Fisher, Jonathan A. Heiss, Vy K. Nguyen, Allan C. Just, Jessica Faul, Erin B. Ware, Colter Mitchell, Justin A. Colacino, Kelly M. Bakulski

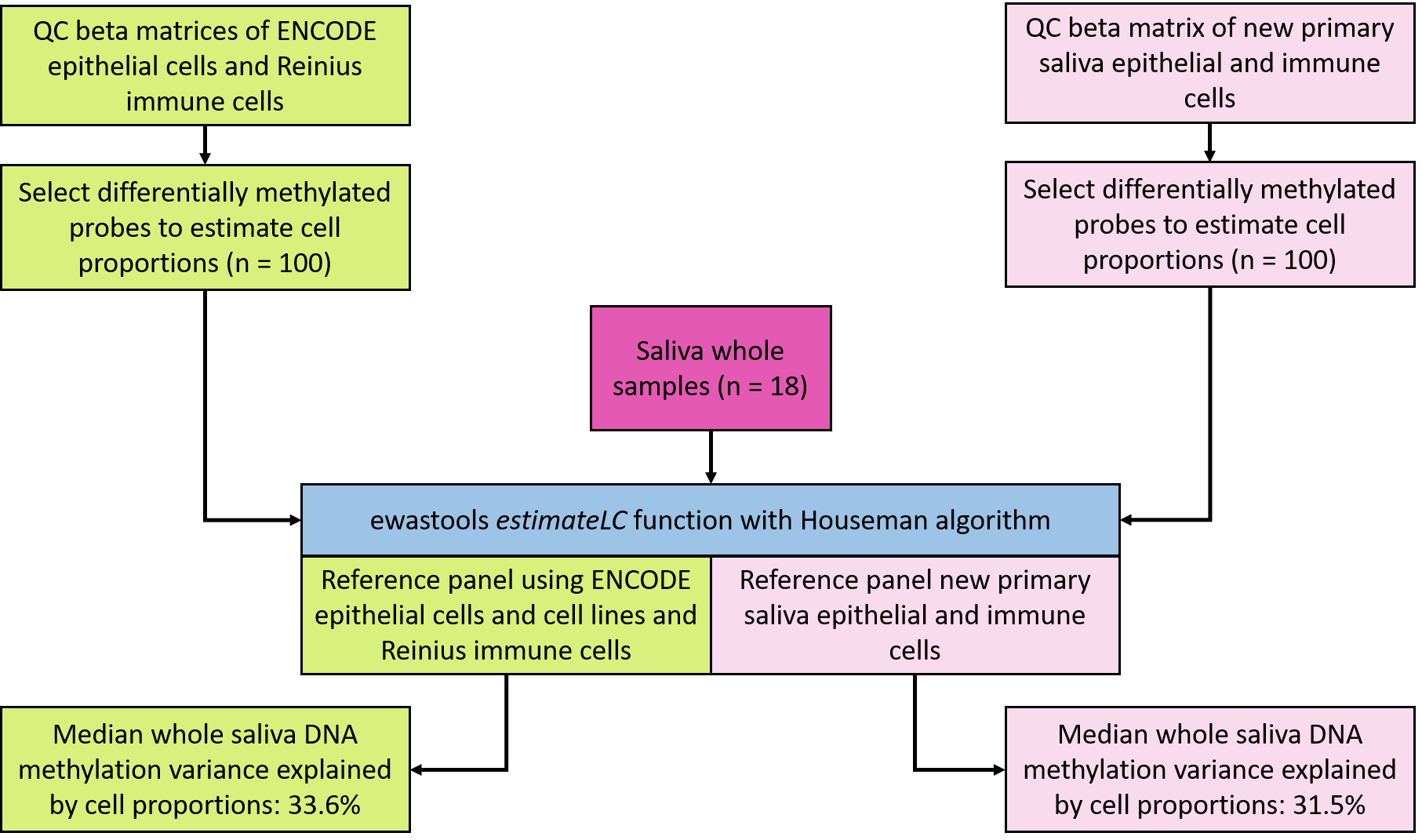
**Supplementary Information**

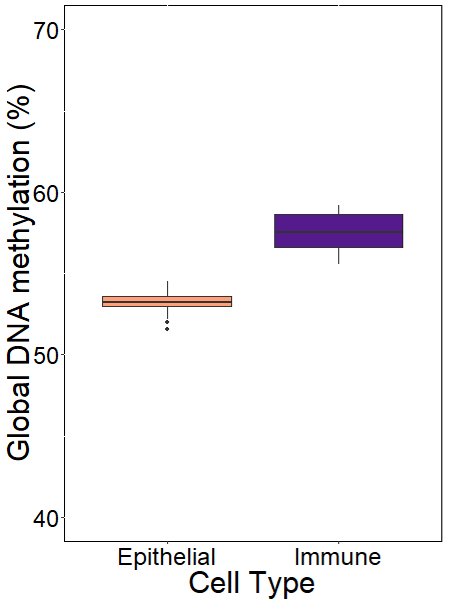
The following supplement contains nine supplemental figures. Four supplemental tables are also available in an excel workbook.

**Supplemental Figures**

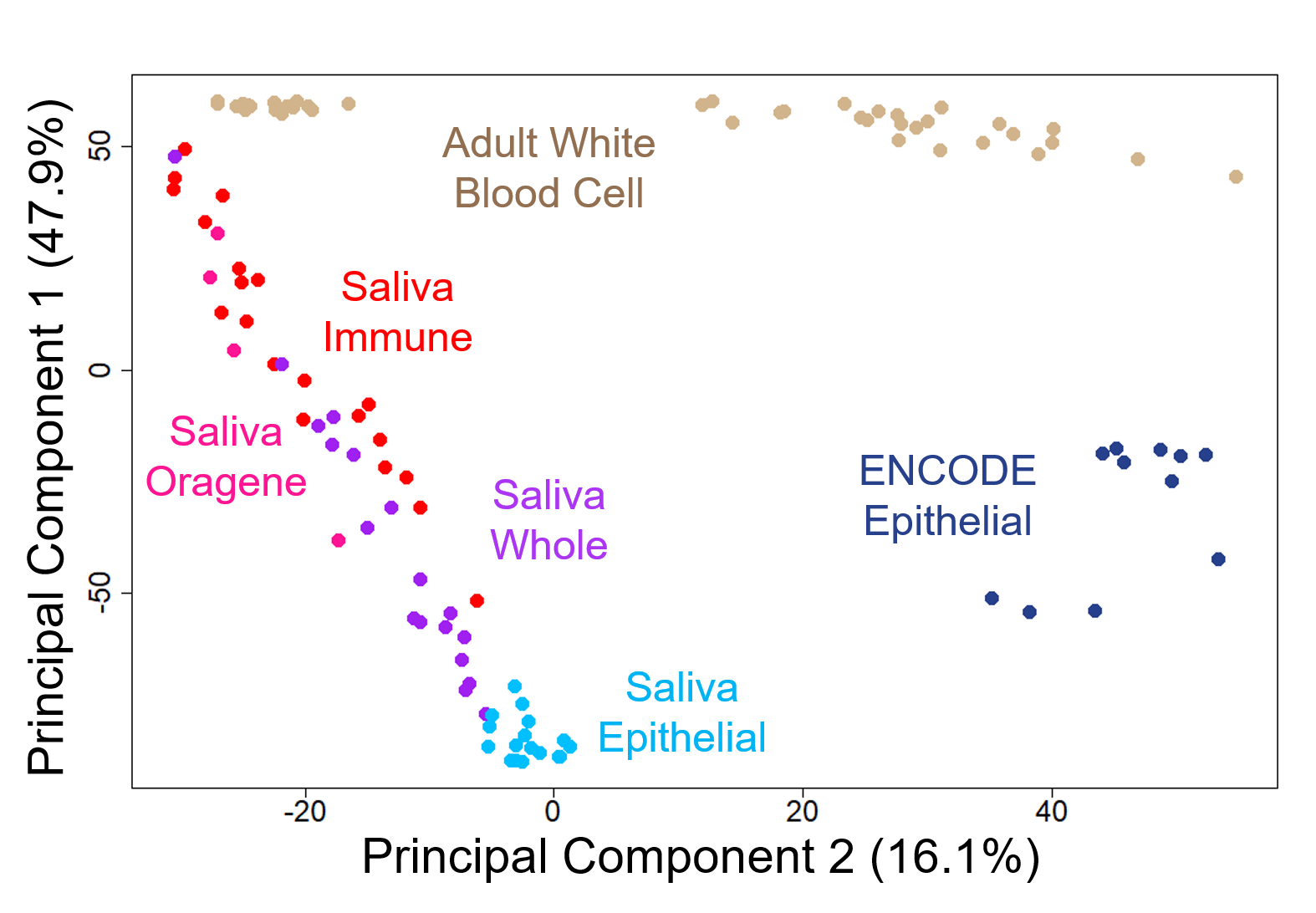
****

**Supplemental Figure 1.** Workflow to integrate three datasets. The immune cell dataset processing is on the left, the saliva dataset processing is in the middle and the ENCODE dataset processing is on the right. The immune cell dataset included seven adult white blood cell types from Reinius *et al* (2012).30 Only the 11 normal epithelial tissue ENCODE DNA methylation data29,31 were included in our analysis. Following data processing, there were 42 immune cell fractions of seven cell types, 11 epithelial cell types, 38 saliva fractions, and 410,287 overlapping probes for analysis.

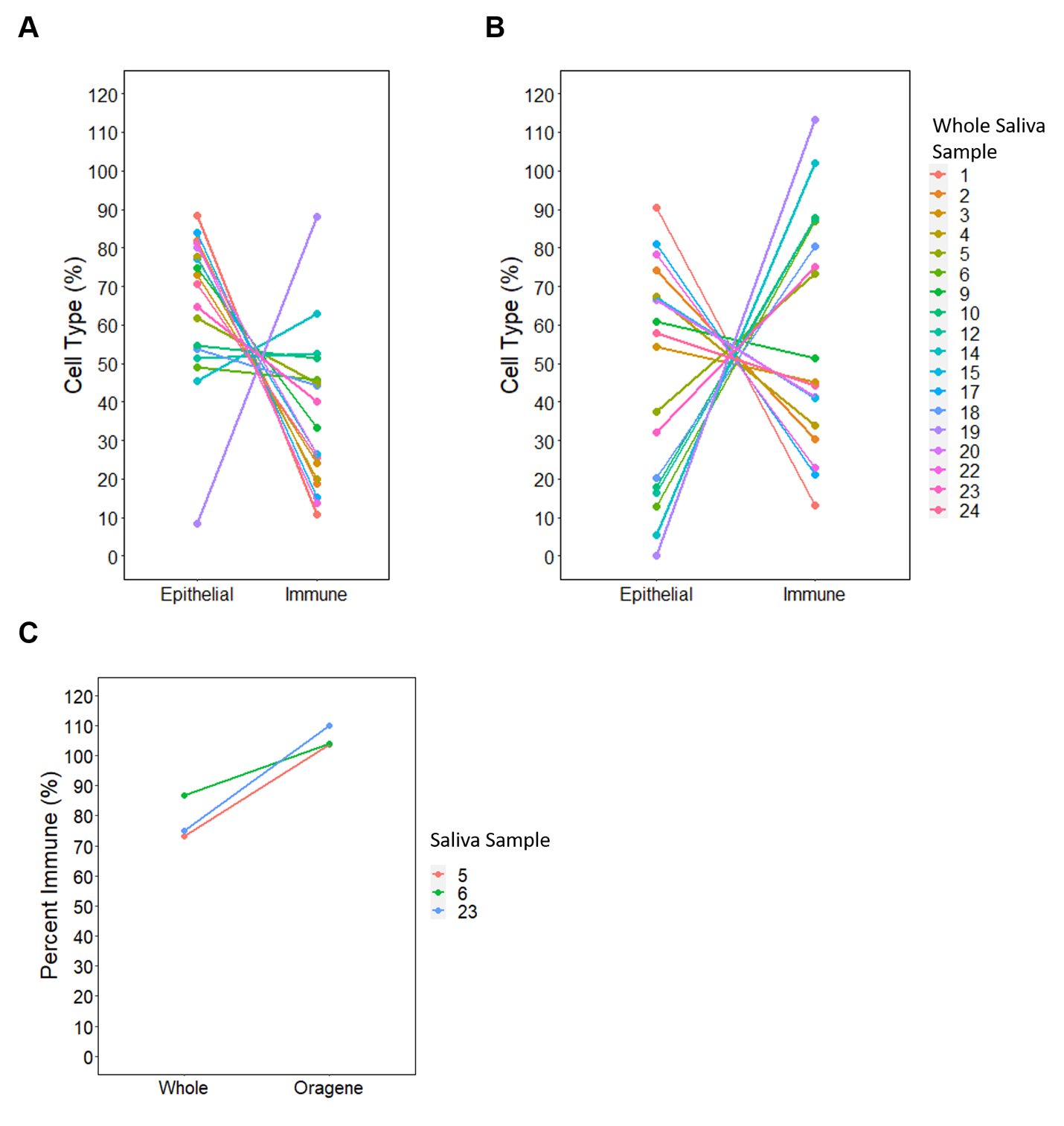
**Supplemental Figure 2.** Workflow to compare our new reference panel to a reference panel generated using an existing dataset. The ENCODE epithelial cell and Reinius adult white blood cell DNA methylation data were passed through our QC pipeline for processing.29–31 The QC steps were described in **Figure 1b**. The reference panels were constructed using the top 100 probes that were statistically significant from a t-test comparing epithelial and immune cell DNA methylation. Both our sorted saliva reference panel and the ENCODE/Reinius panel were integrated into ewastools and used to estimate the median variance explained by cell type in our whole saliva samples.



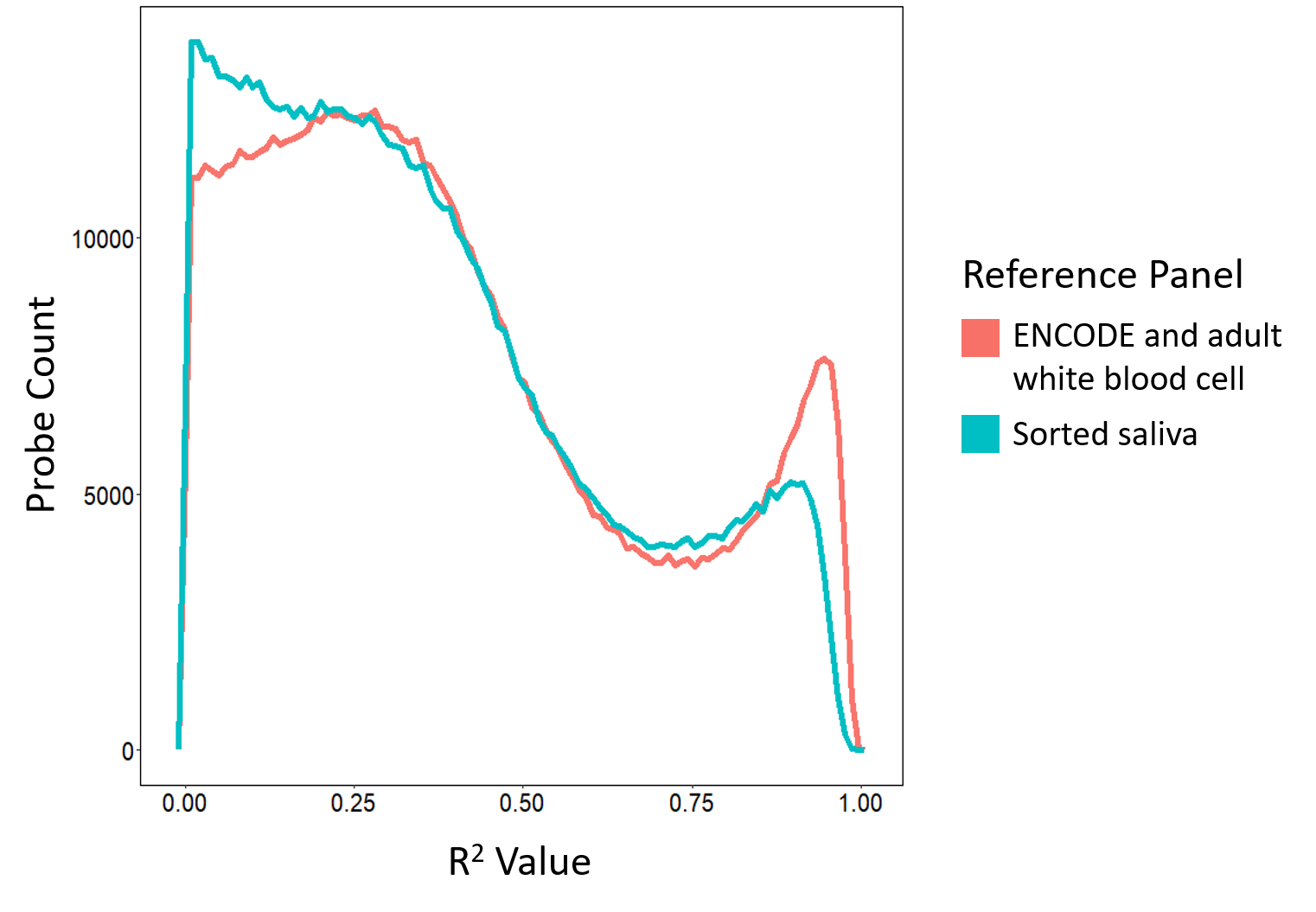
**Supplemental Figure 3.** Comparison of global DNA methylation across all probes in saliva epithelial and immune cell fractions. Box plots were used to visualize the percent global DNA methylation by cell type.

****

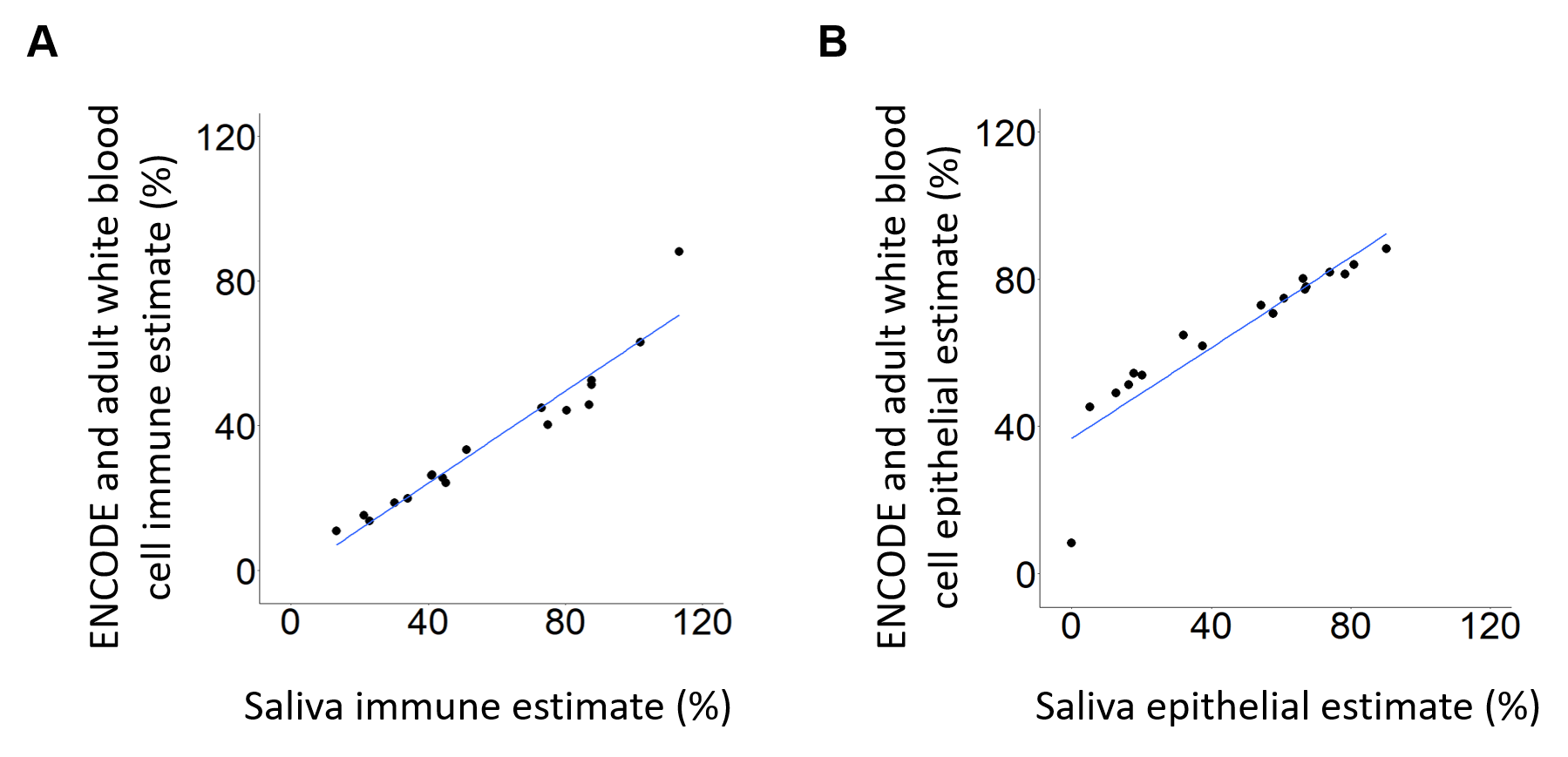
**Supplemental Figure 4.** Principal component analysis of DNA methylation datasets from ENCODE epithelial cells,29,31 adult white blood cells,30 and our saliva samples. The x- and y-axes show DNA methylation principal components one and two, colored by cell type with saliva immune cells in red, saliva epithelial cells in light blue, saliva whole samples in purple, saliva Oragene samples in pink, ENCODE epithelial cells in dark blue, and adult white blood cells in tan. Principal component 1 explained 47.9% of the variation in the DNA methylation data. Principal component 2 explained 16.1% of the variation in the DNA methylation data.

****

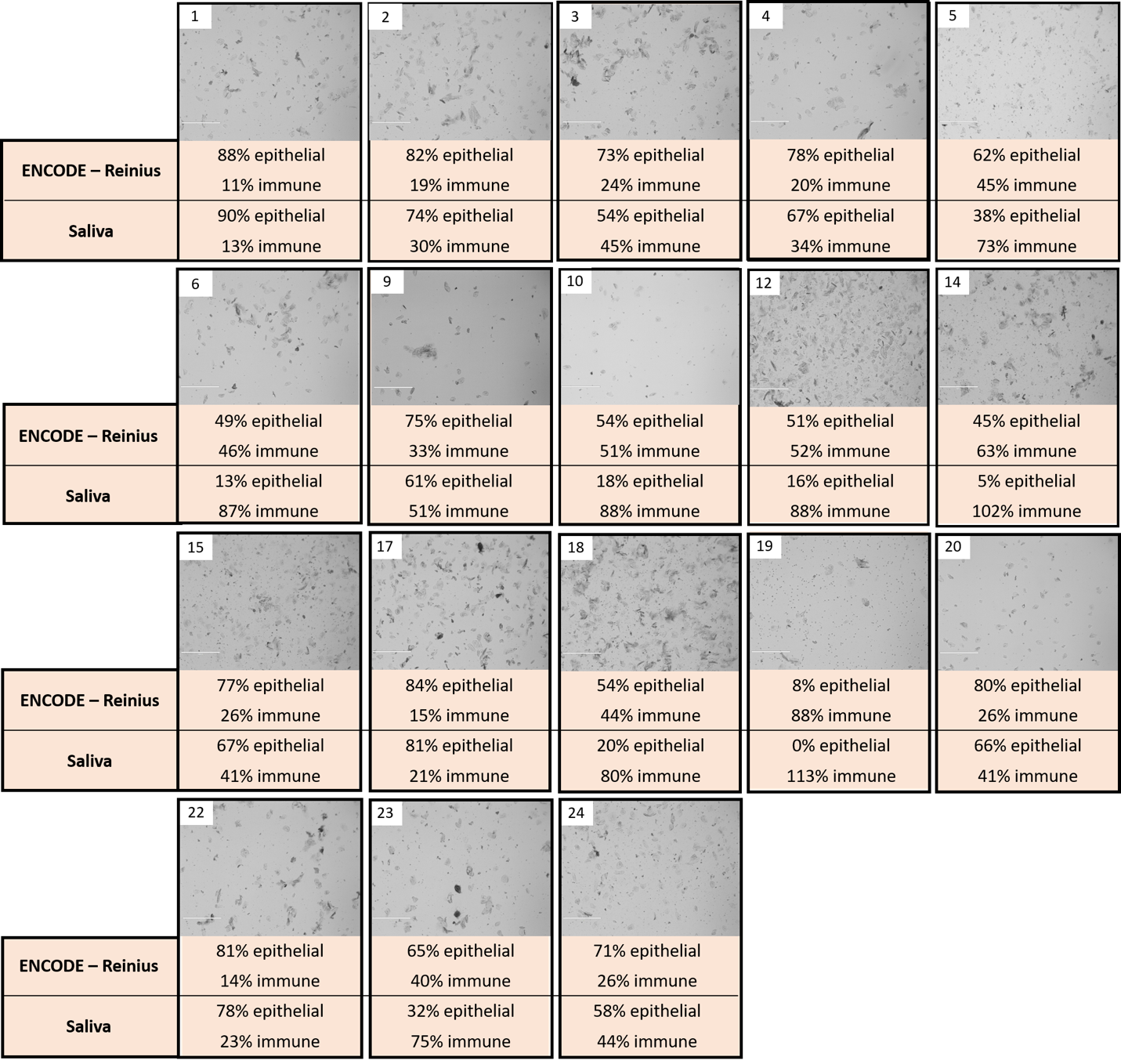
**Supplemental Figure 5.** Estimation of epithelial and immune cell proportions within whole saliva samples. **A)** The ENCODE and adult white blood cell reference panel, implemented through ewastools,23 was used to estimate the epithelial and immune percent of each whole sample. Each set of dots connected by a line represents one whole saliva sample. **B)** Ewastools with saliva as the reference dataset was used to estimate the epithelial and immune percent of each whole sample. Each set of dots connected by a line represents one whole saliva sample. **C)** Comparison of immune cell percent estimates in whole and Oragene saliva samples collected from the same individual. The estimates from the same participant are connected by a line.



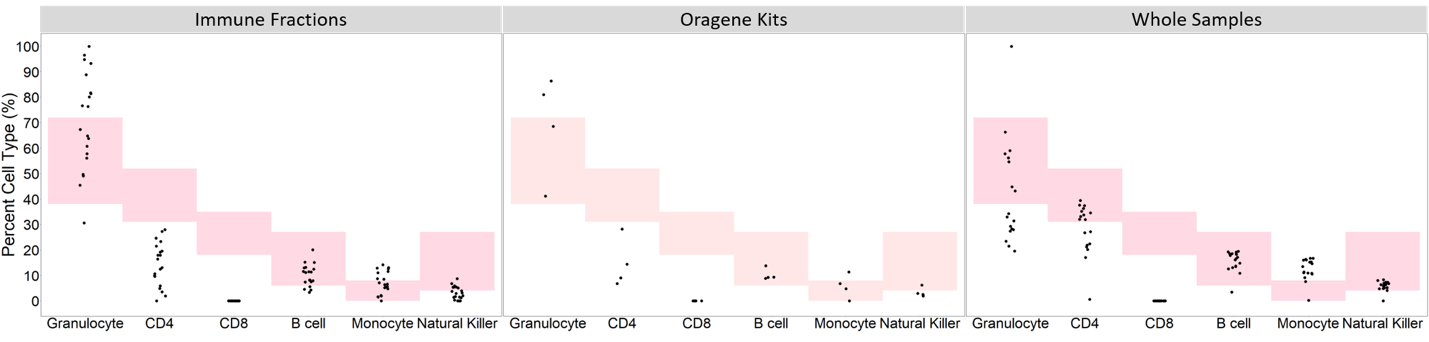
**Supplemental Figure 6**. Frequency polygram plot of variance explained in DNA methylation from whole saliva samples by cell proportions. A reference panel was created using ENCODE epithelial and adult white blood cell DNA methylation data. The frequency of R2 values was calculated for all probes in the whole saliva samples (n = 795,694) based on the ENCODE and adult white blood cell reference panel (pink). Our new reference panel was created using sorted saliva DNA methylation data. The frequency of R2 values was calculated for all probes in the whole saliva samples (n = 795,694) based on the saliva reference panel (blue).



**Supplemental Figure 7.** ENCODE and adult white blood cell reference panel vs sorted saliva reference panel percent estimates from whole saliva samples. The x-axis represents the percent of the cell type that was estimated using our saliva reference panel.23 The y-axis represents the percent of the cell type that was estimated using the ENCODE and adult white blood cell reference panel. **A)** The proportion of immune cells in each whole saliva sample was estimated using the two reference panels integrated into ewastools.23 **B)** The proportion of epithelial cells in each whole saliva sample was the two reference panels integrated into ewastools.23



**Supplemental Figure 8.** Comparison of whole saliva brightfield images (n = 18) to cell proportion estimates using the ENCODE29,31 and adult white blood cell (Reinius)30 reference panel and our saliva reference panel. Reference panels were implemented in ewastools.

****

**Supplemental Figure 9.** Cell types predicted within saliva immune cell fractions. We used the adult white blood cell reference panel30 in the ewastools23 package to deconvolute the immune fractions (n=20), Oragene samples (n=4), and whole samples (n=18). The blood reference panel includes granulocytes, CD4+ T cells, CD8+ T cells, B cells, monocytes, and natural killer cells. Each point in each blood immune cell type is one saliva sample. The height of the pink boxes represents the normal range of each immune cell type in pediatric blood.

**Supplemental Tables** are available in an excel workbook.

**Supplemental Table 1.** Description of ENCODE epithelial tissue data used.

**Supplemental Table 2.** AverageDNA methylation of all included probes by cell type and the mean DNA methylation difference and results of t-tests.

**Supplemental Table 3.** Gene ontology results from top 10,000 hypermethylated probes.

**Supplemental Table 4.** Gene ontology results from top 10,000 hypomethylated probes.