This folder contains the necessary files and R scripts needed to carry out the test benchmarking comparison of PCA-based and CIBERSORT-based cellular deconvolution of whole blood gene expression profiles presented in the Technical Validation section of O'Connell et al.

Note that this analysis requires the CIBERSORTx source code. The CIBERSORTx source code is the property of Stanford University, and thus, we cannot share it here. However, it can be directly requested from its authors at https://cibersortx.stanford.edu. We slightly modified the CIBERSORTx source code for use in our analysis to allow for the direct passage of matrix objects to the CIBERSORT() function as input, as opposed to reading in .txt files as originally written. *cibersort_source_code_modification.r* contains the lines of modified code.

pipeline_comparison_script.r contains the annotated R script used to statistically compare the performance of the two deconvolution pipelines via bootstrapping*. cibersort_reference_signature_matrix.csv* is the reference signature matrix used for deconvolution via CIBERSORTx, which was originally generated via the CIBERSORTx java application as described in the *project-007/cibersort_deconvolution/java_application* folder of this repository. *WBC4_126_marker_genes.csv* contains the WBC4.126 marker gene list used for PCA-based deconvolution.

actual_cell_counts.csv is an abbreviated example input file illustrating the format that the actual circulating leukocyte counts used to benchmark deconvolution performance should be provided in. Data should be in the form of relative cell counts (proportion of total WBC count). *whole_blood_raw_counts.csv* is an abbreviated example input file illustrating the format whole blood RNA sequencing data should be provided in. Data should be in the form of gene-level summarized read counts. The exact RNA sequencing and white blood cell differential data used for the analyses presented in the manuscript can be retrieved from the National Center for Biotechnology Information via BioProject accession number PRJNA949611.