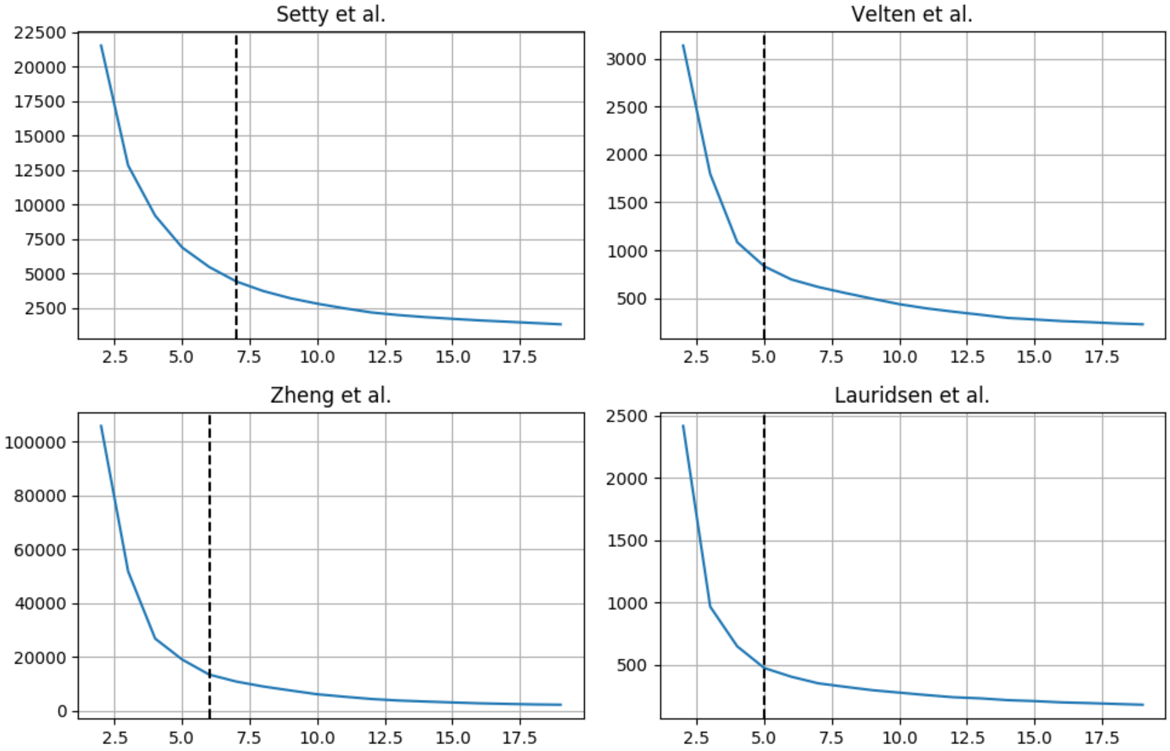
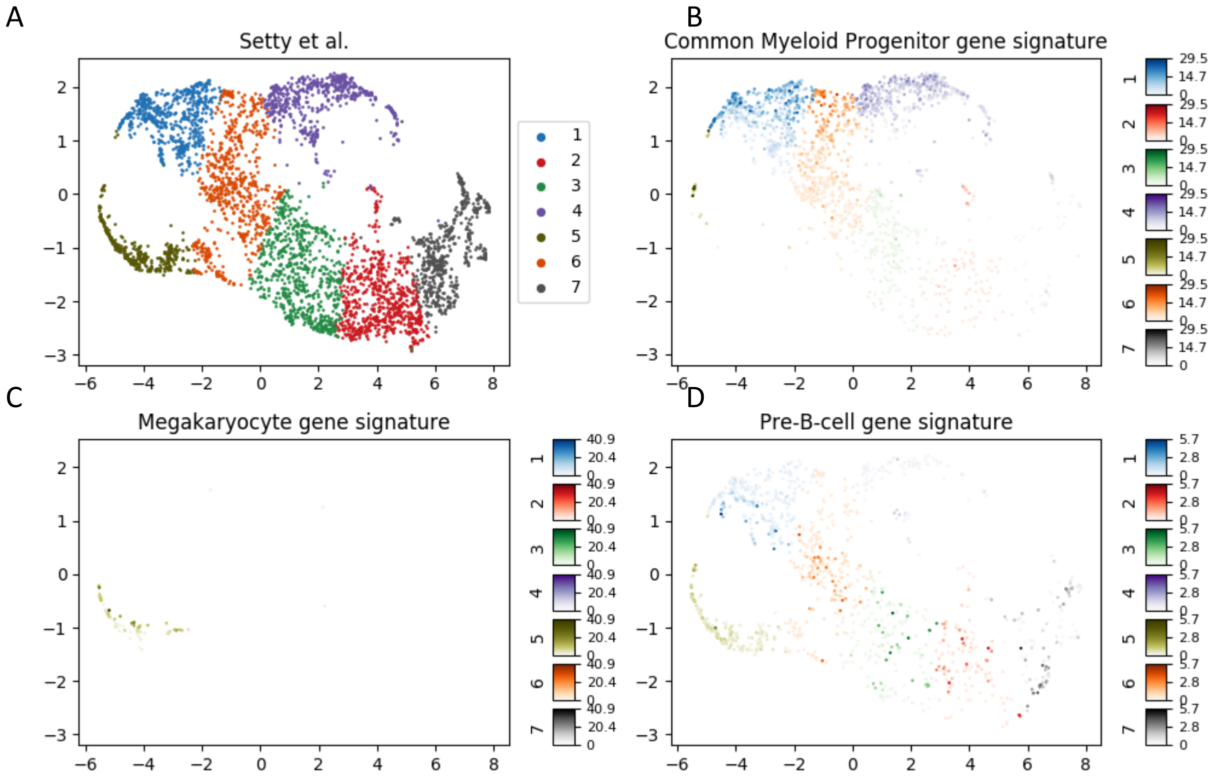
Supplementary figure 1



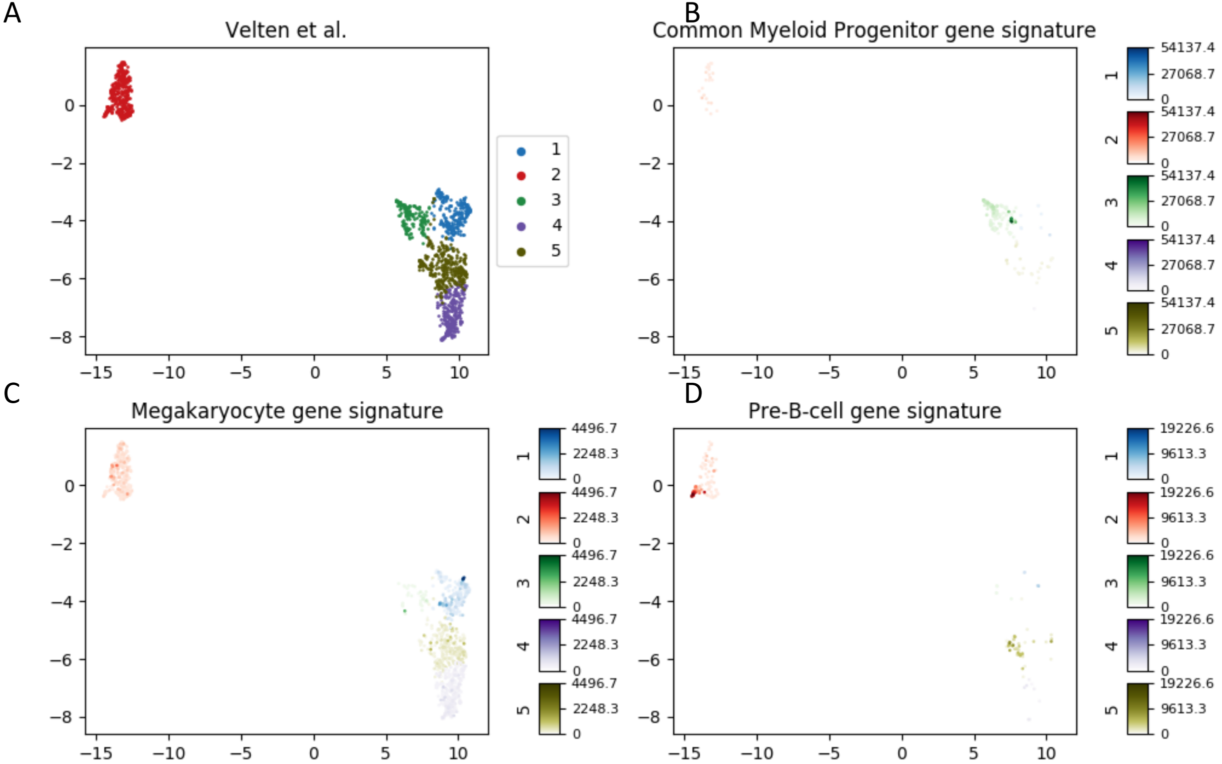
Supplementary Figure 1. Elbow curves of the inertia of the clusters along the number of clusters of the k-means algorithm, for each of the datasets as shown in the corresponding title. The vertical line denotes the k chosen for each dataset.

Supplementary figure 2



See legend after all supplementary figures

Supplementary 3



A Velten et al.

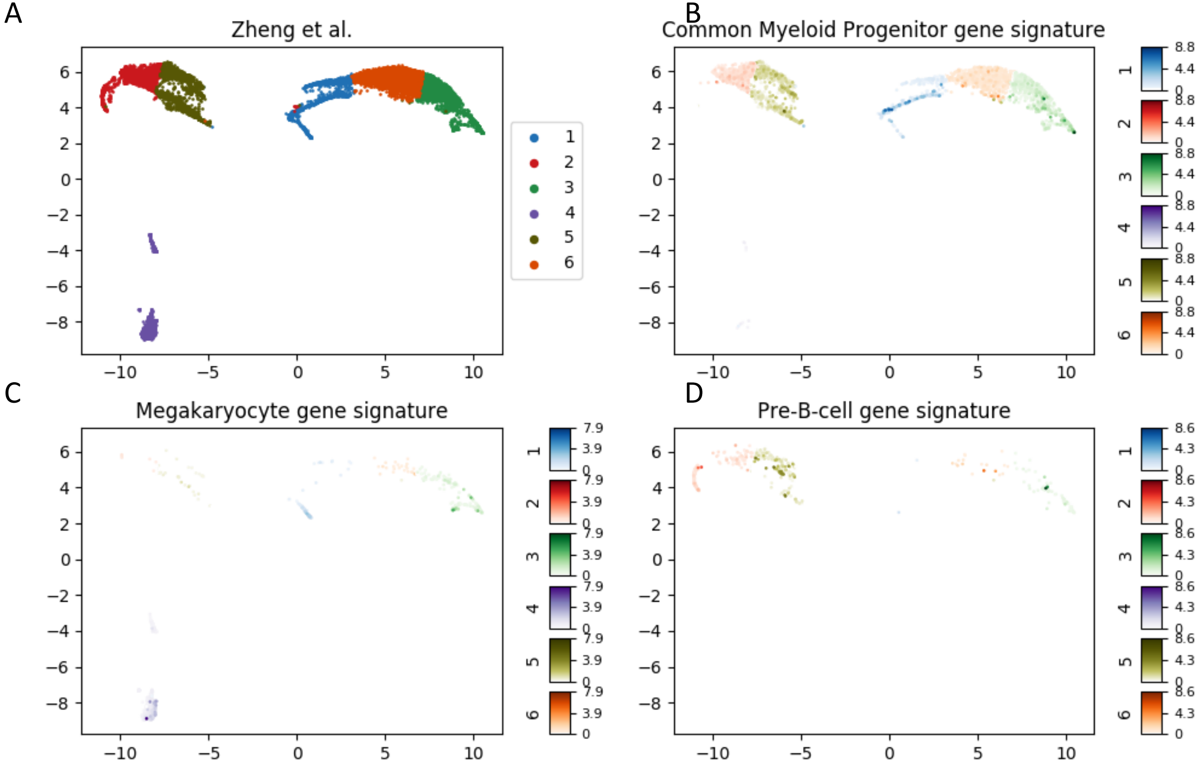
B Common Myeloid Progenitor gene signature

C Megakaryocyte gene signature

D Pre-B-cell gene signature

See legend after all supplementary figures

Supplementary 4



A Zheng et al.

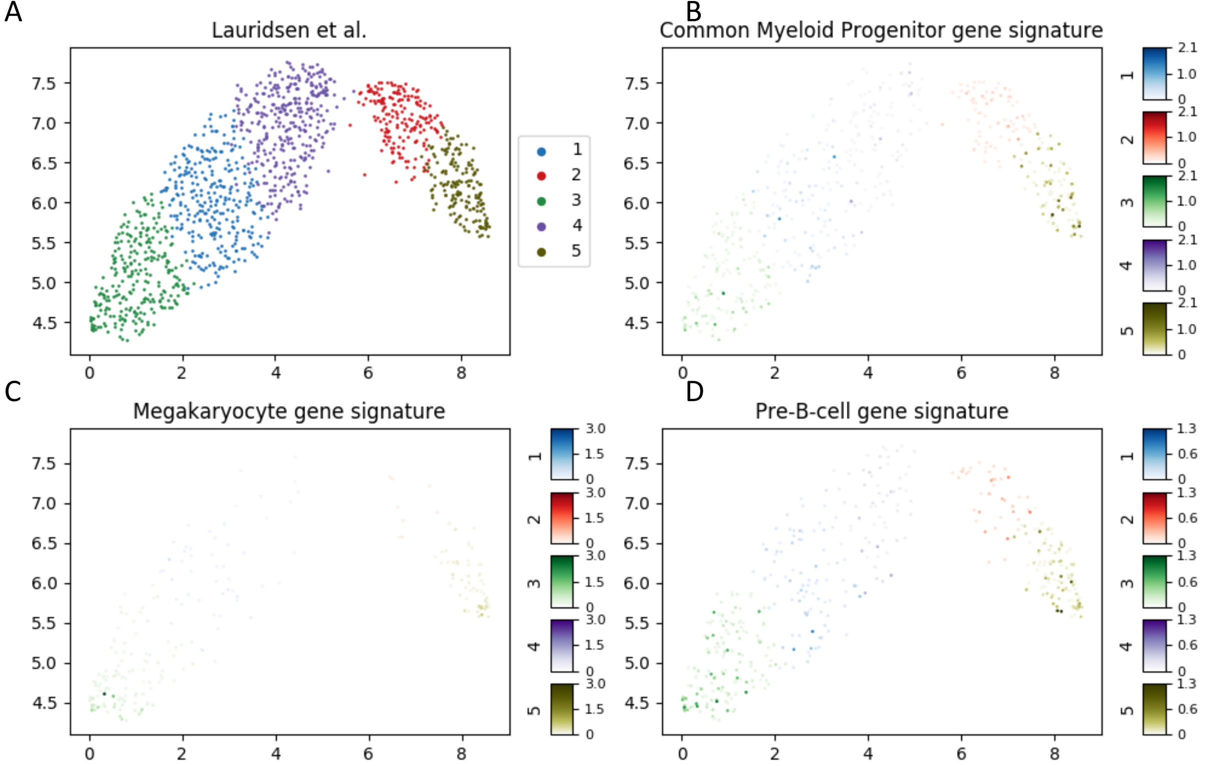
B Common Myeloid Progenitor gene signature

C Megakaryocyte gene signature

D Pre-B-cell gene signature

See legend after all supplementary figures

Supplementary 5



Supplementary 2-5. UMAP embeddings of the expression levels of the cells from 4 studies visualized on two dimensions. A. Plot title denote the name of the dataset. All cells are visualized, colour corresponds to the type, as can be seen on legend. B-D. The intensity of the expression levels of cells is computed as the mean of an expression score function across all genes of the signatures Common Myeloid Progenitor (B.), Megakaryocyte (C.), and Pre-B-cell (D.). As it is shown in the colour bar, more intense colour corresponds to higher expression levels. Colour intensities are logarithm of the expression multiplied by expression (xlogx) and was chosen for visualization of expression, to help differentiate between regions with different expression levels.