**Text S1**

**Algorithm Pseudo code**

A matlab program of this code will be provided by the authors upon request.

**Input: (A)** mixed matrix and reference signatures matrix with non-negative entries, where is the number of genes, is the number of samples and is the number of initial cell-types, **(B)** A collection of cell-type labels where is the label of the column in , **(C)** number of majority voting runs .

**For**

(1) Initialize: with non-negative values ,

Initialize: with ,

Scale columns of to sum to one.

(2) Get and using NMF, as described in Piper et al. 12:

where is used to avoid possible division by zero and refers to   
 the NMF iteration

(3) Determine (*v*) according to (4)

(4) Determine (*v*) and its cell-type labels as the chosen columns in

(5) Set if the label is chosen, otherwise

**End For**

(6) Determine the final cell-type identities:

, where if

(7) Determine the final : for

(8) Set to the number of columns in the matrix

(9) Determine according to Eq. (5)

**Output:** , ,

\* To use classes the algorithm requires the following sets , where is the number of classes and each contains the collection of labels that are affiliated with class . Note that there may be classes that contain a single label. Let be the label of class , then the algoritm outputs which is the estimated labels of classes as follows:

Change (6) such that:

, where if

Change (7) to:

Determine the final : for