Figure S22

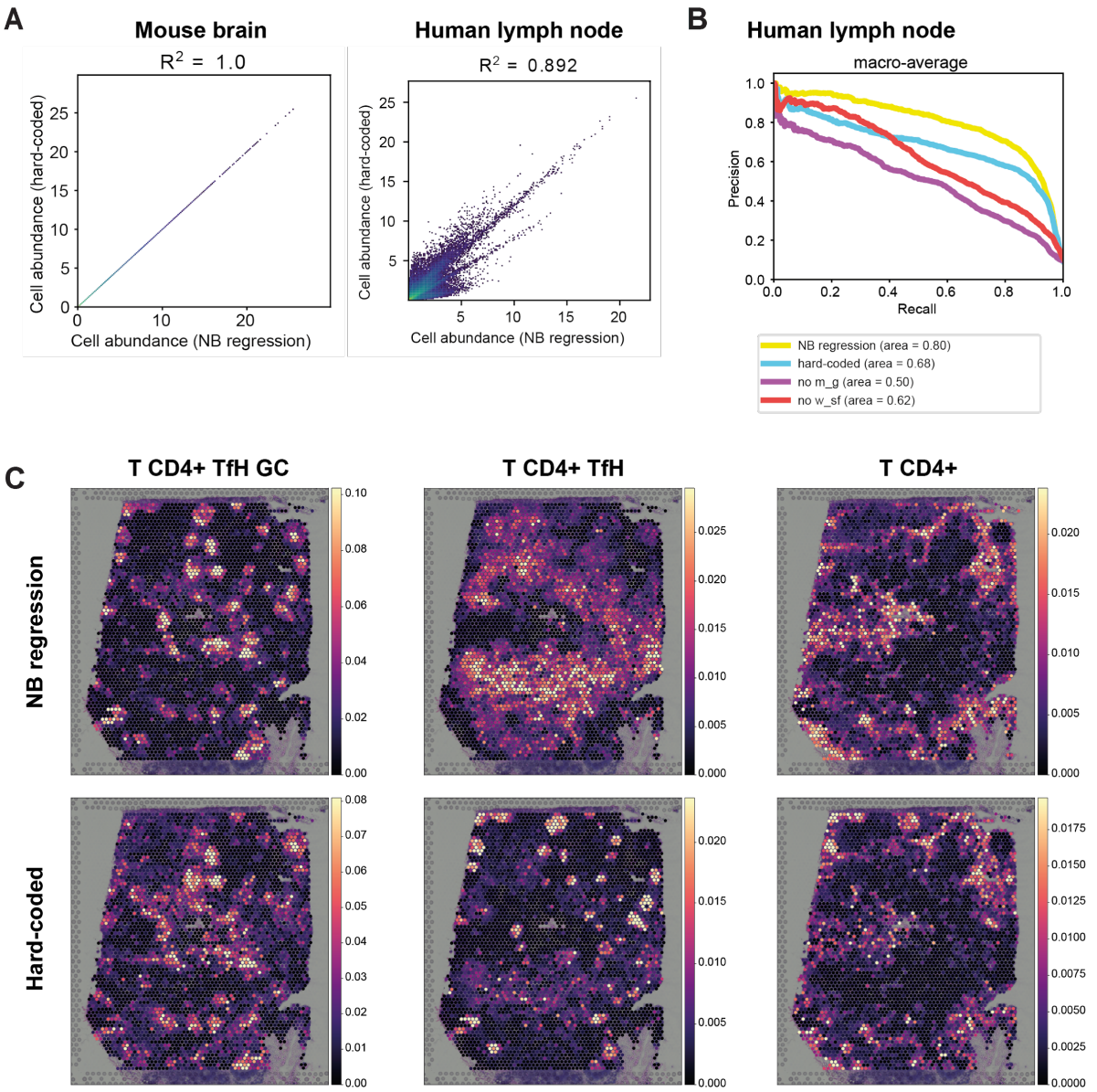

Figure S22. Comparison of alternative methods for estimating reference signatures of cell-types.

Considered are 1) the Negative Binomial (NB) regression model, which accounts for batch and technology effects, and 2) a hard-coded method that computes average count of each gene in each reference cluster (see Suppl. Methods for details).

A. Consistency of cell abundance estimates when using cell2location with signatures derived using the 2 considered methods (X-, Y-axis), pooled across all cell types and locations. Left: mouse brain mapped using paired single nucleus RNA-seq reference of 59 cell types. Right: human lymph nodes mapped using single cell RNA-seq reference composed of multiple batches and 34 cell types (3 organs, 2 technologies). 2D histogram counts (colour) is shown and R2 denotes Pearson correlation.

B. Assessment of the accuracy of alternative variants of cell2location (Suppl. Methods). Considered is the full cell2location model in conjunction with two alternative approaches to estimate reference signatures (NB regression and hard-coded).

Additionally, we considered a simplified version of cell2location without specific features: prior factorisation of cell abundance w_sf and gene-specific technology scaling m_g (used in conjunction with the NB regression model, Suppl. Methods). The assessment was performed using the same procedure as shown in Fig 4F-I and Fig S21.

C. Spatial plots (X, Y-axis) display estimated relative cell abundance (colour intensity) of 3 cell populations (columns) for 2 considered reference expression signature estimation methods (rows). T_CD4+_TfH_GC is expected to be present in the GC zone whereas T_CD4+_TfH and T_CD4+ are expected to be excluded from GC, especially dark zone GC (see Fig 4E).