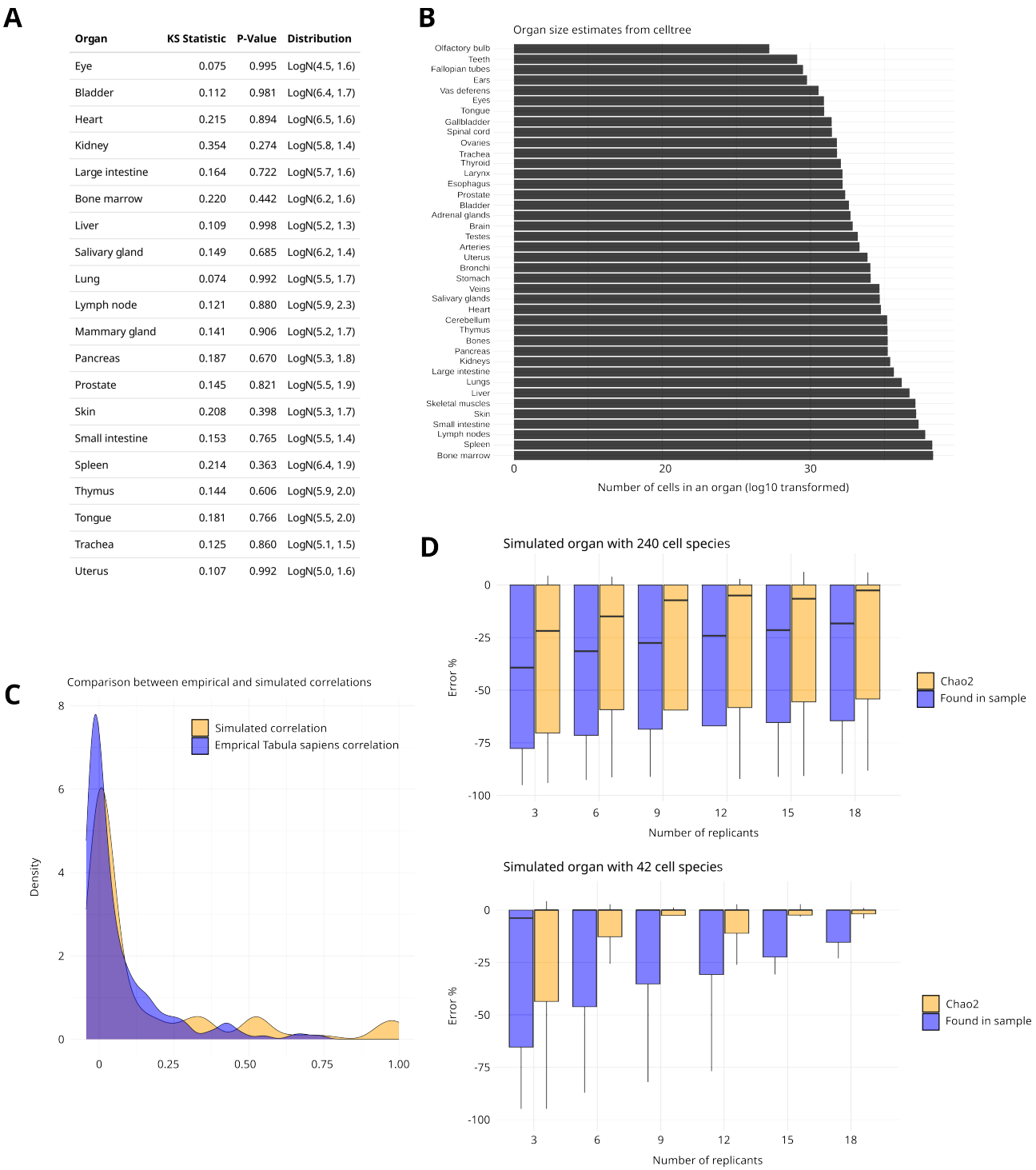
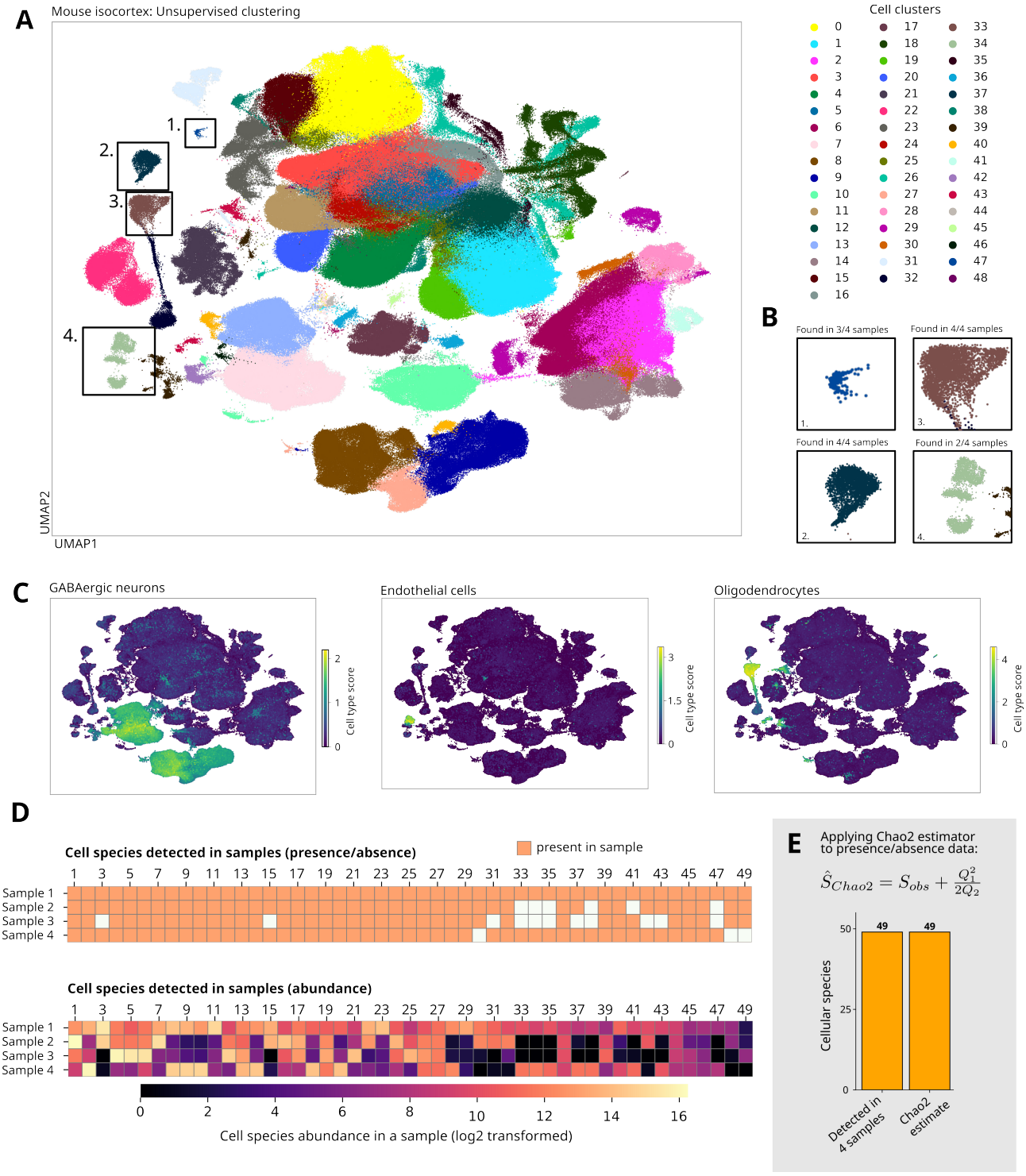
A

Organ	KS Statistic	P-Value	Distribution
Eye	0.075	0.995	LogN(4.5, 1.6)
Bladder	0.112	0.981	LogN(6.4, 1.7)
Heart	0.215	0.894	LogN(6.5, 1.6)
Kidney	0.354	0.274	LogN(5.8, 1.4)
Large intestine	0.164	0.722	LogN(5.7, 1.6)
Bone marrow	0.220	0.442	LogN(6.2, 1.6)
Liver	0.109	0.998	LogN(5.2, 1.3)
Salivary gland	0.149	0.685	LogN(6.2, 1.4)
Lung	0.074	0.992	LogN(5.5, 1.7)
Lymph node	0.121	0.880	LogN(5.9, 2.3)
Mammary gland	0.141	0.906	LogN(5.2, 1.7)
Pancreas	0.187	0.670	LogN(5.3, 1.8)
Prostate	0.145	0.821	LogN(5.5, 1.9)
Skin	0.208	0.398	LogN(5.3, 1.7)
Small intestine	0.153	0.765	LogN(5.5, 1.4)
Spleen	0.214	0.363	LogN(6.4, 1.9)
Thymus	0.144	0.606	LogN(5.9, 2.0)
Tongue	0.181	0.766	LogN(5.5, 2.0)
Trachea	0.125	0.860	LogN(5.1, 1.5)
Uterus	0.107	0.992	LogN(5.0, 1.6)



Supplementary Figure 1. (**A**) Goodness-of-fit testing whether cell type abundances from different organs are log-normally distributed. The table reports test statistics from Kolmogorov-Smirnov (KS) test and the corresponding p-values. Note: high p-values indicate insufficient evidence to reject the null hypothesis of log-normal distribution. (**B**) The estimates of total cell counts in each organ (from https://humancelltreemap.mis.mpg.de/). (**C**) Comparison of empirical correlations between Tabula Sapiens data and simulated human organs. KS test for distribution similarity, with p = 0.42, indicating similar distributions. (**D**) Effect of the number of sample replicates on the Chao2 estimation accuracy. The horizontal lines indicate median error levels from -100% (total

underestimation) to 0% (no error), boxes the interquartile ranges, and the vertical lines indicate the number of samples.



Supplementary Figure 2. Completeness estimation of a large single-cell mouse isocortex dataset. (**A**) one million cell dataset with 4 individual samples analyzed with 10x chromium V2 chemistry of the mouse isocortex from Yao et al., 2023. Leiden clusters are shown. Colors indicate different clusters (49 distinct clusters detected across 4 samples). (**B**) Smaller clusters labelled in panel A within the integrated dataset are detected in multiple samples indicating high dataset

completeness. (**C**) Individual clusters identified with RNA markers and matched to common cell types found in mouse isocortex. The cell type score is calculated by averaging gene expression values of the marker genes separately for each cell type (see Methods). (**D**) Cellular species detection in the four isocortex samples (upper panel: presence/absence and lower panel: abundance data). Cellular species correspond here to the detected cell clusters. (**E**) Chao2 estimate of the cellular richness of mouse isocortex dataset using data from Yao et al. In the Chao2 Equation, Sobs is the observed cellular species count, Q1 is the number of species observed in exactly one sample, and Q2 is the number of species observed in exactly two samples.