



Fig. 5 | scTour reconstructs the transcriptomic space of unobserved time interval in pancreatic endocrinogenesis. **a**, Schematic depicting the scTour model training with the Ngn3^{high} EPs excluded, followed by prediction of the latent space of these cells given their expected developmental time. **b**, UMAP visualization based on the reconstructed latent representations for the held-out Ngn3^{high} EPs (red outline) and those inferred from training cells. Cells are coloured by their developmental pseudotime and cell type identities (top right). **c**, UMAP visualization based on the latent representations of reconstructed Ngn3^{high} EPs (red outline), true Ngn3^{high} EPs (blue outline), and the remaining cells. Cells are coloured by cell types. **d**, Box plot displaying the Euclidean distances calculated between the reconstructed latent representations for Ngn3^{high} EPs and those from each of the cellular states (i.e., true Ngn3^{high} EPs and the remaining states), with the medians, interquartile ranges, and 5th, 95th percentiles indicated by centre lines, hinges, and whiskers, respectively. **e**, Unsupervised hierarchical clustering of the reconstructed Ngn3^{high} EPs along with all the other cells based on their Euclidean distances in the scTour latent space. Column colours of the heatmap mark the cell types and row colours denote the reconstructed (red), true (blue) Ngn3^{high} EPs, and remaining cells (light grey). The colour gradient of the heatmap indicates the Euclidean distance.