

Fig. S1| Preprocessing of multi-species training data and statistics of prior knowledge. a, Preprocessing flowchart for scCompass-126M. **b,** Illustration of diversity from single-cell perturbations, cancer cell identification, gender distribution, and cell differentiation time for humans (first row) and mice (second row). **c,** Number of homologous, human-specific, and mouse-specific genes in the token dictionary. **d,** Histogram of the Pearson correlation coefficient for all gene pairs in humans (left) and mice (right). **e,** Histogram of gene numbers in each gene family for humans (left) and mice (right). **f,** Multi-tissue developmental stage samples in mouse ENCODE and the number of PECA2 TF-TG edges in each mouse GRN.