nature machine intelligence

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s42256-020-0160-y

In the format provided by the authors and unedited.

Generative molecular design in low data regimes

Michael Moret[1], Lukas Friedrich[1], Francesca Grisoni[1], Daniel Merk[1,2] and Gisbert Schneider[1]✉

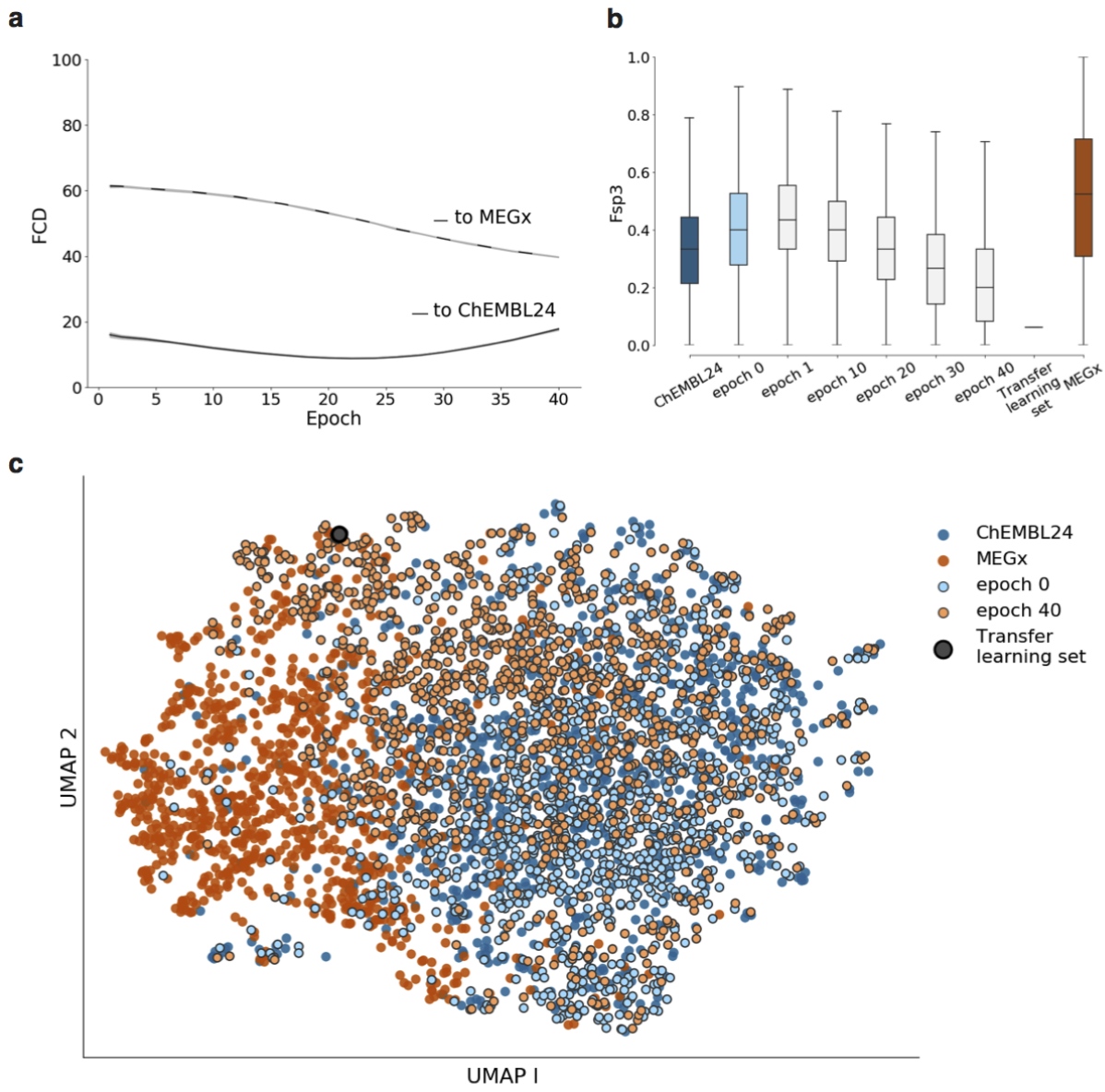
[1]Department of Chemistry and Applied Biosciences, RETHINK, ETH Zurich, Zurich, Switzerland. [2]Goethe University Frankfurt, Institute of Pharmaceutical Chemistry, Frankfurt, Germany. ✉e-mail: gisbert@ethz.ch

Supplementary Table

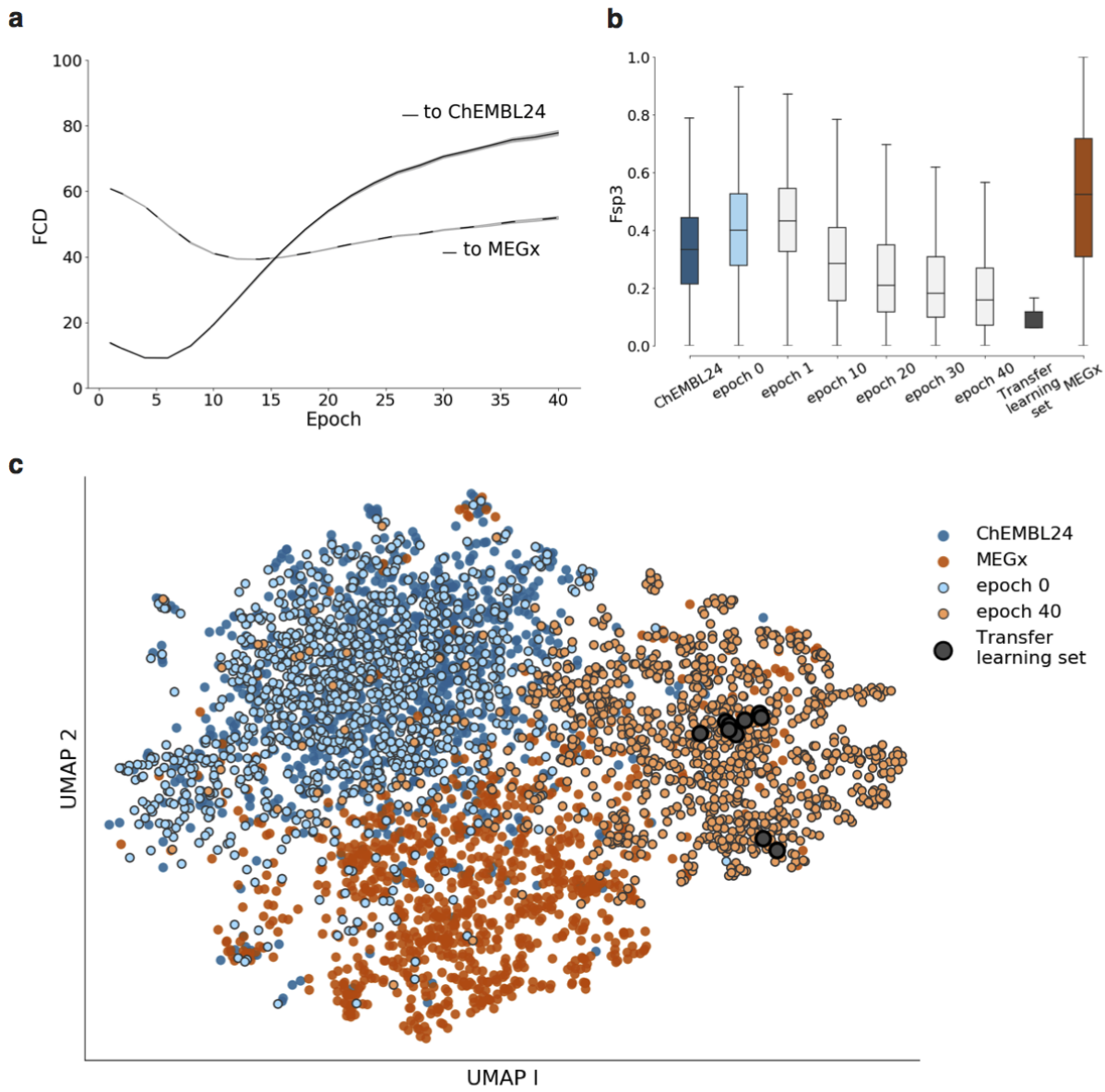
Supplementary Table 1. Comparison of the number of novel molecules generated during transfer learning when keeping the first layer of the chemical language model constant (i.e. frozen layer) or without keeping them constant (i.e. fine-tuning all weights). 10,000 molecules were sampled at each epoch.

Source	Novel molecules with first layer frozen	Novel molecules without first layer frozen
Transfer learning with five similar compounds		
Epoch 1	9035	9337
Epoch 20	6543	461
Epoch 40	3373	181
Transfer learning with five dissimilar compounds		
Epoch 1	9459	9391
Epoch 20	8702	6244
Epoch 40	8184	4383

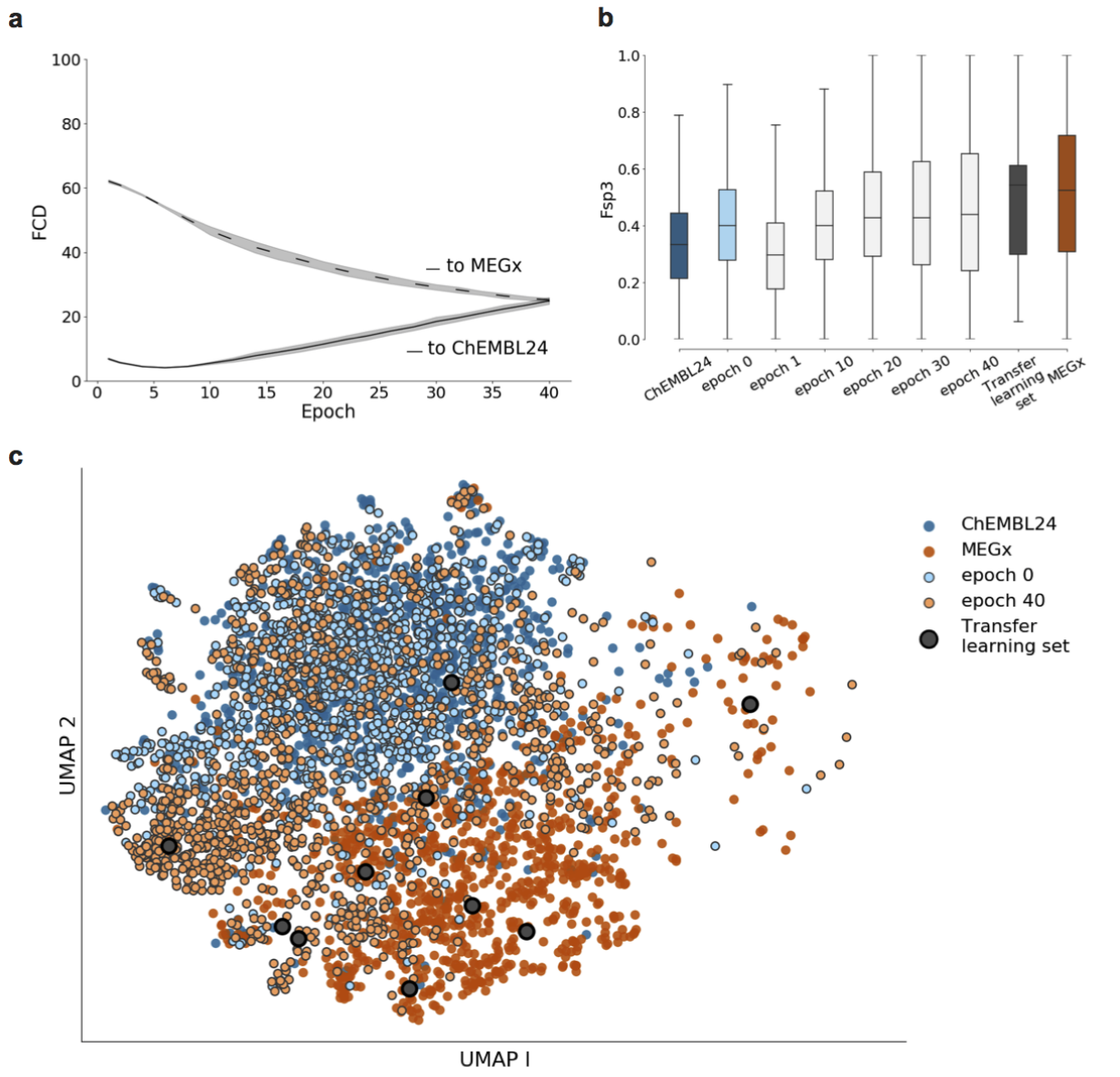
Supplementary Figures



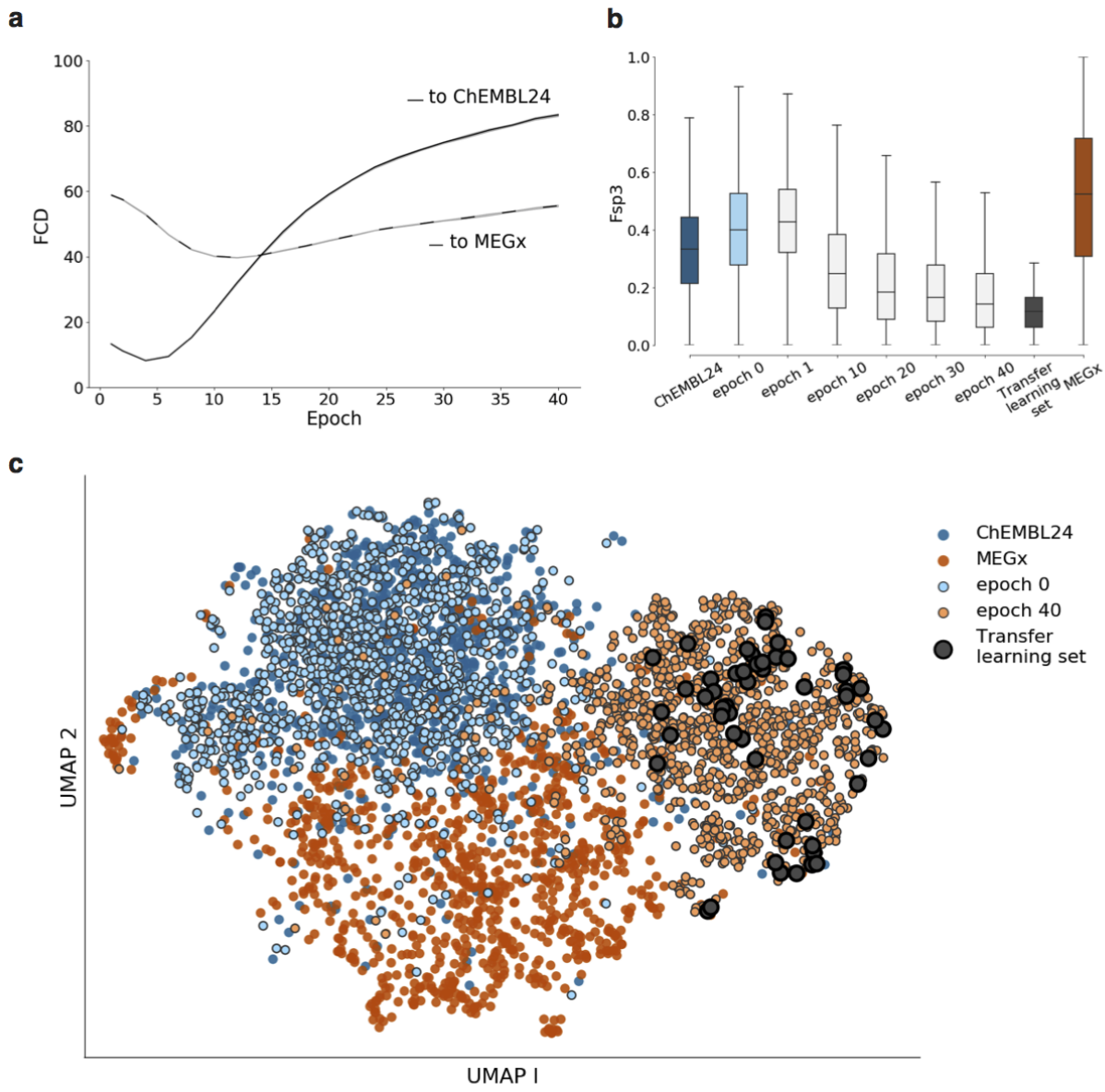
Supplementary Figure 1. Chemical space navigation by transfer learning with one molecule. **a,** Fréchet ChemNet Distance (FCD) to ChEMBL24 and MEGx of generated molecules during chemical space navigation. Mean and 0.95 confidence interval for ten repeats are shown in shaded area. **b,** Evolution of the fraction of sp^3-hybridized carbon atoms (Fsp3) during the chemical space navigation. **c,** UMAP plot of molecules. For each group, 1k molecules were randomly selected. Dark blue: ChEMBL24. Dark orange: MEGx. Light blue: molecules generated from the pretrained model (*i.e.*, epoch zero). Light orange: molecules generated at epoch 40. Gray circles: transfer learning set.



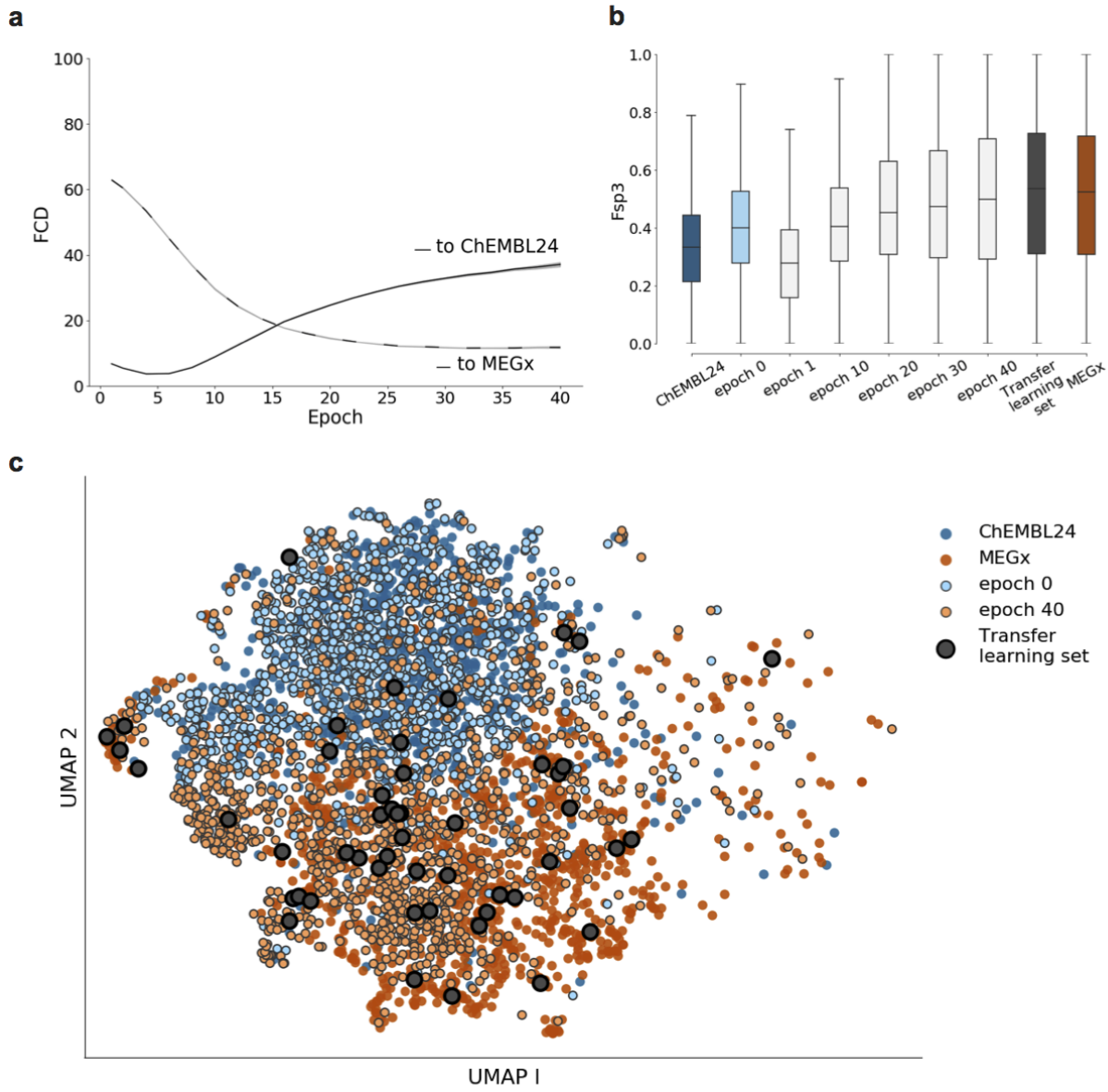
Supplementary Figure 2. Chemical space navigation by transfer learning with 10 similar molecules. **a,** Fréchet ChemNet Distance (FCD) to ChEMBL24 and MEGx of generated molecules during chemical space navigation. Mean and 0.95 confidence interval for ten repeats are shown in shaded area. **b,** Evolution of the fraction of sp3-hybridized carbon atoms (Fsp3) during the chemical space navigation. **c,** UMAP plot of molecules. For each group, 1k molecules were randomly selected. Dark blue: ChEMBL24. Dark orange: MEGx. Light blue: molecules generated from the pretrained model (*i.e.*, epoch zero). Light orange: molecules generated at epoch 40. Gray circles: transfer learning set.



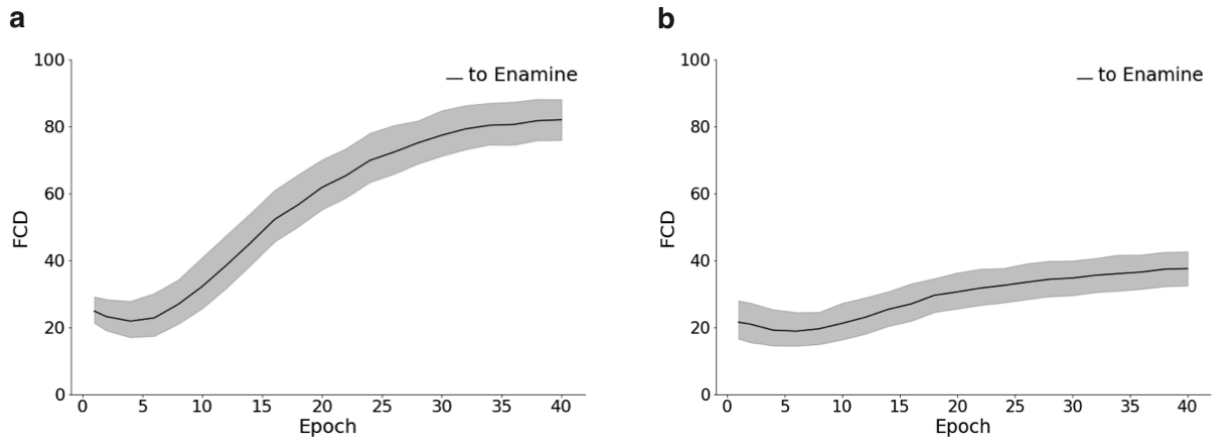
Supplementary Figure 3. Chemical space navigation by transfer learning with 10 dissimilar molecules. **a,** Fréchet ChemNet Distance (FCD) to ChEMBL24 and MEGx of generated molecules during chemical space navigation. Mean and 0.95 confidence interval for ten repeats are shown in shaded area. **b,** Evolution of the fraction of sp^3-hybridized carbon atoms (Fsp3) during the chemical space navigation. **c,** UMAP plot of molecules. For each group, 1k molecules were randomly selected. Dark blue: ChEMBL24. Dark orange: MEGx. Light blue: molecules generated from the pretrained model (*i.e.*, epoch zero). Light orange: molecules generated at epoch 40. Gray circles: transfer learning set.



Supplementary Figure 4. Chemical space navigation by transfer learning with 50 similar molecules. **a,** Fréchet ChemNet Distance (FCD) to ChEMBL24 and MEGx of generated molecules during chemical space navigation. Mean and 0.95 confidence interval for ten repeats are shown in shaded area. **b,** Evolution of the fraction of sp3-hybridized carbon atoms (Fsp3) during the chemical space navigation. **c,** UMAP plot of molecules. For each group, 1k molecules were randomly selected. Dark blue: ChEMBL24. Dark orange: MEGx. Light blue: molecules generated from the pretrained model (*i.e.*, epoch zero). Light orange: molecules generated at epoch 40. Gray circles: transfer learning set.



Supplementary Figure 5. Chemical space navigation by transfer learning with 50 dissimilar molecules. **a,** Fréchet ChemNet Distance (FCD) to ChEMBL24 and MEGx of generated molecules during chemical space navigation. Mean and 0.95 confidence interval for ten repeats are shown in shaded area. **b,** Evolution of the fraction of sp^3-hybridized carbon atoms (Fsp3) during the chemical space navigation. **c,** UMAP plot of molecules. For each group, 1k molecules were randomly selected. Dark blue: ChEMBL24. Dark orange: MEGx. Light blue: molecules generated from the pretrained model (*i.e.*, epoch zero). Light orange: molecules generated at epoch 40. Gray circles: transfer learning set.



Supplementary Figure 6. Fréchet ChemNet Distance (FCD) of generated molecules to the Enamine compound collection. **a,** FCD of five similar molecules; **b,** FCD of five dissimilar molecules. The mean values and corresponding 0.95 confidence intervals (shaded area) are shown ($N = 5$; independent random subsets).