

Supplementary Table 1. Default hyper-parameter settings

Hyperparameter		Value
Token adapter	Fully connected layer 1 input dimension	512
	Fully connected layer 1 output dimension	128
	Fully connected layer 2 input dimension	128
	Fully connected layer 2 output dimension	512
	Activation function	GeLU
	Residual connection	True
Encoder adapter	Fully connected layer 1 input dimension	512
	Fully connected layer 1 output dimension	128
	Fully connected layer 2 input dimension	128
	Fully connected layer 2 output dimension	512
	Activation function	GeLU
	Residual connection	True
Adapter configuration to transformer layers		first 6 transformer blocks
Prefix adapter	Number of prefix tokens	64
	Initial value of prefix token	0
	Adapter configuration to transformer layers	first 6 transformer blocks
LoRA	LoRA rank	[8, 16, 32, 128, 256]
	Lora Alpha	[1, 16, 32, 128, 256, 512]
	LoRA input dimension	512
	LoRA output dimension	512
	Use LoRA on Q, K, and V, respectively	[True, False, True]
	Adapter configuration to transformer layers	first 6 transformer blocks
Training	Optimizer	Adam
	Optimizer momentum	$\beta_1 = 0.9, \beta_2 = 0.999$
	Maximum learning rate	1e-5
	Ratio of epochs for learning rate schedule	0.9
	Early stop patience	5
	Maximum epoch	50
	Batch size	32
	Automatic mixed precision	True
scBERT	Default adapter	Prefix adapter
Geneformer		Prefix adapter
scGPT		Encoder adapter