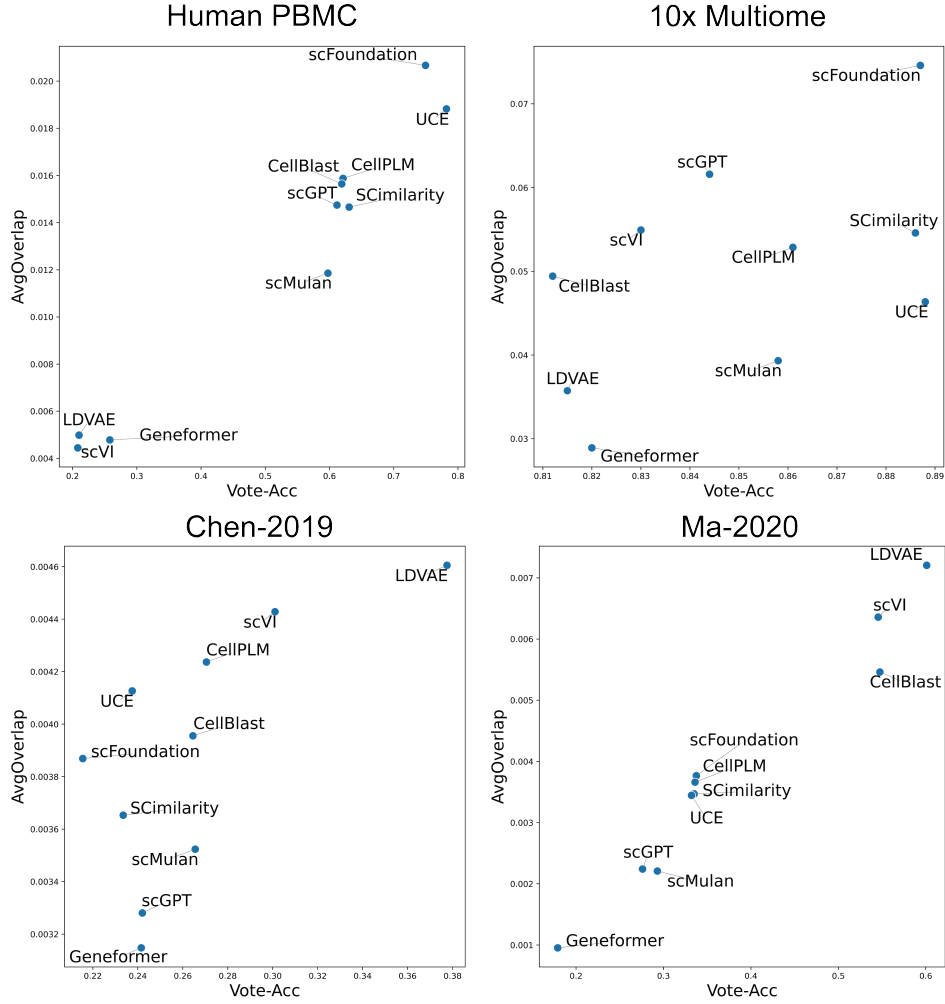


A APPENDIX

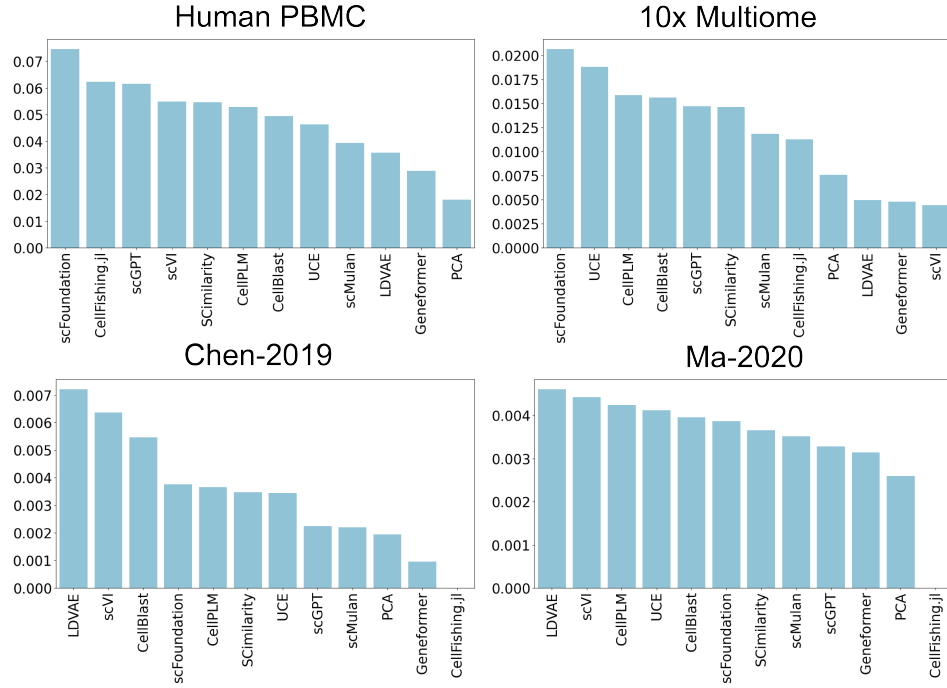
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755



Appendix Figure 1: Correlation between *Vote-Acc* and *AvgOverlap* on four benchmarking datasets

Appendix Table 1: Evaluating single-cell FMs in cross-platform retrieval setting on human pancreas dataset (*Vote-Acc* and batch diversity). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query

| K              | 1        |                 | 5        |                 | 10       |                 | 20       |                 | 50       |                 | 100      |                 |
|----------------|----------|-----------------|----------|-----------------|----------|-----------------|----------|-----------------|----------|-----------------|----------|-----------------|
| Metric         | Vote Acc | Batch Diversity | Vote Acc | Batch Diversity | Vote Acc | Batch Diversity | Vote Acc | Batch Diversity | Vote Acc | Batch Diversity | Vote Acc | Batch Diversity |
| PCA            | 0.838    | 0.000           | 0.870    | 0.079           | 0.888    | 0.102           | 0.880    | 0.134           | 0.848    | 0.159           | 0.820    | 0.173           |
| CellFishing.jl | 0.950    | 0.000           | 0.952    | 0.165           | 0.954    | 0.220           | 0.948    | 0.279           | 0.934    | 0.369           | 0.925    | 0.457           |
| scVI           | 0.901    | 0.000           | 0.920    | 0.078           | 0.918    | 0.106           | 0.910    | 0.130           | 0.898    | 0.159           | 0.890    | 0.177           |
| LDVAE          | 0.824    | 0.000           | 0.842    | 0.061           | 0.839    | 0.077           | 0.835    | 0.095           | 0.825    | 0.124           | 0.815    | 0.153           |
| CellBlast      | 0.218    | 0.000           | 0.235    | 0.115           | 0.242    | 0.159           | 0.265    | 0.184           | 0.302    | 0.198           | 0.328    | 0.196           |
| scFoundation   | 0.799    | 0.000           | 0.823    | 0.064           | 0.829    | 0.101           | 0.830    | 0.133           | 0.823    | 0.170           | 0.814    | 0.203           |
| scGPT          | 0.736    | 0.000           | 0.776    | 0.319           | 0.785    | 0.387           | 0.807    | 0.433           | 0.812    | 0.483           | 0.804    | 0.513           |
| SCimilarity    | 0.913    | 0.000           | 0.934    | 0.224           | 0.934    | 0.278           | 0.931    | 0.331           | 0.927    | 0.399           | 0.923    | 0.471           |
| UCE            | 0.670    | 0.000           | 0.707    | 0.232           | 0.708    | 0.285           | 0.703    | 0.326           | 0.710    | 0.387           | 0.735    | 0.437           |
| Geneformer     | 0.481    | 0.000           | 0.518    | 0.112           | 0.521    | 0.133           | 0.498    | 0.146           | 0.451    | 0.182           | 0.412    | 0.219           |
| scMulán        | 0.748    | 0.000           | 0.760    | 0.316           | 0.752    | 0.370           | 0.728    | 0.406           | 0.698    | 0.444           | 0.682    | 0.478           |
| CellPLM        | 0.594    | 0.000           | 0.585    | 0.249           | 0.581    | 0.314           | 0.585    | 0.389           | 0.579    | 0.427           | 0.564    | 0.447           |



Appendix Figure 2: AvgOverlap score on four benchmarking datasets (K=50)

Appendix Table 2: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 1.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.822       | 0.752         | 0.782         | 0.548       | 0.350        | 0.591        | 0.472        | 0.213          | 0.475   | 0.556   |
| CellFishing.jl | 0.875       | 0.871         | 0.888         | 0.854       | 0.760        | 0.811        | 0.710        | 0.823          | 0.719   | 0.812   |
| scVI           | 0.877       | 0.855         | 0.873         | 0.891       | 0.599        | 0.746        | 0.665        | 0.791          | 0.706   | 0.778   |
| LDVAE          | 0.869       | 0.811         | 0.818         | 0.792       | 0.684        | 0.765        | 0.586        | 0.738          | 0.726   | 0.754   |
| CellBlast      | 0.818       | 0.831         | 0.858         | 0.792       | 0.757        | 0.677        | 0.630        | 0.778          | 0.714   | 0.762   |
| scFoundation   | 0.899       | 0.890         | 0.895         | 0.905       | 0.802        | 0.773        | 0.782        | 0.869          | 0.823   | 0.849   |
| scGPT          | 0.913       | 0.888         | 0.887         | 0.893       | 0.728        | 0.739        | 0.716        | 0.722          | 0.736   | 0.803   |
| SCimilarity    | 0.920       | 0.886         | 0.890         | 0.913       | 0.823        | 0.818        | 0.743        | 0.848          | 0.807   | 0.850   |
| UCE            | 0.914       | 0.887         | 0.896         | 0.895       | 0.800        | 0.825        | 0.760        | 0.871          | 0.823   | 0.852   |
| Geneformer     | 0.837       | 0.827         | 0.833         | 0.711       | 0.705        | 0.767        | 0.704        | 0.749          | 0.781   | 0.768   |
| scMulan        | 0.894       | 0.882         | 0.885         | 0.882       | 0.738        | 0.796        | 0.712        | 0.823          | 0.786   | 0.822   |
| CellPLM        | 0.892       | 0.885         | 0.876         | 0.886       | 0.667        | 0.770        | 0.755        | 0.719          | 0.807   | 0.806   |

Appendix Table 3: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 5.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.793       | 0.732         | 0.758         | 0.564       | 0.317        | 0.570        | 0.446        | 0.216          | 0.469   | 0.541   |
| CellFishing.jl | 0.875       | 0.872         | 0.887         | 0.852       | 0.759        | 0.808        | 0.708        | 0.831          | 0.704   | 0.811   |
| scVI           | 0.891       | 0.855         | 0.862         | 0.886       | 0.611        | 0.748        | 0.680        | 0.783          | 0.709   | 0.781   |
| LDVAE          | 0.858       | 0.807         | 0.815         | 0.791       | 0.688        | 0.748        | 0.613        | 0.745          | 0.731   | 0.755   |
| CellBlast      | 0.819       | 0.829         | 0.847         | 0.791       | 0.742        | 0.695        | 0.612        | 0.740          | 0.706   | 0.753   |
| scFoundation   | 0.910       | 0.894         | 0.896         | 0.903       | 0.787        | 0.771        | 0.781        | 0.855          | 0.830   | 0.848   |
| scGPT          | 0.916       | 0.887         | 0.881         | 0.884       | 0.702        | 0.731        | 0.713        | 0.723          | 0.727   | 0.796   |
| SCimilarity    | 0.919       | 0.895         | 0.883         | 0.906       | 0.804        | 0.821        | 0.755        | 0.854          | 0.817   | 0.850   |
| UCE            | 0.913       | 0.887         | 0.888         | 0.889       | 0.784        | 0.830        | 0.747        | 0.851          | 0.825   | 0.846   |
| Geneformer     | 0.825       | 0.819         | 0.815         | 0.704       | 0.683        | 0.759        | 0.705        | 0.724          | 0.771   | 0.756   |
| scMulan        | 0.891       | 0.877         | 0.878         | 0.867       | 0.730        | 0.791        | 0.709        | 0.833          | 0.787   | 0.818   |
| CellPLM        | 0.901       | 0.883         | 0.875         | 0.883       | 0.673        | 0.763        | 0.744        | 0.666          | 0.813   | 0.800   |

Appendix Table 4: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 10.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.776       | 0.717         | 0.745         | 0.567       | 0.310        | 0.557        | 0.431        | 0.227          | 0.462   | 0.532   |
| CellFishing.jl | 0.876       | 0.873         | 0.885         | 0.852       | 0.751        | 0.808        | 0.709        | 0.829          | 0.699   | 0.809   |
| scVI           | 0.895       | 0.855         | 0.857         | 0.884       | 0.619        | 0.748        | 0.682        | 0.791          | 0.710   | 0.782   |
| LDVAE          | 0.856       | 0.806         | 0.809         | 0.785       | 0.675        | 0.742        | 0.621        | 0.742          | 0.733   | 0.752   |
| CellBlast      | 0.820       | 0.828         | 0.839         | 0.789       | 0.735        | 0.701        | 0.609        | 0.738          | 0.698   | 0.751   |
| scFoundation   | 0.913       | 0.892         | 0.894         | 0.899       | 0.789        | 0.775        | 0.778        | 0.853          | 0.830   | 0.847   |
| scGPT          | 0.914       | 0.887         | 0.878         | 0.879       | 0.703        | 0.730        | 0.711        | 0.725          | 0.720   | 0.794   |
| SCimilarity    | 0.918       | 0.897         | 0.879         | 0.903       | 0.799        | 0.824        | 0.754        | 0.857          | 0.816   | 0.850   |
| UCE            | 0.912       | 0.886         | 0.886         | 0.888       | 0.780        | 0.830        | 0.746        | 0.846          | 0.826   | 0.845   |
| Geneformer     | 0.819       | 0.814         | 0.806         | 0.702       | 0.670        | 0.756        | 0.700        | 0.717          | 0.767   | 0.750   |
| scMulan        | 0.891       | 0.875         | 0.874         | 0.857       | 0.733        | 0.787        | 0.703        | 0.822          | 0.785   | 0.814   |
| CellPLM        | 0.903       | 0.882         | 0.871         | 0.882       | 0.679        | 0.763        | 0.740        | 0.667          | 0.816   | 0.800   |

Appendix Table 5: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 20.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.755       | 0.698         | 0.727         | 0.569       | 0.304        | 0.539        | 0.409        | 0.234          | 0.450   | 0.521   |
| CellFishing.jl | 0.877       | 0.873         | 0.883         | 0.853       | 0.748        | 0.807        | 0.711        | 0.828          | 0.692   | 0.808   |
| scVI           | 0.898       | 0.852         | 0.851         | 0.879       | 0.621        | 0.745        | 0.687        | 0.789          | 0.709   | 0.781   |
| LDVAE          | 0.853       | 0.805         | 0.805         | 0.781       | 0.677        | 0.738        | 0.629        | 0.737          | 0.733   | 0.751   |
| CellBlast      | 0.818       | 0.825         | 0.831         | 0.785       | 0.731        | 0.705        | 0.604        | 0.737          | 0.688   | 0.747   |
| scFoundation   | 0.912       | 0.891         | 0.890         | 0.898       | 0.789        | 0.778        | 0.775        | 0.852          | 0.831   | 0.846   |
| scGPT          | 0.912       | 0.884         | 0.875         | 0.877       | 0.712        | 0.728        | 0.706        | 0.708          | 0.711   | 0.790   |
| SCimilarity    | 0.916       | 0.898         | 0.876         | 0.899       | 0.804        | 0.825        | 0.754        | 0.856          | 0.815   | 0.849   |
| UCE            | 0.912       | 0.886         | 0.883         | 0.884       | 0.774        | 0.830        | 0.743        | 0.838          | 0.827   | 0.842   |
| Geneformer     | 0.811       | 0.808         | 0.795         | 0.693       | 0.651        | 0.750        | 0.696        | 0.701          | 0.763   | 0.741   |
| scMulan        | 0.891       | 0.873         | 0.869         | 0.850       | 0.731        | 0.782        | 0.698        | 0.818          | 0.783   | 0.811   |
| CellPLM        | 0.903       | 0.880         | 0.865         | 0.878       | 0.678        | 0.763        | 0.733        | 0.678          | 0.816   | 0.799   |

Appendix Table 6: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 50.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.712       | 0.668         | 0.692         | 0.569       | 0.287        | 0.507        | 0.379        | 0.244          | 0.430   | 0.499   |
| CellFishing.jl | 0.879       | 0.870         | 0.875         | 0.851       | 0.743        | 0.804        | 0.710        | 0.808          | 0.680   | 0.802   |
| scVI           | 0.893       | 0.845         | 0.837         | 0.869       | 0.608        | 0.738        | 0.691        | 0.770          | 0.706   | 0.773   |
| LDVAE          | 0.843       | 0.796         | 0.796         | 0.771       | 0.667        | 0.733        | 0.640        | 0.723          | 0.733   | 0.745   |
| CellBlast      | 0.811       | 0.816         | 0.814         | 0.780       | 0.717        | 0.703        | 0.596        | 0.729          | 0.670   | 0.737   |
| scFoundation   | 0.909       | 0.890         | 0.881         | 0.895       | 0.788        | 0.784        | 0.772        | 0.844          | 0.831   | 0.844   |
| scGPT          | 0.907       | 0.878         | 0.868         | 0.871       | 0.712        | 0.722        | 0.696        | 0.688          | 0.699   | 0.782   |
| SCimilarity    | 0.914       | 0.896         | 0.869         | 0.890       | 0.806        | 0.826        | 0.752        | 0.850          | 0.813   | 0.846   |
| UCE            | 0.909       | 0.885         | 0.878         | 0.876       | 0.764        | 0.829        | 0.739        | 0.828          | 0.828   | 0.837   |
| Geneformer     | 0.796       | 0.795         | 0.776         | 0.681       | 0.625        | 0.741        | 0.690        | 0.666          | 0.758   | 0.725   |
| scMulan        | 0.888       | 0.869         | 0.860         | 0.843       | 0.718        | 0.773        | 0.689        | 0.803          | 0.776   | 0.802   |
| CellPLM        | 0.900       | 0.876         | 0.853         | 0.872       | 0.673        | 0.763        | 0.724        | 0.692          | 0.812   | 0.796   |

Appendix Table 7: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc ). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 100.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.660       | 0.637         | 0.657         | 0.559       | 0.277        | 0.476        | 0.349        | 0.248          | 0.411   | 0.475   |
| CellFishing.jl | 0.877       | 0.867         | 0.864         | 0.848       | 0.729        | 0.798        | 0.704        | 0.785          | 0.666   | 0.793   |
| scVI           | 0.884       | 0.835         | 0.818         | 0.859       | 0.558        | 0.731        | 0.688        | 0.740          | 0.701   | 0.757   |
| LDVAE          | 0.834       | 0.785         | 0.786         | 0.763       | 0.658        | 0.727        | 0.650        | 0.706          | 0.731   | 0.738   |
| CellBlast      | 0.801       | 0.805         | 0.799         | 0.772       | 0.708        | 0.699        | 0.590        | 0.723          | 0.652   | 0.728   |
| scFoundation   | 0.904       | 0.889         | 0.869         | 0.893       | 0.784        | 0.787        | 0.767        | 0.834          | 0.830   | 0.840   |
| scGPT          | 0.897       | 0.867         | 0.857         | 0.867       | 0.706        | 0.713        | 0.687        | 0.658          | 0.688   | 0.771   |
| SCimilarity    | 0.913       | 0.895         | 0.863         | 0.883       | 0.807        | 0.825        | 0.751        | 0.845          | 0.810   | 0.844   |
| UCE            | 0.904       | 0.882         | 0.873         | 0.870       | 0.754        | 0.826        | 0.735        | 0.820          | 0.828   | 0.833   |
| Geneformer     | 0.784       | 0.782         | 0.759         | 0.670       | 0.597        | 0.731        | 0.685        | 0.626          | 0.752   | 0.710   |
| scMulan        | 0.880       | 0.860         | 0.849         | 0.837       | 0.704        | 0.759        | 0.680        | 0.793          | 0.765   | 0.792   |
| CellPLM        | 0.892       | 0.870         | 0.840         | 0.862       | 0.669        | 0.759        | 0.714        | 0.702          | 0.804   | 0.790   |

Appendix Table 8: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc ). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 1.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.822       | 0.752         | 0.782         | 0.548       | 0.350        | 0.591        | 0.472        | 0.213          | 0.475   | 0.556   |
| CellFishing.jl | 0.875       | 0.871         | 0.888         | 0.854       | 0.760        | 0.811        | 0.710        | 0.823          | 0.719   | 0.812   |
| scVI           | 0.877       | 0.855         | 0.873         | 0.891       | 0.599        | 0.746        | 0.665        | 0.791          | 0.706   | 0.778   |
| LDVAE          | 0.869       | 0.811         | 0.818         | 0.792       | 0.684        | 0.765        | 0.586        | 0.738          | 0.726   | 0.754   |
| CellBlast      | 0.818       | 0.831         | 0.858         | 0.792       | 0.757        | 0.677        | 0.630        | 0.778          | 0.714   | 0.762   |
| scFoundation   | 0.899       | 0.890         | 0.895         | 0.905       | 0.802        | 0.773        | 0.782        | 0.869          | 0.823   | 0.849   |
| scGPT          | 0.913       | 0.888         | 0.887         | 0.893       | 0.728        | 0.739        | 0.716        | 0.722          | 0.736   | 0.803   |
| SCimilarity    | 0.920       | 0.886         | 0.890         | 0.913       | 0.823        | 0.818        | 0.743        | 0.848          | 0.807   | 0.850   |
| UCE            | 0.914       | 0.887         | 0.896         | 0.895       | 0.800        | 0.825        | 0.760        | 0.871          | 0.823   | 0.852   |
| Geneformer     | 0.837       | 0.827         | 0.833         | 0.711       | 0.705        | 0.767        | 0.704        | 0.749          | 0.781   | 0.768   |
| scMulan        | 0.894       | 0.882         | 0.885         | 0.882       | 0.738        | 0.796        | 0.712        | 0.823          | 0.786   | 0.822   |
| CellPLM        | 0.892       | 0.885         | 0.876         | 0.886       | 0.667        | 0.770        | 0.755        | 0.719          | 0.807   | 0.806   |

Appendix Table 9: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 5.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.833       | 0.782         | 0.804         | 0.660       | 0.375        | 0.628        | 0.484        | 0.234          | 0.524   | 0.592   |
| CellFishing.jl | 0.883       | 0.901         | 0.902         | 0.895       | 0.814        | 0.849        | 0.739        | 0.875          | 0.731   | 0.843   |
| scVI           | 0.916       | 0.895         | 0.873         | 0.904       | 0.641        | 0.784        | 0.719        | 0.814          | 0.744   | 0.810   |
| LDVAE          | 0.886       | 0.851         | 0.837         | 0.839       | 0.745        | 0.801        | 0.646        | 0.814          | 0.783   | 0.800   |
| CellBlast      | 0.837       | 0.880         | 0.869         | 0.858       | 0.774        | 0.746        | 0.675        | 0.819          | 0.760   | 0.802   |
| scFoundation   | 0.927       | 0.926         | 0.908         | 0.925       | 0.846        | 0.811        | 0.798        | 0.882          | 0.864   | 0.876   |
| scGPT          | 0.930       | 0.930         | 0.891         | 0.914       | 0.793        | 0.769        | 0.754        | 0.800          | 0.764   | 0.838   |
| SCimilarity    | 0.936       | 0.931         | 0.891         | 0.931       | 0.833        | 0.859        | 0.781        | 0.863          | 0.856   | 0.876   |
| UCE            | 0.935       | 0.917         | 0.905         | 0.911       | 0.829        | 0.879        | 0.770        | 0.895          | 0.858   | 0.878   |
| Geneformer     | 0.866       | 0.876         | 0.845         | 0.769       | 0.757        | 0.806        | 0.762        | 0.785          | 0.826   | 0.810   |
| scMulan        | 0.906       | 0.918         | 0.894         | 0.912       | 0.798        | 0.846        | 0.746        | 0.856          | 0.838   | 0.857   |
| CellPLM        | 0.931       | 0.918         | 0.885         | 0.913       | 0.732        | 0.815        | 0.779        | 0.745          | 0.846   | 0.841   |

Appendix Table 10: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 10.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.833       | 0.788         | 0.801         | 0.687       | 0.413        | 0.630        | 0.487        | 0.257          | 0.545   | 0.604   |
| CellFishing.jl | 0.889       | 0.912         | 0.907         | 0.910       | 0.827        | 0.857        | 0.747        | 0.882          | 0.730   | 0.851   |
| scVI           | 0.922       | 0.901         | 0.872         | 0.908       | 0.677        | 0.789        | 0.722        | 0.863          | 0.749   | 0.822   |
| LDVAE          | 0.891       | 0.859         | 0.838         | 0.845       | 0.740        | 0.807        | 0.677        | 0.817          | 0.795   | 0.808   |
| CellBlast      | 0.846       | 0.886         | 0.866         | 0.878       | 0.778        | 0.759        | 0.689        | 0.848          | 0.764   | 0.812   |
| scFoundation   | 0.929       | 0.934         | 0.909         | 0.930       | 0.850        | 0.824        | 0.800        | 0.888          | 0.870   | 0.882   |
| scGPT          | 0.930       | 0.939         | 0.893         | 0.916       | 0.797        | 0.775        | 0.760        | 0.825          | 0.762   | 0.844   |
| SCimilarity    | 0.939       | 0.939         | 0.893         | 0.936       | 0.852        | 0.864        | 0.787        | 0.871          | 0.864   | 0.883   |
| UCE            | 0.938       | 0.923         | 0.908         | 0.922       | 0.835        | 0.890        | 0.775        | 0.899          | 0.868   | 0.884   |
| Geneformer     | 0.877       | 0.889         | 0.842         | 0.794       | 0.770        | 0.816        | 0.772        | 0.817          | 0.837   | 0.824   |
| scMulan        | 0.913       | 0.925         | 0.897         | 0.911       | 0.827        | 0.856        | 0.748        | 0.840          | 0.846   | 0.863   |
| CellPLM        | 0.940       | 0.920         | 0.888         | 0.915       | 0.740        | 0.826        | 0.780        | 0.793          | 0.864   | 0.852   |

Appendix Table 11: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 20.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.822       | 0.789         | 0.798         | 0.719       | 0.428        | 0.633        | 0.478        | 0.304          | 0.546   | 0.613   |
| CellFishing.jl | 0.898       | 0.915         | 0.908         | 0.918       | 0.844        | 0.862        | 0.759        | 0.895          | 0.723   | 0.858   |
| scVI           | 0.930       | 0.908         | 0.872         | 0.916       | 0.688        | 0.789        | 0.748        | 0.886          | 0.754   | 0.832   |
| LDVAE          | 0.898       | 0.870         | 0.834         | 0.845       | 0.738        | 0.816        | 0.689        | 0.825          | 0.797   | 0.812   |
| CellBlast      | 0.844       | 0.899         | 0.863         | 0.882       | 0.793        | 0.770        | 0.695        | 0.842          | 0.761   | 0.817   |
| scFoundation   | 0.932       | 0.937         | 0.911         | 0.932       | 0.848        | 0.836        | 0.799        | 0.886          | 0.876   | 0.884   |
| scGPT          | 0.932       | 0.939         | 0.892         | 0.916       | 0.812        | 0.781        | 0.760        | 0.850          | 0.762   | 0.849   |
| SCimilarity    | 0.938       | 0.943         | 0.890         | 0.938       | 0.850        | 0.872        | 0.789        | 0.875          | 0.870   | 0.885   |
| UCE            | 0.945       | 0.933         | 0.903         | 0.923       | 0.829        | 0.897        | 0.774        | 0.897          | 0.872   | 0.886   |
| Geneformer     | 0.870       | 0.897         | 0.832         | 0.806       | 0.760        | 0.825        | 0.774        | 0.840          | 0.844   | 0.828   |
| scMulan        | 0.922       | 0.928         | 0.892         | 0.908       | 0.840        | 0.856        | 0.742        | 0.825          | 0.849   | 0.863   |
| CellPLM        | 0.941       | 0.925         | 0.888         | 0.913       | 0.749        | 0.832        | 0.782        | 0.817          | 0.874   | 0.858   |

Appendix Table 12: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 50.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.800       | 0.781         | 0.773         | 0.768       | 0.386        | 0.623        | 0.443        | 0.319          | 0.538   | 0.604   |
| CellFishing.jl | 0.905       | 0.921         | 0.904         | 0.930       | 0.840        | 0.866        | 0.777        | 0.897          | 0.710   | 0.861   |
| scVI           | 0.933       | 0.910         | 0.864         | 0.914       | 0.675        | 0.790        | 0.757        | 0.875          | 0.755   | 0.830   |
| LDVAE          | 0.899       | 0.876         | 0.826         | 0.839       | 0.736        | 0.815        | 0.710        | 0.838          | 0.800   | 0.815   |
| CellBlast      | 0.846       | 0.893         | 0.842         | 0.875       | 0.778        | 0.772        | 0.701        | 0.848          | 0.755   | 0.812   |
| scFoundation   | 0.934       | 0.937         | 0.903         | 0.934       | 0.857        | 0.851        | 0.799        | 0.886          | 0.879   | 0.887   |
| scGPT          | 0.934       | 0.939         | 0.887         | 0.914       | 0.816        | 0.780        | 0.761        | 0.816          | 0.750   | 0.844   |
| SCimilarity    | 0.938       | 0.952         | 0.885         | 0.930       | 0.852        | 0.883        | 0.792        | 0.867          | 0.875   | 0.886   |
| UCE            | 0.949       | 0.936         | 0.905         | 0.920       | 0.831        | 0.905        | 0.773        | 0.901          | 0.874   | 0.888   |
| Geneformer     | 0.861       | 0.886         | 0.813         | 0.808       | 0.738        | 0.822        | 0.776        | 0.837          | 0.844   | 0.820   |
| scMulan        | 0.924       | 0.928         | 0.890         | 0.898       | 0.825        | 0.858        | 0.732        | 0.821          | 0.849   | 0.858   |
| CellPLM        | 0.943       | 0.925         | 0.878         | 0.912       | 0.751        | 0.839        | 0.778        | 0.837          | 0.883   | 0.861   |



Appendix Table 13: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 100.

|                | 10x<br>(v2) | 10x<br>(v2) A | 10x<br>(v2) B | 10x<br>(v3) | CEL<br>-Seq2 | Drop<br>-seq | Seq<br>-Well | Smart<br>-seq2 | inDrops | Average |
|----------------|-------------|---------------|---------------|-------------|--------------|--------------|--------------|----------------|---------|---------|
| PCA            | 0.772       | 0.774         | 0.755         | 0.746       | 0.350        | 0.602        | 0.408        | 0.312          | 0.524   | 0.583   |
| CellFishing.jl | 0.910       | 0.921         | 0.901         | 0.933       | 0.829        | 0.871        | 0.778        | 0.895          | 0.695   | 0.859   |
| scVI           | 0.929       | 0.907         | 0.849         | 0.908       | 0.715        | 0.788        | 0.760        | 0.857          | 0.754   | 0.830   |
| LDVAE          | 0.889       | 0.872         | 0.814         | 0.829       | 0.738        | 0.816        | 0.724        | 0.808          | 0.802   | 0.810   |
| CellBlast      | 0.848       | 0.885         | 0.818         | 0.864       | 0.793        | 0.769        | 0.714        | 0.865          | 0.748   | 0.812   |
| scFoundation   | 0.933       | 0.939         | 0.898         | 0.931       | 0.861        | 0.859        | 0.803        | 0.892          | 0.885   | 0.889   |
| scGPT          | 0.932       | 0.932         | 0.880         | 0.913       | 0.802        | 0.775        | 0.762        | 0.817          | 0.742   | 0.840   |
| SCimilarity    | 0.935       | 0.951         | 0.876         | 0.926       | 0.846        | 0.887        | 0.791        | 0.865          | 0.877   | 0.884   |
| UCE            | 0.949       | 0.936         | 0.904         | 0.917       | 0.825        | 0.905        | 0.777        | 0.886          | 0.875   | 0.886   |
| Geneformer     | 0.855       | 0.875         | 0.800         | 0.804       | 0.722        | 0.813        | 0.777        | 0.802          | 0.840   | 0.810   |
| scMulan        | 0.925       | 0.926         | 0.887         | 0.889       | 0.823        | 0.852        | 0.720        | 0.819          | 0.844   | 0.854   |
| CellPLM        | 0.944       | 0.928         | 0.867         | 0.908       | 0.743        | 0.840        | 0.772        | 0.856          | 0.883   | 0.860   |

Appendix Table 14: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 1.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.888   | 0.882    | 0.906       | 0.674      | 0.838   |
| CellFishing.jl | 0.938   | 0.968    | 0.944       | 0.950      | 0.950   |
| scVI           | 0.915   | 0.919    | 0.875       | 0.893      | 0.901   |
| LDVAE          | 0.848   | 0.909    | 0.743       | 0.795      | 0.824   |
| CellBlast      | 0.301   | 0.249    | 0.132       | 0.189      | 0.218   |
| scFoundation   | 0.912   | 0.878    | 0.671       | 0.736      | 0.799   |
| scGPT          | 0.844   | 0.792    | 0.475       | 0.832      | 0.736   |
| SCimilarity    | 0.919   | 0.914    | 0.912       | 0.906      | 0.913   |
| UCE            | 0.830   | 0.658    | 0.453       | 0.739      | 0.670   |
| Geneformer     | 0.603   | 0.599    | 0.401       | 0.320      | 0.481   |
| scMulan        | 0.890   | 0.858    | 0.404       | 0.839      | 0.748   |

Appendix Table 15: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 5.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.833   | 0.851    | 0.887       | 0.737      | 0.827   |
| CellFishing.jl | 0.929   | 0.963    | 0.929       | 0.934      | 0.939   |
| scVI           | 0.899   | 0.922    | 0.877       | 0.882      | 0.895   |
| LDVAE          | 0.841   | 0.889    | 0.670       | 0.798      | 0.799   |
| CellBlast      | 0.243   | 0.301    | 0.131       | 0.198      | 0.218   |
| scFoundation   | 0.879   | 0.858    | 0.688       | 0.739      | 0.791   |
| scGPT          | 0.815   | 0.768    | 0.514       | 0.824      | 0.730   |
| SCimilarity    | 0.908   | 0.908    | 0.911       | 0.904      | 0.908   |
| UCE            | 0.809   | 0.670    | 0.404       | 0.719      | 0.651   |
| Geneformer     | 0.518   | 0.545    | 0.375       | 0.318      | 0.439   |
| scMulan        | 0.860   | 0.828    | 0.429       | 0.811      | 0.732   |

Appendix Table 16: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 10.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.807   | 0.830    | 0.881       | 0.779      | 0.824   |
| CellFishing.jl | 0.921   | 0.958    | 0.904       | 0.925      | 0.927   |
| scVI           | 0.889   | 0.916    | 0.877       | 0.877      | 0.890   |
| LDVAE          | 0.833   | 0.888    | 0.653       | 0.804      | 0.795   |
| CellBlast      | 0.216   | 0.287    | 0.144       | 0.204      | 0.213   |
| scFoundation   | 0.869   | 0.834    | 0.685       | 0.739      | 0.782   |
| scGPT          | 0.803   | 0.760    | 0.519       | 0.817      | 0.724   |
| SCimilarity    | 0.901   | 0.902    | 0.904       | 0.901      | 0.902   |
| UCE            | 0.799   | 0.667    | 0.398       | 0.707      | 0.643   |
| Geneformer     | 0.461   | 0.510    | 0.376       | 0.317      | 0.416   |
| scMulan        | 0.847   | 0.800    | 0.421       | 0.794      | 0.716   |

Appendix Table 17: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 20.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.774   | 0.808    | 0.874       | 0.793      | 0.812   |
| CellFishing.jl | 0.913   | 0.947    | 0.875       | 0.914      | 0.912   |
| scVI           | 0.877   | 0.902    | 0.876       | 0.871      | 0.881   |
| LDVAE          | 0.826   | 0.873    | 0.640       | 0.807      | 0.786   |
| CellBlast      | 0.229   | 0.305    | 0.129       | 0.214      | 0.220   |
| scFoundation   | 0.865   | 0.794    | 0.692       | 0.735      | 0.771   |
| scGPT          | 0.783   | 0.746    | 0.534       | 0.807      | 0.717   |
| SCimilarity    | 0.896   | 0.894    | 0.899       | 0.896      | 0.896   |
| UCE            | 0.790   | 0.663    | 0.401       | 0.699      | 0.638   |
| Geneformer     | 0.410   | 0.479    | 0.364       | 0.307      | 0.390   |
| scMulan        | 0.828   | 0.767    | 0.409       | 0.770      | 0.694   |

Appendix Table 18: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 50.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.722   | 0.754    | 0.864       | 0.792      | 0.783   |
| CellFishing.jl | 0.894   | 0.920    | 0.827       | 0.890      | 0.883   |
| scVI           | 0.861   | 0.868    | 0.866       | 0.856      | 0.863   |
| LDVAE          | 0.824   | 0.839    | 0.636       | 0.810      | 0.777   |
| CellBlast      | 0.200   | 0.279    | 0.117       | 0.206      | 0.200   |
| scFoundation   | 0.845   | 0.735    | 0.674       | 0.729      | 0.746   |
| scGPT          | 0.748   | 0.712    | 0.545       | 0.788      | 0.698   |
| SCimilarity    | 0.888   | 0.868    | 0.891       | 0.888      | 0.884   |
| UCE            | 0.774   | 0.652    | 0.417       | 0.684      | 0.632   |
| Geneformer     | 0.346   | 0.428    | 0.348       | 0.288      | 0.352   |
| scMulan        | 0.791   | 0.713    | 0.392       | 0.734      | 0.658   |

Appendix Table 19: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Avg-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 100.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.647   | 0.683    | 0.852       | 0.779      | 0.741   |
| CellFishing.jl | 0.869   | 0.872    | 0.781       | 0.863      | 0.846   |
| scVI           | 0.844   | 0.804    | 0.849       | 0.840      | 0.834   |
| LDVAE          | 0.822   | 0.801    | 0.651       | 0.808      | 0.770   |
| CellBlast      | 0.206   | 0.236    | 0.139       | 0.234      | 0.204   |
| scFoundation   | 0.819   | 0.686    | 0.640       | 0.721      | 0.716   |
| scGPT          | 0.717   | 0.676    | 0.545       | 0.767      | 0.676   |
| SCimilarity    | 0.876   | 0.831    | 0.877       | 0.878      | 0.866   |
| UCE            | 0.757   | 0.625    | 0.444       | 0.669      | 0.624   |
| Geneformer     | 0.295   | 0.366    | 0.336       | 0.274      | 0.318   |
| scMulan        | 0.754   | 0.656    | 0.370       | 0.705      | 0.621   |

Appendix Table 20: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 1.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.888   | 0.882    | 0.906       | 0.674      | 0.838   |
| CellFishing.jl | 0.938   | 0.968    | 0.944       | 0.950      | 0.950   |
| scVI           | 0.915   | 0.919    | 0.875       | 0.893      | 0.901   |
| LDVAE          | 0.848   | 0.909    | 0.743       | 0.795      | 0.824   |
| CellBlast      | 0.301   | 0.249    | 0.132       | 0.189      | 0.218   |
| scFoundation   | 0.912   | 0.878    | 0.671       | 0.736      | 0.799   |
| scGPT          | 0.844   | 0.792    | 0.475       | 0.832      | 0.736   |
| SCimilarity    | 0.919   | 0.914    | 0.912       | 0.906      | 0.913   |
| UCE            | 0.830   | 0.658    | 0.453       | 0.739      | 0.670   |
| Geneformer     | 0.603   | 0.599    | 0.401       | 0.320      | 0.481   |
| scMulan        | 0.890   | 0.858    | 0.404       | 0.839      | 0.748   |

Appendix Table 21: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 5.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.880   | 0.875    | 0.912       | 0.811      | 0.870   |
| CellFishing.jl | 0.938   | 0.968    | 0.955       | 0.949      | 0.952   |
| scVI           | 0.917   | 0.957    | 0.901       | 0.903      | 0.920   |
| LDVAE          | 0.896   | 0.941    | 0.710       | 0.821      | 0.842   |
| CellBlast      | 0.300   | 0.323    | 0.124       | 0.195      | 0.235   |
| scFoundation   | 0.920   | 0.904    | 0.730       | 0.739      | 0.823   |
| scGPT          | 0.846   | 0.829    | 0.586       | 0.842      | 0.776   |
| SCimilarity    | 0.918   | 0.942    | 0.950       | 0.926      | 0.934   |
| UCE            | 0.839   | 0.767    | 0.483       | 0.740      | 0.707   |
| Geneformer     | 0.616   | 0.646    | 0.469       | 0.341      | 0.518   |
| scMulan        | 0.905   | 0.886    | 0.422       | 0.829      | 0.760   |

Appendix Table 22: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 10.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.885   | 0.857    | 0.918       | 0.893      | 0.888   |
| CellFishing.jl | 0.938   | 0.969    | 0.959       | 0.948      | 0.954   |
| scVI           | 0.912   | 0.961    | 0.895       | 0.902      | 0.918   |
| LDVAE          | 0.900   | 0.947    | 0.679       | 0.828      | 0.839   |
| CellBlast      | 0.302   | 0.317    | 0.160       | 0.191      | 0.242   |
| scFoundation   | 0.918   | 0.888    | 0.774       | 0.735      | 0.829   |
| scGPT          | 0.850   | 0.834    | 0.614       | 0.840      | 0.785   |
| SCimilarity    | 0.920   | 0.939    | 0.948       | 0.929      | 0.934   |
| UCE            | 0.840   | 0.784    | 0.469       | 0.738      | 0.708   |
| Geneformer     | 0.574   | 0.613    | 0.552       | 0.346      | 0.521   |
| scMulan        | 0.898   | 0.870    | 0.415       | 0.823      | 0.752   |

Appendix Table 23: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 20.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.887   | 0.828    | 0.918       | 0.888      | 0.880   |
| CellFishing.jl | 0.929   | 0.970    | 0.951       | 0.940      | 0.948   |
| scVI           | 0.906   | 0.951    | 0.892       | 0.891      | 0.910   |
| LDVAE          | 0.908   | 0.946    | 0.657       | 0.830      | 0.835   |
| CellBlast      | 0.306   | 0.471    | 0.064       | 0.221      | 0.265   |
| scFoundation   | 0.920   | 0.857    | 0.801       | 0.740      | 0.830   |
| scGPT          | 0.849   | 0.835    | 0.708       | 0.834      | 0.807   |
| SCimilarity    | 0.919   | 0.933    | 0.948       | 0.921      | 0.931   |
| UCE            | 0.832   | 0.786    | 0.466       | 0.731      | 0.703   |
| Geneformer     | 0.511   | 0.601    | 0.567       | 0.313      | 0.498   |
| scMulan        | 0.896   | 0.853    | 0.370       | 0.794      | 0.728   |

Appendix Table 24: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 50.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.865   | 0.763    | 0.908       | 0.858      | 0.848   |
| CellFishing.jl | 0.915   | 0.959    | 0.945       | 0.918      | 0.934   |
| scVI           | 0.899   | 0.911    | 0.897       | 0.883      | 0.898   |
| LDVAE          | 0.904   | 0.912    | 0.658       | 0.824      | 0.825   |
| CellBlast      | 0.302   | 0.451    | 0.058       | 0.395      | 0.302   |
| scFoundation   | 0.922   | 0.786    | 0.851       | 0.735      | 0.823   |
| scGPT          | 0.833   | 0.835    | 0.755       | 0.826      | 0.812   |
| SCimilarity    | 0.911   | 0.930    | 0.950       | 0.919      | 0.927   |
| UCE            | 0.823   | 0.792    | 0.505       | 0.720      | 0.710   |
| Geneformer     | 0.422   | 0.573    | 0.494       | 0.314      | 0.451   |
| scMulan        | 0.880   | 0.807    | 0.328       | 0.777      | 0.698   |

Appendix Table 25: Evaluating single-cell FMs in cross-platform retrieval setting on human PBMC dataset (Vote-Acc ). Setting: Leave-one-out. Given scRNA-seq sequencing results from  $N$  platforms. The sequencing results from  $N - 1$  platforms are used as reference the remaining platform is used as query. We set the retrieval cell number for each query cell as 100.

|                | CEL-Seq | CEL-Seq2 | Fluidigm C1 | Smart-Seq2 | Average |
|----------------|---------|----------|-------------|------------|---------|
| PCA            | 0.816   | 0.717    | 0.898       | 0.849      | 0.820   |
| CellFishing.jl | 0.913   | 0.955    | 0.944       | 0.889      | 0.925   |
| scVI           | 0.882   | 0.906    | 0.895       | 0.876      | 0.890   |
| LDVAE          | 0.911   | 0.856    | 0.672       | 0.821      | 0.815   |
| CellBlast      | 0.304   | 0.472    | 0.091       | 0.444      | 0.328   |
| scFoundation   | 0.917   | 0.741    | 0.867       | 0.730      | 0.814   |
| scGPT          | 0.811   | 0.823    | 0.771       | 0.813      | 0.804   |
| SCimilarity    | 0.865   | 0.797    | 0.301       | 0.763      | 0.682   |
| UCE            | 0.813   | 0.782    | 0.639       | 0.707      | 0.735   |
| Geneformer     | 0.365   | 0.546    | 0.429       | 0.309      | 0.412   |
| scMulan        | 0.865   | 0.797    | 0.301       | 0.763      | 0.682   |

Appendix Table 26: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on large intestine.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.692        | 0.652 | 0.645 | 0.633 | 0.625 | 0.599 | 0.776        | 0.763 | 0.750 | 0.734 | 0.718 | 0.704 |
| CellFishing.jl | 0.509        | 0.510 | 0.510 | 0.512 | 0.527 | 0.541 | 0.825        | 0.815 | 0.813 | 0.811 | 0.813 | 0.812 |
| scVI           | 0.279        | 0.303 | 0.312 | 0.313 | 0.312 | 0.304 | 0.639        | 0.643 | 0.635 | 0.618 | 0.577 | 0.564 |
| LDVAE          | 0.366        | 0.288 | 0.290 | 0.298 | 0.314 | 0.326 | 0.662        | 0.661 | 0.653 | 0.651 | 0.643 | 0.634 |
| CellBlast      | 0.525        | 0.573 | 0.554 | 0.582 | 0.579 | 0.573 | 0.427        | 0.335 | 0.316 | 0.299 | 0.288 | 0.282 |
| scFoundation   | 0.904        | 0.888 | 0.883 | 0.877 | 0.863 | 0.874 | 0.508        | 0.525 | 0.526 | 0.529 | 0.548 | 0.607 |
| scGPT          | 0.744        | 0.809 | 0.830 | 0.845 | 0.853 | 0.855 | 0.831        | 0.849 | 0.854 | 0.871 | 0.873 | 0.874 |
| SCimilarity    | 0.796        | 0.844 | 0.870 | 0.898 | 0.928 | 0.945 | 0.871        | 0.880 | 0.880 | 0.885 | 0.903 | 0.904 |
| UCE            | 0.732        | 0.890 | 0.895 | 0.873 | 0.833 | 0.814 | 0.742        | 0.831 | 0.832 | 0.835 | 0.846 | 0.858 |
| Geneformer     | 0.609        | 0.707 | 0.757 | 0.803 | 0.839 | 0.853 | 0.444        | 0.495 | 0.511 | 0.532 | 0.531 | 0.523 |
| scMulan        | 0.513        | 0.579 | 0.615 | 0.651 | 0.695 | 0.727 | 0.632        | 0.626 | 0.615 | 0.594 | 0.544 | 0.504 |
| CellPLM        | 0.713        | 0.758 | 0.790 | 0.841 | 0.889 | 0.907 | 0.693        | 0.702 | 0.704 | 0.711 | 0.720 | 0.744 |

Appendix Table 27: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on Adipose.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.992        | 0.992 | 0.992 | 0.991 | 0.996 | 0.994 | 0.745        | 0.739 | 0.738 | 0.735 | 0.728 | 0.723 |
| CellFishing.jl | 0.928        | 0.931 | 0.930 | 0.925 | 0.917 | 0.909 | 0.736        | 0.732 | 0.730 | 0.727 | 0.718 | 0.707 |
| scVI           | 0.997        | 0.998 | 0.996 | 0.994 | 0.991 | 0.979 | 0.758        | 0.756 | 0.753 | 0.749 | 0.746 | 0.740 |
| LDVAE          | 0.999        | 0.998 | 0.998 | 0.998 | 0.997 | 0.997 | 0.751        | 0.751 | 0.751 | 0.751 | 0.751 | 0.751 |
| CellBlast      | 0.770        | 0.753 | 0.737 | 0.725 | 0.740 | 0.740 | 0.895        | 0.882 | 0.892 | 0.849 | 0.751 | 0.751 |
| scFoundation   | 0.995        | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 | 0.751        | 0.751 | 0.751 | 0.751 | 0.751 | 0.751 |
| scGPT          | 0.858        | 0.870 | 0.863 | 0.860 | 0.842 | 0.808 | 0.742        | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 |
| SCimilarity    | 0.995        | 0.995 | 0.995 | 0.996 | 0.996 | 0.996 | 0.885        | 0.760 | 0.755 | 0.752 | 0.751 | 0.751 |
| UCE            | 0.997        | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.751        | 0.751 | 0.751 | 0.751 | 0.751 | 0.751 |
| Geneformer     | 0.907        | 0.924 | 0.929 | 0.928 | 0.929 | 0.927 | 0.738        | 0.740 | 0.740 | 0.738 | 0.735 | 0.733 |
| scMulan        | 0.992        | 0.981 | 0.980 | 0.980 | 0.976 | 0.948 | 0.772        | 0.758 | 0.751 | 0.751 | 0.751 | 0.750 |
| CellPLM        | 0.995        | 0.996 | 0.996 | 0.996 | 0.995 | 0.995 | 0.752        | 0.751 | 0.751 | 0.751 | 0.751 | 0.751 |

Appendix Table 28: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on subcutaneous adipose.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.975        | 0.995 | 0.995 | 0.995 | 0.994 | 0.989 | 0.659        | 0.589 | 0.600 | 0.602 | 0.555 | 0.518 |
| CellFishing.jl | 0.867        | 0.858 | 0.852 | 0.832 | 0.798 | 0.767 | 0.613        | 0.592 | 0.592 | 0.579 | 0.535 | 0.476 |
| scVI           | 0.906        | 0.984 | 0.998 | 0.998 | 0.998 | 0.998 | 0.641        | 0.645 | 0.649 | 0.649 | 0.650 | 0.650 |
| LDVAE          | 0.987        | 0.997 | 0.999 | 0.998 | 0.998 | 0.998 | 0.643        | 0.649 | 0.649 | 0.649 | 0.650 | 0.650 |
| CellBlast      | 0.460        | 0.509 | 0.534 | 0.551 | 0.581 | 0.601 | 0.801        | 0.921 | 0.650 | 0.648 | 0.648 | 0.648 |
| scFoundation   | 0.985        | 0.997 | 0.996 | 0.997 | 0.996 | 0.995 | 0.964        | 0.952 | 0.695 | 0.650 | 0.649 | 0.649 |
| scGPT          | 0.825        | 0.845 | 0.839 | 0.839 | 0.808 | 0.787 | 0.737        | 0.649 | 0.648 | 0.648 | 0.649 | 0.649 |
| SCimilarity    | 0.957        | 0.990 | 0.993 | 0.999 | 0.999 | 0.999 | 0.960        | 0.973 | 0.977 | 0.650 | 0.650 | 0.650 |
| UCE            | 0.986        | 0.993 | 0.998 | 0.999 | 0.999 | 0.999 | 0.961        | 0.771 | 0.650 | 0.649 | 0.650 | 0.650 |
| Geneformer     | 0.744        | 0.728 | 0.706 | 0.704 | 0.666 | 0.637 | 0.659        | 0.627 | 0.629 | 0.629 | 0.624 | 0.478 |
| scMulan        | 0.949        | 0.940 | 0.939 | 0.936 | 0.935 | 0.935 | 0.969        | 0.969 | 0.967 | 0.650 | 0.650 | 0.649 |
| CellPLM        | 0.910        | 0.903 | 0.933 | 0.938 | 0.938 | 0.933 | 0.880        | 0.711 | 0.650 | 0.650 | 0.650 | 0.650 |

Appendix Table 29: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on spleen.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.833        | 0.844 | 0.838 | 0.840 | 0.838 | 0.842 | 0.557        | 0.598 | 0.660 | 0.659 | 0.663 | 0.659 |
| CellFishing.jl | 0.598        | 0.610 | 0.614 | 0.619 | 0.626 | 0.644 | 0.692        | 0.703 | 0.690 | 0.670 | 0.645 | 0.635 |
| scVI           | 0.784        | 0.804 | 0.815 | 0.829 | 0.836 | 0.856 | 0.675        | 0.718 | 0.713 | 0.699 | 0.692 | 0.663 |
| LDVAE          | 0.837        | 0.846 | 0.845 | 0.842 | 0.842 | 0.850 | 0.522        | 0.672 | 0.811 | 0.823 | 0.829 | 0.818 |
| CellBlast      | 0.508        | 0.447 | 0.405 | 0.428 | 0.320 | 0.246 | 0.428        | 0.470 | 0.484 | 0.492 | 0.519 | 0.500 |
| scFoundation   | 0.877        | 0.876 | 0.877 | 0.876 | 0.877 | 0.879 | 0.864        | 0.830 | 0.833 | 0.833 | 0.837 | 0.863 |
| scGPT          | 0.890        | 0.894 | 0.894 | 0.896 | 0.899 | 0.900 | 0.826        | 0.829 | 0.822 | 0.819 | 0.811 | 0.802 |
| SCimilarity    | 0.904        | 0.907 | 0.908 | 0.908 | 0.909 | 0.909 | 0.749        | 0.780 | 0.836 | 0.835 | 0.850 | 0.882 |
| UCE            | 0.850        | 0.890 | 0.901 | 0.904 | 0.908 | 0.909 | 0.757        | 0.815 | 0.822 | 0.831 | 0.830 | 0.828 |
| Geneformer     | 0.585        | 0.590 | 0.584 | 0.579 | 0.554 | 0.527 | 0.705        | 0.726 | 0.724 | 0.702 | 0.670 | 0.644 |
| scMulan        | 0.728        | 0.823 | 0.833 | 0.852 | 0.865 | 0.880 | 0.760        | 0.757 | 0.744 | 0.595 | 0.429 | 0.703 |
| CellPLM        | 0.868        | 0.873 | 0.873 | 0.874 | 0.872 | 0.870 | 0.571        | 0.726 | 0.751 | 0.752 | 0.751 | 0.757 |

Appendix Table 30: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on kidney.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.944        | 0.933 | 0.939 | 0.929 | 0.924 | 0.902 | 0.928        | 0.993 | 0.995 | 0.995 | 0.968 | 0.924 |
| CellFishing.jl | 0.834        | 0.872 | 0.876 | 0.890 | 0.901 | 0.904 | 0.815        | 0.894 | 0.908 | 0.918 | 0.926 | 0.929 |
| scVI           | 0.957        | 0.964 | 0.967 | 0.971 | 0.970 | 0.970 | 0.990        | 0.991 | 0.992 | 0.991 | 0.991 | 0.988 |
| LDVAE          | 0.971        | 0.966 | 0.966 | 0.966 | 0.970 | 0.977 | 0.959        | 0.985 | 0.991 | 0.994 | 0.990 | 0.929 |
| CellBlast      | 0.964        | 0.964 | 0.964 | 0.965 | 0.969 | 0.969 | 0.523        | 0.524 | 0.533 | 0.545 | 0.557 | 0.599 |
| scFoundation   | 0.976        | 0.978 | 0.980 | 0.981 | 0.981 | 0.981 | 0.962        | 0.977 | 0.975 | 0.976 | 0.976 | 0.970 |
| scGPT          | 0.974        | 0.974 | 0.975 | 0.975 | 0.974 | 0.975 | 0.983        | 0.992 | 0.992 | 0.993 | 0.993 | 0.988 |
| SCimilarity    | 0.967        | 0.963 | 0.963 | 0.963 | 0.964 | 0.964 | 0.976        | 0.997 | 0.997 | 0.998 | 0.998 | 0.993 |
| UCE            | 0.974        | 0.976 | 0.976 | 0.977 | 0.978 | 0.981 | 0.984        | 0.997 | 0.997 | 0.998 | 0.998 | 0.998 |
| Geneformer     | 0.879        | 0.924 | 0.920 | 0.884 | 0.846 | 0.842 | 0.882        | 0.885 | 0.883 | 0.877 | 0.857 | 0.829 |
| scMulan        | 0.961        | 0.961 | 0.961 | 0.961 | 0.961 | 0.960 | 0.989        | 0.991 | 0.992 | 0.994 | 0.988 | 0.897 |
| CellPLM        | 0.971        | 0.970 | 0.970 | 0.966 | 0.964 | 0.963 | 0.955        | 0.969 | 0.970 | 0.971 | 0.972 | 0.929 |

Appendix Table 31: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on thymus.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.637        | 0.633 | 0.633 | 0.624 | 0.601 | 0.558 | 0.100        | 0.123 | 0.122 | 0.113 | 0.092 | 0.103 |
| CellFishing.jl | 0.605        | 0.609 | 0.609 | 0.608 | 0.604 | 0.598 | 0.067        | 0.059 | 0.057 | 0.057 | 0.055 | 0.051 |
| scVI           | 0.638        | 0.641 | 0.636 | 0.631 | 0.611 | 0.536 | 0.101        | 0.086 | 0.078 | 0.071 | 0.052 | 0.045 |
| LDVAE          | 0.624        | 0.643 | 0.642 | 0.638 | 0.635 | 0.623 | 0.109        | 0.092 | 0.079 | 0.071 | 0.066 | 0.063 |
| CellBlast      | 0.270        | 0.232 | 0.226 | 0.200 | 0.159 | 0.105 | 0.204        | 0.175 | 0.156 | 0.153 | 0.151 | 0.149 |
| scFoundation   | 0.599        | 0.597 | 0.594 | 0.593 | 0.588 | 0.576 | 0.232        | 0.238 | 0.247 | 0.244 | 0.314 | 0.392 |
| scGPT          | 0.591        | 0.581 | 0.577 | 0.569 | 0.550 | 0.489 | 0.245        | 0.289 | 0.302 | 0.309 | 0.310 | 0.306 |
| SCimilarity    | 0.684        | 0.689 | 0.693 | 0.697 | 0.699 | 0.700 | 0.196        | 0.201 | 0.219 | 0.274 | 0.409 | 0.452 |
| UCE            | 0.627        | 0.623 | 0.612 | 0.602 | 0.601 | 0.601 | 0.220        | 0.203 | 0.212 | 0.236 | 0.411 | 0.434 |
| Geneformer     | 0.068        | 0.031 | 0.017 | 0.012 | 0.006 | 0.001 | 0.044        | 0.036 | 0.033 | 0.030 | 0.028 | 0.027 |
| scMulan        | 0.650        | 0.650 | 0.649 | 0.649 | 0.638 | 0.630 | 0.207        | 0.212 | 0.198 | 0.203 | 0.207 | 0.447 |
| CellPLM        | 0.627        | 0.613 | 0.610 | 0.607 | 0.596 | 0.571 | 0.174        | 0.185 | 0.189 | 0.187 | 0.194 | 0.217 |

Appendix Table 32: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on lung.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.264        | 0.463 | 0.619 | 0.634 | 0.598 | 0.572 | 0.377        | 0.351 | 0.341 | 0.336 | 0.329 | 0.310 |
| CellFishing.jl | 0.604        | 0.700 | 0.698 | 0.693 | 0.690 | 0.695 | 0.377        | 0.418 | 0.431 | 0.438 | 0.436 | 0.410 |
| scVI           | 0.431        | 0.529 | 0.672 | 0.679 | 0.681 | 0.678 | 0.514        | 0.524 | 0.527 | 0.527 | 0.514 | 0.499 |
| LDVAE          | 0.462        | 0.582 | 0.619 | 0.624 | 0.626 | 0.676 | 0.220        | 0.202 | 0.203 | 0.196 | 0.204 | 0.197 |
| CellBlast      | 0.411        | 0.621 | 0.673 | 0.674 | 0.667 | 0.652 | 0.075        | 0.073 | 0.071 | 0.068 | 0.058 | 0.050 |
| scFoundation   | 0.768        | 0.788 | 0.786 | 0.783 | 0.782 | 0.775 | 0.486        | 0.549 | 0.566 | 0.574 | 0.584 | 0.583 |
| scGPT          | 0.750        | 0.742 | 0.733 | 0.723 | 0.713 | 0.705 | 0.343        | 0.369 | 0.377 | 0.381 | 0.384 | 0.378 |
| SCimilarity    | 0.788        | 0.810 | 0.785 | 0.764 | 0.752 | 0.747 | 0.498        | 0.538 | 0.547 | 0.548 | 0.543 | 0.529 |
| UCE            | 0.780        | 0.795 | 0.787 | 0.786 | 0.781 | 0.780 | 0.503        | 0.554 | 0.564 | 0.571 | 0.569 | 0.557 |
| Geneformer     | 0.399        | 0.399 | 0.368 | 0.300 | 0.198 | 0.152 | 0.279        | 0.309 | 0.326 | 0.340 | 0.346 | 0.340 |
| scMulan        | 0.751        | 0.766 | 0.771 | 0.774 | 0.751 | 0.736 | 0.452        | 0.511 | 0.525 | 0.529 | 0.526 | 0.517 |
| CellPLM        | 0.732        | 0.749 | 0.747 | 0.739 | 0.723 | 0.698 | 0.398        | 0.438 | 0.452 | 0.467 | 0.484 | 0.486 |

Appendix Table 33: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on liver.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.264        | 0.463 | 0.619 | 0.634 | 0.598 | 0.572 | 0.377        | 0.351 | 0.341 | 0.336 | 0.329 | 0.310 |
| CellFishing.jl | 0.604        | 0.700 | 0.698 | 0.693 | 0.690 | 0.695 | 0.377        | 0.418 | 0.431 | 0.438 | 0.436 | 0.410 |
| scVI           | 0.431        | 0.529 | 0.672 | 0.679 | 0.681 | 0.678 | 0.514        | 0.524 | 0.527 | 0.527 | 0.514 | 0.499 |
| LDVAE          | 0.462        | 0.582 | 0.619 | 0.624 | 0.626 | 0.676 | 0.220        | 0.202 | 0.203 | 0.196 | 0.204 | 0.197 |
| CellBlast      | 0.411        | 0.621 | 0.673 | 0.674 | 0.667 | 0.652 | 0.075        | 0.073 | 0.071 | 0.068 | 0.058 | 0.050 |
| scFoundation   | 0.768        | 0.788 | 0.786 | 0.783 | 0.782 | 0.775 | 0.486        | 0.549 | 0.566 | 0.574 | 0.584 | 0.583 |
| scGPT          | 0.750        | 0.742 | 0.733 | 0.723 | 0.713 | 0.705 | 0.343        | 0.369 | 0.377 | 0.381 | 0.384 | 0.378 |
| SCimilarity    | 0.788        | 0.810 | 0.785 | 0.764 | 0.752 | 0.747 | 0.498        | 0.538 | 0.547 | 0.548 | 0.543 | 0.529 |
| UCE            | 0.780        | 0.795 | 0.787 | 0.786 | 0.781 | 0.780 | 0.503        | 0.554 | 0.564 | 0.571 | 0.569 | 0.557 |
| Geneformer     | 0.399        | 0.399 | 0.368 | 0.300 | 0.198 | 0.152 | 0.279        | 0.309 | 0.326 | 0.340 | 0.346 | 0.340 |
| scMulan        | 0.751        | 0.766 | 0.771 | 0.774 | 0.751 | 0.736 | 0.452        | 0.511 | 0.525 | 0.529 | 0.526 | 0.517 |
| CellPLM        | 0.732        | 0.749 | 0.747 | 0.739 | 0.723 | 0.698 | 0.398        | 0.438 | 0.452 | 0.467 | 0.484 | 0.486 |

Appendix Table 34: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on skin of body.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.842        | 0.998 | 1.000 | 0.947 | 0.668 | 0.659 | 1.000        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| CellFishing.jl | 0.779        | 0.783 | 0.747 | 0.541 | 0.341 | 0.341 | 0.909        | 0.948 | 0.970 | 0.980 | 0.980 | 0.973 |
| scVI           | 0.874        | 0.923 | 0.676 | 0.676 | 0.663 | 0.659 | 1.000        | 0.988 | 0.988 | 0.877 | 0.277 | 0.420 |
| LDVAE          | 0.612        | 0.670 | 0.675 | 0.565 | 0.659 | 0.659 | 0.933        | 0.894 | 0.894 | 0.881 | 0.825 | 0.812 |
| CellBlast      | 0.520        | 0.402 | 0.362 | 0.436 | 0.665 | 0.659 | 0.696        | 0.753 | 0.694 | 0.733 | 0.640 | 0.573 |
| scFoundation   | 0.800        | 0.711 | 0.761 | 0.650 | 0.659 | 0.659 | 0.822        | 0.830 | 0.822 | 0.802 | 0.778 | 0.736 |
| scGPT          | 0.970        | 0.929 | 0.929 | 0.964 | 0.663 | 0.659 | 0.988        | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| SCimilarity    | 1.000        | 0.999 | 0.999 | 0.993 | 0.659 | 0.659 | 0.980        | 0.946 | 0.941 | 0.941 | 0.928 | 0.909 |
| UCE            | 1.000        | 0.999 | 0.996 | 0.957 | 0.745 | 0.659 | 1.000        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Geneformer     | 0.366        | 0.354 | 0.348 | 0.393 | 0.610 | 0.659 | 0.479        | 0.427 | 0.407 | 0.333 | 0.227 | 0.178 |
| scMulan        | 0.944        | 0.943 | 0.943 | 0.938 | 0.659 | 0.659 | 0.958        | 0.968 | 0.980 | 0.970 | 0.960 | 0.941 |
| CellPLM        | 0.989        | 0.985 | 0.986 | 0.875 | 0.659 | 0.659 | 1.000        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Appendix Table 35: Evaluating single-cell FMs in cross-species retrieval setting (Vote-Acc ) on trachea.

| Setting        | Human->Mouse |       |       |       |       |       | Mouse->Human |       |       |       |       |       |
|----------------|--------------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|-------|
| K              | 1            | 5     | 10    | 20    | 50    | 100   | 1            | 5     | 10    | 20    | 50    | 100   |
| PCA            | 0.902        | 0.919 | 0.921 | 0.918 | 0.916 | 0.914 | 0.931        | 0.922 | 0.910 | 0.904 | 0.880 | 0.865 |
| CellFishing.jl | 0.982        | 0.983 | 0.984 | 0.984 | 0.991 | 0.990 | 0.943        | 0.949 | 0.952 | 0.951 | 0.952 | 0.952 |
| scVI           | 0.929        | 0.934 | 0.929 | 0.920 | 0.907 | 0.891 | 0.842        | 0.867 | 0.868 | 0.867 | 0.866 | 0.884 |
| LDVAE          | 0.932        | 0.939 | 0.938 | 0.937 | 0.936 | 0.937 | 0.966        | 0.962 | 0.964 | 0.963 | 0.961 | 0.961 |
| CellBlast      | 0.551        | 0.727 | 0.766 | 0.785 | 0.824 | 0.815 | 0.656        | 0.636 | 0.618 | 0.587 | 0.601 | 0.572 |
| scFoundation   | 0.919        | 0.926 | 0.925 | 0.926 | 0.920 | 0.913 | 0.950        | 0.947 | 0.939 | 0.931 | 0.925 | 0.923 |
| scGPT          | 0.948        | 0.945 | 0.943 | 0.935 | 0.918 | 0.903 | 0.944        | 0.948 | 0.950 | 0.949 | 0.946 | 0.925 |
| SCimilarity    | 0.976        | 0.991 | 0.991 | 0.992 | 0.993 | 0.988 | 0.959        | 0.964 | 0.963 | 0.962 | 0.961 | 0.959 |
| UCE            | 0.987        | 0.985 | 0.987 | 0.986 | 0.978 | 0.964 | 0.971        | 0.968 | 0.966 | 0.964 | 0.962 | 0.962 |
| Geneformer     | 0.707        | 0.718 | 0.716 | 0.705 | 0.669 | 0.657 | 0.791        | 0.784 | 0.772 | 0.756 | 0.720 | 0.685 |
| scMulan        | 0.981        | 0.985 | 0.980 | 0.964 | 0.962 | 0.966 | 0.969        | 0.970 | 0.968 | 0.966 | 0.966 | 0.954 |
| CellPLM        | 0.927        | 0.912 | 0.897 | 0.886 | 0.879 | 0.869 | 0.963        | 0.964 | 0.964 | 0.965 | 0.966 | 0.963 |

Appendix Table 36: Evaluating single-cell FMs in cross-omic retrieval setting (Vote-Acc ) on 10x multi-omics ( $K = 1$ ).  $K$  indicates the number of cells retrieved for each query cell.

| Setting        | scRNA->scATAC |         |          | scATAC->scRNA |         |          |
|----------------|---------------|---------|----------|---------------|---------|----------|
| Method         | Recall        | Avg-Acc | Vote-Acc | Recall        | Avg-Acc | Vote-Acc |
| PCA            | 0.000         | 0.170   | 0.170    | 0.002         | 0.505   | 0.505    |
| CellFishing.jl | 0.001         | 0.520   | 0.520    | 0.002         | 0.540   | 0.540    |
| scVI           | 0.000         | 0.180   | 0.180    | 0.000         | 0.183   | 0.183    |
| LDVAE          | 0.00          | 0.181   | 0.181    | 0.000         | 0.153   | 0.153    |
| CellBlast      | 0.003         | 0.537   | 0.537    | 0.002         | 0.609   | 0.609    |
| scFoundation   | 0.004         | 0.732   | 0.732    | 0.003         | 0.639   | 0.639    |
| scGPT          | 0.001         | 0.551   | 0.551    | 0.002         | 0.516   | 0.516    |
| SCimilarity    | 0.002         | 0.599   | 0.599    | 0.002         | 0.507   | 0.507    |
| UCE            | 0.003         | 0.783   | 0.783    | 0.003         | 0.643   | 0.643    |
| Geneformer     | 0.000         | 0.271   | 0.271    | 0.000         | 0.184   | 0.184    |
| scMulan        | 0.002         | 0.578   | 0.578    | 0.002         | 0.508   | 0.508    |
| CellPLM        | 0.003         | 0.607   | 0.607    | 0.002         | 0.507   | 0.507    |



Appendix Table 37: Evaluating single-cell FMs in cross-omic retrieval setting (Vote-Acc) on 10x multi-omics ( $K = 5$ ).  $K$  indicates the number of cells retrieved for each query cell.

| Setting        | scRNA->scATAC |         |          | scATAC->scRNA |         |          |
|----------------|---------------|---------|----------|---------------|---------|----------|
| Method         | Recall        | Avg-Acc | Vote-Acc | Recall        | Avg-Acc | Vote-Acc |
| PCA            | 0.001         | 0.173   | 0.190    | 0.008         | 0.485   | 0.533    |
| CellFishing.jl | 0.004         | 0.510   | 0.605    | 0.010         | 0.537   | 0.588    |
| scVI           | 0.001         | 0.186   | 0.190    | 0.000         | 0.192   | 0.199    |
| LDVAE          | 0.001         | 0.183   | 0.202    | 0.000         | 0.157   | 0.165    |
| CellBlast      | 0.011         | 0.524   | 0.552    | 0.012         | 0.601   | 0.633    |
| scFoundation   | 0.018         | 0.716   | 0.794    | 0.011         | 0.637   | 0.685    |
| scGPT          | 0.008         | 0.546   | 0.619    | 0.004         | 0.504   | 0.540    |
| SCimilarity    | 0.010         | 0.590   | 0.689    | 0.009         | 0.499   | 0.517    |
| UCE            | 0.014         | 0.766   | 0.843    | 0.012         | 0.640   | 0.686    |
| Geneformer     | 0.001         | 0.261   | 0.325    | 0.001         | 0.179   | 0.197    |
| scMulan        | 0.008         | 0.563   | 0.639    | 0.006         | 0.497   | 0.549    |
| CellPLM        | 0.010         | 0.614   | 0.681    | 0.008         | 0.496   | 0.527    |

Appendix Table 38: Evaluating single-cell FMs in cross-omic retrieval setting (Vote-Acc) on 10x multi-omics ( $K = 10$ ).  $K$  indicates the number of cells retrieved for each query cell.

| Setting        | scRNA->scATAC |         |          | scATAC->scRNA |         |          |
|----------------|---------------|---------|----------|---------------|---------|----------|
| Method         | Recall        | Avg-Acc | Vote-Acc | Recall        | Avg-Acc | Vote-Acc |
| PCA            | 0.002         | 0.170   | 0.201    | 0.015         | 0.473   | 0.534    |
| CellFishing.jl | 0.007         | 0.500   | 0.627    | 0.016         | 0.536   | 0.603    |
| scVI           | 0.002         | 0.187   | 0.195    | 0.002         | 0.195   | 0.204    |
| LDVAE          | 0.002         | 0.188   | 0.214    | 0.001         | 0.160   | 0.174    |
| CellBlast      | 0.021         | 0.524   | 0.567    | 0.020         | 0.598   | 0.640    |
| scFoundation   | 0.031         | 0.709   | 0.811    | 0.022         | 0.635   | 0.691    |
| scGPT          | 0.015         | 0.537   | 0.641    | 0.008         | 0.501   | 0.548    |
| SCimilarity    | 0.019         | 0.587   | 0.712    | 0.014         | 0.498   | 0.521    |
| UCE            | 0.029         | 0.755   | 0.857    | 0.020         | 0.639   | 0.696    |
| Geneformer     | 0.003         | 0.253   | 0.342    | 0.001         | 0.177   | 0.198    |
| scMulan        | 0.014         | 0.550   | 0.650    | 0.010         | 0.490   | 0.558    |
| CellPLM        | 0.018         | 0.608   | 0.699    | 0.015         | 0.491   | 0.538    |

Appendix Table 39: Evaluating single-cell FMs in cross-omic retrieval setting (Vote-Acc) on 10x multi-omics ( $K = 20$ ).  $K$  indicates the number of cells retrieved for each query cell.

| Setting        | scRNA->scATAC |         |          | scATAC->scRNA |         |          |
|----------------|---------------|---------|----------|---------------|---------|----------|
| Method         | Recall        | Avg-Acc | Vote-Acc | Recall        | Avg-Acc | Vote-Acc |
| PCA            | 0.003         | 0.168   | 0.211    | 0.025         | 0.459   | 0.536    |
| CellFishing.jl | 0.012         | 0.487   | 0.625    | 0.030         | 0.536   | 0.619    |
| scVI           | 0.003         | 0.190   | 0.198    | 0.003         | 0.196   | 0.206    |
| LDVAE          | 0.004         | 0.195   | 0.219    | 0.002         | 0.164   | 0.180    |
| CellBlast      | 0.038         | 0.522   | 0.571    | 0.035         | 0.597   | 0.652    |
| scFoundation   | 0.054         | 0.699   | 0.817    | 0.039         | 0.631   | 0.694    |
| scGPT          | 0.028         | 0.528   | 0.664    | 0.014         | 0.496   | 0.550    |
| SCimilarity    | 0.039         | 0.582   | 0.721    | 0.024         | 0.498   | 0.526    |
| UCE            | 0.049         | 0.737   | 0.864    | 0.036         | 0.637   | 0.704    |
| Geneformer     | 0.006         | 0.244   | 0.345    | 0.002         | 0.174   | 0.198    |
| scMulan        | 0.023         | 0.533   | 0.652    | 0.016         | 0.482   | 0.555    |
| CellPLM        | 0.032         | 0.597   | 0.710    | 0.029         | 0.484   | 0.534    |

Appendix Table 40: Evaluating single-cell FMs in cross-omic retrieval setting (Vote-Acc) on 10x multi-omics ( $K = 100$ ).  $K$  indicates the number of cells retrieved for each query cell.

| Setting        | scRNA->scATAC |         |          | scATAC->scRNA |         |          |
|----------------|---------------|---------|----------|---------------|---------|----------|
| Method         | Recall        | Avg-Acc | Vote-Acc | Recall        | Avg-Acc | Vote-Acc |
| PCA            | 0.012         | 0.159   | 0.224    | 0.083         | 0.412   | 0.506    |
| CellFishing.jl | 0.047         | 0.444   | 0.598    | 0.115         | 0.528   | 0.624    |
| scVI           | 0.016         | 0.197   | 0.214    | 0.012         | 0.199   | 0.217    |
| LDVAE          | 0.020         | 0.217   | 0.245    | 0.012         | 0.169   | 0.190    |
| CellBlast      | 0.140         | 0.517   | 0.587    | 0.128         | 0.588   | 0.654    |
| scFoundation   | 0.171         | 0.663   | 0.795    | 0.136         | 0.618   | 0.689    |
| scGPT          | 0.093         | 0.491   | 0.664    | 0.063         | 0.479   | 0.547    |
| SCimilarity    | 0.127         | 0.554   | 0.722    | 0.093         | 0.497   | 0.539    |
| UCE            | 0.162         | 0.679   | 0.845    | 0.138         | 0.628   | 0.705    |
| Geneformer     | 0.023         | 0.215   | 0.302    | 0.012         | 0.168   | 0.196    |
| scMulan        | 0.078         | 0.493   | 0.619    | 0.064         | 0.456   | 0.553    |
| CellPLM        | 0.115         | 0.555   | 0.686    | 0.109         | 0.467   | 0.542    |

Appendix Table 41: Dataset statistics

|                | Dataset              | Number of cells | Number of unique cell types |
|----------------|----------------------|-----------------|-----------------------------|
| Multi-platform |                      |                 |                             |
|                | PBMC                 | 31021           | 10                          |
|                | Human Pancreas       | 6321            | 13                          |
| Multi-species  |                      |                 |                             |
|                | adipose              | 5705            | 3                           |
|                | kidney               | 5429            | 2                           |
|                | large intestine      | 10622           | 3                           |
|                | liver                | 8087            | 5                           |
|                | lung                 | 39231           | 20                          |
|                | skin of body         | 910             | 2                           |
|                | spleen               | 12114           | 5                           |
|                | subcutaneous adipose | 3422            | 3                           |
|                | thymus               | 14627           | 5                           |
|                | trachea              | 6924            | 5                           |
| Multi-omics    |                      |                 |                             |
|                | 10X PBMC             | 9631            | 19                          |
|                | Chen-2019            | 9190            | 22                          |
|                | Ma-2020              | 32231           | 22                          |