OXFORD

Briefings in Bioinformatics, 2023, **24(4)**, 1–16
https://doi.org/10.1093/bib/bbad254
Advance access publication date 14 July 2023
Problem Solving Protocol

Deep transfer learning for clinical decision-making based on high-throughput data: comprehensive survey with benchmark results

Muhammad Toseef, Olutomilayo Olayemi Petinrin, Fuzhou Wang, Saifur Rahaman, Zhe Liu, Xiangtao Li and Ka-Chun Wong

Corresponding authors. Xiangtao Li, Professor in the School of Artificial Intelligence, Jilin University, Jilin, China. Email: lixt314@jlu.edu.cn; Ka-Chun Wong, Associate Professor at City University of Hong Kong, Hong Kong SAR. Email: kc.w@cityu.edu.hk

Abstract

The rapid growth of omics-based data has revolutionized biomedical research and precision medicine, allowing machine learning models to be developed for cutting-edge performance. However, despite the wealth of high-throughput data available, the performance of these models is hindered by the lack of sufficient training data, particularly in clinical research (*in vivo* experiments). As a result, translating this knowledge into clinical practice, such as predicting drug responses, remains a challenging task. Transfer learning is a promising tool that bridges the gap between data domains by transferring knowledge from the source to the target domain. Researchers have proposed transfer learning to predict clinical outcomes by leveraging pre-clinical data (mouse, zebrafish), highlighting its vast potential. In this work, we present a comprehensive literature review of deep transfer learning methods for health informatics and clinical decision-making, focusing on high-throughput molecular data. Previous reviews mostly covered image-based transfer learning works, while we present a more detailed analysis of transfer learning papers. Furthermore, we evaluated original studies based on different evaluation settings across cross-validations, data splits and model architectures. The result shows that those transfer learning methods have great potential; high-throughput sequencing data and state-of-the-art deep learning models lead to significant insights and conclusions. Additionally, we explored various datasets in transfer learning papers with statistics and visualization.

Keywords: transfer learning, domain adaptation, health informatics, clinical decision-making, cross-species methods

INTRODUCTION

With the advent of high-throughput sequencing (HTS) technologies, the field has been transformed significantly, which has made it possible to gather vast amounts of data at the single-cell level [1–3]. In particular, high-throughput molecular data are generated from omics-based technologies, such as genomics, proteomics, transcriptomics, epigenomics and metabolomics [2, 4]. With the help of multi-omics data, classical machine learning and deep learning methods made significant advances in biomedical research and precision medicine, overcoming key problems that were previously considered a big challenge [5]. High-throughput data analysis can be broadly categorized as whole genome sequencing (WGS), whole exome sequencing (WES), RNA-seq, ChIP-seq and microarray, all of which help stakeholders in the diagnosis, prognosis, classification and treatment of human and animal diseases. We are not discussing these sequencing methods in detail, but high-throughput molecular data analysis can answer critical questions, such as the genomic locations of mutations responsible for a particular disease [6, 7].

Although the availability of biomedical data has facilitated the development of state-of-the-art machine learning models, however, one of the assumptions in machine learning is the availability of sufficient training data for robust model performance. Unfortunately, in some cases in human health informatics, machine learning models performance is hindered by the unequal

Muhammad Toseef is currently a PhD student in the Department of Computer science, City University of Hong Kong, Hong Kong SAR. His research interests include bioinformatics, transfer learning and computational biology.
Olutomilayo Olayemi Petinrin is currently a PhD student in the Department of Computer science, City University of Hong Kong, Hong Kong SAR. Her research interests include medical informatics, chemoinformatics and deep learning.
Fuzhou Wang is currently a PhD student in the Department of Computer science, City University of Hong Kong, Hong Kong SAR. His research interests include 3D Genomics, epigenetics and machine learning.
Saifur Rahaman is currently a PhD student in the Department of Computer Science, City University of Hong Kong, Hong Kong SAR, and a Visiting Graduate student at the Broad Institute of MIT and Harvard, Cambridge, MA, United States. His research interests include cancer genomics, cancer detection, applied machine learning to bioinformatics and computational intelligence.
Zhe Liu is currently a PhD student in the Department of Computer Science, City University of Hong Kong, Hong Kong SAR. Her research interests include cancer genomics, bioinformatics and deep learning.
Xiangtao Li is a professor in the School of Artificial Intelligence, Jilin University, Jilin, China. His research interests include bioinformatics, computational biology and evolutionary data mining.
Ka-Chun Wong assumed his duty as an associate professor at City University of Hong Kong, Hong Kong SAR. His research interests include bioinformatics, computational biology, evolutionary computation, data mining, machine learning and interdisciplinary research. He is merited as the associate editor of Bio Data Mining in 2016. In addition, he is on the editorial board of Applied Soft Computing since 2016. Remarkably, he has solely edited two books published by Springer and CRC Press, attracting 30 peer-reviewed book chapters around the world.
Received: March 16, 2023. **Revised:** June 4, 2023. **Accepted:** June 20, 2023

distribution of training data, creating data discrepancies across domains [8, 9].

In vivo experiments for humans are challenging due to ethical considerations [10] or limited availability of tissues [11, 12], making it difficult to obtain data. To overcome these challenges, novel deep transfer learning methods have been developed, achieving breakthrough results in knowledge transfer from source domain to target domain using high-throughput data [13–16]. These studies mostly used preclinical data (Mus Musculus) in the source domain to transfer the learned knowledge and representations to the clinical (Homo sapiens) target domain.

Taxonomy of transfer learning algorithms

Transfer learning has been adapted with different names, such as knowledge transfer, multi-task learning, domain adaptation, life-long learning and context-sensitive learning [17]. Transfer learning helps to transfer the learned representation or features from a source domain to a predictive function in the target domain considering the following three points: (i) what to transfer, (ii) when to transfer and (iii) how to transfer [18]. The first question '*what to transfer*' significantly sorts out the critical information about the transfer learning task [19]. With the condition of labeled or unlabeled training data in the source and target domains, and whether the source and target tasks are related or different, transfer learning algorithms can be defined into three major paradigms [18]. The three main scenarios of transfer learning on the basis of source and target domains tasks and data availability are: (i) unsupervised transfer learning, (ii) transductive transfer learning and (iii) inductive transfer learning. The source and target domains, or source and target tasks must have shared denominators to transfer the knowledge to the predictive function in the target domain. Inductive and transductive transfer learning use labeled data in the source domain, while unsupervised transfer learning helps the clustering and dimensionality reduction tasks with no labeled data either in source or target tasks. In inductive transfer learning method, we have labeled training data in target domain. Based on labeled training data availability in source domain, inductive transfer learning can be further categorized into two sub-domains: multi-task learning and self-taught learning. On the other hand, transductive transfer learning helps to solve the problems with labeled training data only in the source domain. In particular, domain adaptation is an important sub-domain of transductive learning where we have the same task but in different domains.

Importance of deep transfer learning in human health informatics

Deep transfer learning has become a crucial tool in human health informatics due to its ability to leverage existing knowledge to improve the accuracy and reliability of computational models. In particular, it has shown great potential in addressing the challenges of limited training data and data inequality in biomedical research and clinical practice. The use of deep learning methods in *in vivo* experiments remains challenging due to data scarcity and financial constraints; Ravi *et al.* [20] outlined the problem of limited disease-specific data.

One of the key benefits of transfer learning is its capability of knowledge transfer from one task (with enough training data) to another (with limited data). This allows the reuse of previously learned features, resulting in more reliable, accurate and efficient models. For human disease informatics, it is critical to train the models with accurate data to transfer the shared denominators to the predictive function [21, 22]. Transfer learning has shown great impact in clinical research and health informatics in recent years for image-based models, time series data, text, tabular and audio-based applications [23].

Another important application of deep transfer learning in human health informatics is the ability to overcome the challenges of data inequality. This is particularly important for rare diseases and multiple ethnic races [8], with limited data available for training computational models. Deep transfer learning can help to overcome this challenge by leveraging existing knowledge and data from related diseases or populations [24], improving the accuracy and reliability of models. The COVID-19 pandemic has highlighted the importance of deep transfer learning in human health informatics. During the pandemic, deep transfer learning has been used to develop computational models for image-based diagnosis [25–28], predicting disease severity and patient outcomes and developing personalized treatment recommendations [29].

TRANSFER LEARNING STUDIES

In this section, we have explained the survey papers according to the sub-paradigms of transfer learning (Pan and Yang [18]). We have summarized transfer learning methods covered in our survey in Table 1, and statistics of published high-throughput studies with machine learning methods in the last five years has been shown in Figure 1. Furthermore, we have shown the word cloud representation from the abstracts of published articles covered in our survey, as shown in Figure 2. The word cloud shown in Figure 2 is generated with the top 30 words; and the word size represented the frequency of any word. For article search and selection, we followed the standard Systematic Literature Review (SLR) steps including (i) research questions formulation, (ii) search strategy, (iii) article selection, (iv) quality assessments and (v) data extraction (as shown in Supplementary Materials section 1). We searched from four biomedical databases including Web of Science (WOS), PubMed, Medline and Scopus with these keywords: gene expression, single-cell RNA sequencing (scRNA-seq), RNA-seq, transcriptomics, domain adaptation and transfer learning. A detailed explanation has been given in Supplementary Materials.

Unsupervised transfer learning

Feature representation such as latent variables (LV) is one of the answers of '*what to transfer*' in transfer learning. In similar settings of unsupervised transfer learning, Taroni *et al.* [24] presented the pathway-level information extractor method called MultiPLIER. The main focus of the study is to learn predictive function (to identify perturbed molecular processes across different organs systems) in the target domain for rare human diseases. The authors utilized a large compendium of publicly available gene expression dataset, Recount2 [30]. The feature representations of correlated genes were calculated with matrix factorization [31], to find the associated patterns of LV with known pathways. The learned features from the source domain were then projected onto the target domain of rare disease scRNA-seq data. The authors employed the PLIER [31] R package and the limma (Linear Models for Microarray Data) Bioconductor package for differential expression analysis. They selected subsets of the recount2 data, ranging from 500 to 32 000 samples, to train the PLIER. The authors observed that training MultiPLIER with a larger sample size led to better performance. The limitations of MultiPLIER include the potential for limited generalizability to specific tasks and datasets, as well as the need for careful evaluation of its performance. Additionally, it may not capture all relevant biological information because of the relatively small dataset in the target domain.

Table 1. Recent Transfer Learning studies for high-throughput molecular data for human disease informatics

Method	Description	Pros	Cons	References
MultiPLIER	The authors trained an unsupervised TL model to identify perturbed molecular processes in complex human disease across organ systems. They used a large publicly available dataset Recount2, for pre-training of transfer learning model in the source domain. (Software is available at https://github.com/greenelab/multi-plier)	It can be useful in case of rare diseases where data are too limited	It may not capture all information because of small target domain datasets	Cell Systems [24]
projectR	The authors used the scRNA datasets to define latent spaces in the source domain with mouse retina data. The proposed method projectR used transfer learning and evaluate the latent spaces in source domain using human retina data. (Software is available at https://github.com/genesofeve/projectR)	This method may help in new cell annotation, cross-species analysis, and linking genomic regulatory and transcriptional signatures	The efficacy of the projection method needs further evaluation in other orthogonal and non-orthogonal projection methods	Cell Systems [15]
pro-jectR_ICI	To identify transcriptional changes in tumors across different datasets via immunotherapy responses. The authors found the NK cell activation in mouse and human tumors in anti-CTLA-4 treatments (Data and software is available at https://github.com/edavis71/projectR_ICI)	The proposed matrix factorization can detect the signature of NK cell activation without any need for clustering, differential expression analyses, or additional technologies in response to treatments	Expression of CLTA-4 in NK cells is disputed in both mouse and human cells and further explanations are required	Genome Medicine [16]
FIT	The authors created 170 CSPs from disease and control datasets (for 28 diseases) to predict human effect size from mouse effect size and then used these predictions to find human gene effect from target data in the target domain. (Software is available at https://github.com/shenorrLab/FIT.mouse2man)	As compared with the direct exploration of mouse data, FIT can detect 20-50% more relevant genes	The training data compendium should have a large training set of human data for a specific disease(s)	Nature Method [32]
CaSTLe	The proposed model is used for the classification of single cells from scRNA-seq datasets using transfer learning. The model was trained on previously labeled data by selecting informative features. The classification task was done using the XGBoost classifier. (Data and software is available at https://github.com/yuvallb/CaSTLe)	The main strength of this study is the scalability and enhanced performance on larger and imbalance datasets as compared with benchmark models	Cannot detect novel cell types, technical variability between datasets	PLOS ONE [37]
TF-Binding-Matrix	A two-step (pre-training and fine-tuning) transfer learning for TF binding using a CNN model (Software is available at https://github.com/wassermanlab/TF-Binding-Matrix)	As compared with previous methods, this method is trained the TF with the same binding mode using transfer learning	It may not be computational effective and the model performance can still be optimized	Genome Biology [33]
scJoint	This study adapted the domain adaptation task in transfer learning to transfer label knowledge from the source domain (RNA-labeled data) to the target domain (scRNA-seq, ATAC-seq unlabeled data). (Software is available at https://github.com/SydneyBioX/scJoint)	Proven more effective than previous studies, even in the case of highly complex data where important biological information was mixed with technical variations	If extended to other epigenomic data, a separate encoder development will be required	Nature Biotechnology [14]
BERMUDA	This study presented a novel batch effect removal using deep autoencoders using transfer learning for multiple batches of scRNA data from various cell populations. This helped to transfer information among these batches by amplifying batch signals. (Software is available at https://github.com/txWang/BERMUDA)	BERMUDA can effectively remove batch effects even in cases across vastly different batches among cell populations	Clustering algorithm is not robust and only tested on small-scale scRNA-seq datasets	Genome biology [13]
trVAE	The transfer of conditions across domains is achieved by adapting a transfer variational autoencoder. The trVAE has been adapted to two sets of problems: smiling condition images for males and females and cell type infections in source and target domains. (Software is available at https://github.com/theislab/trvae)	It can be useful and applicable to multimodal methods in biomedical research	Two gene sets from different species in source and target domains	Bioinformatics [38]
XGSEA	CROSS-species gene set enrichment with a three-step transfer learning approach was proposed. They used five regression and one classification method to evaluate the model performance using four datasets in source and target domains. (Software is available at https://github.com/LiminLi-xjtu/XGSEA)	It can provide more robust and focused results as compared with other approaches for the prediction of enriched pathways	They used traditional machine learning methods for regression and classification tasks, which may cause model underperformance	Briefings in Bioinformatics [39]

(continued)

Table 1. Continued

Method	Description	Pros	Cons	References
PRECISE	This is a domain adaptation transfer learning study to transfer the drug response predictors from pre-clinical models to human tumors. To train the predictor, they used the GDSC1000 dataset, while human datasets were adapted from TCGA, using breast and multiple melanoma cancers. (Software is available at https://github.com/NKI-CCB/PRECISE)	Performance comparison of known biomarkers in skin cancer and breast cancer showed a strong association between biomarkers and relevant drugs	Only available for gene expression data, so a multi-omics approach could fill this gap	Bioinformatics [40]
AITL	An adversarial inductive transfer learning method for input and output space adaptation for pharmacogenomics. The model consisted of a multi-task deep learning model to address the source domain discrepancies and predicted the target domain's drug response. (Software is available at https://github.com/hosseinshn/AITL)	The first approach of adversarial and inductive transfer learning by adapting both input and output spaces and outperforming recent state-of-the-art methods	The pharmacogenomics patient data for drug responses is not enough, and they only consider the gene expression data	Bioinformatics [41]
ssNN	A semi-supervised transfer learning for mouse-to-human genomic insight translation for 36 human disease transcriptomics case studies. (Software is available at https://ww2.mathworks.cn/matlabcentral/fileexchange/69718-semi-supervised-learning-functions)	successfully predicted human pathways and phenotype associated-genes for inter-species molecular translation, with no need for predicted humans labels	Mostly suitable for regression problems;	PLoS Computational Biology [36]
TransComp-R	Cross-species knowledge transfer, additionally with different omics types, such as transcriptomics to proteomics. (Software is available at https://ww2.mathworks.cn/matlabcentral/fileexchange/77987-transcompr)	Transfer knowledge from one omic data to a different omic data	The main issue is that is only applicable to homologous mouse-to-human proteins/genes	Science Signaling [42]
scDEAL	Prediction of drug response in high throughput data using deep transfer learning, with a domain-adaptive neural network using bulk RNA data as source data to learn and predict drug responses in the target domain. (Software is available at https://github.com/OSU-BMBL/scDEAL)	Intelligent model structure to maintain the heterogeneity of single cell while training the model using scRNA and bulk RNA data	Because of the unavailability of drug-treated mouse data, it is difficult to evaluate and optimize the cross-species model reliability	Nature Communications [43]
CaSee	pan-Cancer Seeker (CaSee) is proposed to discriminate normal and cancer cells in scRNA data. They trained the model on 18 types of pan-cancer bulk RNA-seq data. For training purposes, they adapted a capsule network with transfer learning. (Software is available at https://github.com/yuansh3354/CaSee)	Shown better performance against copy number variations (CNVs) and other existing methods, and successfully differentiated tissues, cell lines source, and xenograft cells	Feature space and sample needed to user selected before model training	Oncogene [44]

In general, analysis of scRNA-seq to learn the meaningful representations is a challenging task because of the low-dimensional latent space. Recently, Stein-O'Brien *et al.* [15, 16] developed projectR and scCoGAPS to address this issue using transfer learning. Firstly, they used the CoGAPS (Coordinated Gene Activity in Pattern Sets) package from Bioconductor to learn the latent spaces from sparse scRNA-sec data of mouse retina. The learned projections were transferred to the target task using both scRNA and bulk RNA datasets for predictions in human data. The target domain consisted of multiple datasets, including developing (age, cell type, sex, disease status) the human brain cortex and developing mouse midbrain, datasets details are provided in Table 2. In the source task, latent spaces were learned using dimensionality reduction with UMAP (Uniform Manifold Approximation and Projection) (https://umap-learn.readthedocs.io/en/latest/). They implemented the software using R and Scanpy (https://scanpy.readthedocs.io/en/stable/). Two years later, Davis-Marcisak *et al.* [16] developed an unsupervised transfer learning extended study of projectR, projectR-ICI, to identify the nature killer (NK) cells activation in human tumors using a source domain task for anti-CTLA-4 treatment respondents. The authors chose pre-clinical (mice-scRNA) as source domain data and clinical (human-scRNA, bulk RNA, cyTOF) datasets for target domain data. Gene regulations and cell types were learned with matrix factorization using UMAP in source task. To transfer these feature representations to the target predictive function, they used projectR using an independent human tumor dataset in the target task. The authors found positive results with NK cell activation in source mouse data treated with anti-CTLA-4 and then validated it in human tumor (metastatic melanoma) data in the target task.

Normand *et al.* (2018) [32] proposed a novel method called Found in Translation (FIT) to predict relevant genes associated with 28 human diseases (target domain), using both mouse RNA-seq data and human mouse-model disease-versus-control datasets (source domain). The authors collected the source and target datasets from Gene Expression Omnibus (GEO) datasets, including microarray and RNA data, to create 170 cross-species pairings (CSPs). For each CSP, they employed a lasso regression model to fit α and β parameters for all genes using a linear model. FIT used a manually annotated 170 microarray and RNA-seq datasets (for every GEO dataset, the authors considered each sub-dataset as a new dataset). To predict the translation

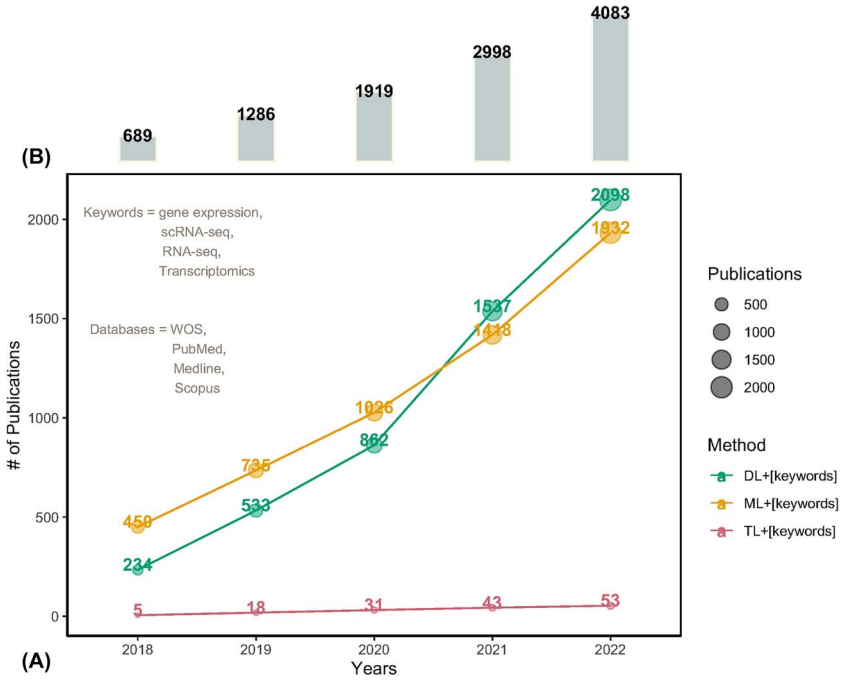

Figure 1. (**a**) A statistical representation of literature published in the last 5 years for gene-expression data with different computational methods; each search query has method name in the title and keywords in all fields. The overall results are retrieved from Web of Science (WOS), PubMed, Medline and Scopus (DL: deep learning, ML: machine learning, TL: transfer learning.) (**b**) a bar plot showing the total number of publications every year for all methods combined. (**Note**: The year 2022 includes the publications until December 2022 and overall comparison shows the gap significant gap between transfer learning studies and deep/machine learning works in the past 5 years)

Figure 2. A word cloud based on the published transfer learning papers for high-throughput data

from mouse to human, the authors trained a support vector machine (SVM) model, where principal component analysis (PCA) was performed on mouse gene expression data in the source domain. The classifier was trained using the hold-out strategy, where authors split the 80% data as training data and 20% data as test data. Furthermore, the first 50 principal components were provided as input to the SVM model, capturing more than 80% variation. The limitation associated with FIT is the quality of the reference compendium used to train the model. The authors compiled a dataset of 170 CSPs for 28 different diseases, but it is possible that these pairs are not representative of the full range of human diseases. In addition, the quality of the training data is dependent on the quality of the original studies, which may vary in terms of sample size, experimental design and data analysis methods.

A multi-task learning model is proposed by Novakovsky *et al.* [33] where authors used transfer learning to predict transcription factor (TF) binding sites in DNA sequences based on position weight matrices of the TF binding motifs. The authors adapted a two-step strategy of pre-training and fine-tuning for TF binding prediction in the target domain. The authors used the ChIP-seq peaks datasets from ReMap [34] and UniBind [35] studies as source data for TF binding events. To predict the TF binding prediction, a convolutional neural network (CNN) was adapted with three convolutional layers and two fully connected layers. The overall performance of two-step transfer learning was evaluated using the area under the precision-recall (AUCPR) curve. One of the limitations mentioned by the authors is the dependency on the quality and quantity of ChIP-seq data available for a particular TF. If there are only a few ChIP-seq peaks available, the model may not perform well. The authors suggested several steps to improve the performance of the model, including pre-training a larger multi-model with representative TF from each binding mode and fine-tuning the model with different learning rates.

After 1 year of the previous study in 2019, snNN [36], Brubaker *et al.* proposed a new transfer learning method Trans Comp-R, for the prediction of treatment (with infliximab) resistance to inflammatory bowel disease (IBD). The proposed study is based on the previous attempt by the authors for the transfer of cross-species knowledge. Additionally, this method added the novelty of the adaption of one data space to another, such as transcriptomics to proteomics. In the source domain, they selected labeled human transcriptomics data to learn the corresponding mouse proteomics data for responsive or non-responsive phenotypes in humans. During the first training step, the human gene expression data were provided as input to find the associated genes for the responder phenotype. After this training step, they provided the mouse proteomics data for homologous human genes found in the previous step. Then they performed the PCA on these genes. The next step was to project the human transcriptomic data into the PCA space to perform a regression task against the human responder phenotype. This prediction enabled the finding of new mouse proteins to help human phenotype prediction. After these experiments, the model predicted a collagen-binding integrin to be involved in resistance to treatment.

Table 2. Source and target domains datasets description

Method	Key datasets with organism		Reference
	Source domain	Target domain	
MultiPLIER	Large public data compendium comprising multiple experiments, tissues, and biological conditions. Figsharelink, https://github.com/greenelab/rheum-plier-data	Human (E-GEOD-65391, E-GEOD-11907, E-GEOD-49454, E-GEOD-39088, E-GEOD-72747, E-GEOD-78193, E-GEOD-61635, E-MTAB-2452, GSE119136, GSE104948, GSE37382, GSE37418)	[24]
projectR	Mouse GSE118614 https://github.com/gofflab/developing_mouse_retina_scRNASeq	Human GSE104827, GSE104276, Mouse, Human, and Stem Cells GSE76381	[15, 51]
projectR ICI	Mouse (GSE119352)	Human (GSE120575, GSE139249)	[16, 52]
FIT	Human and Mouse (GitHublink)	Human (Microarraysample, RNAseqsample)	[32]
CaSTLe*	Mouse (GSE59114, GSE81682), (Mouse and Human (GSE63473))	Human (EMTAB5061 GSE81608)	[37]
TF-Binding-Matrix	Human (GitHublink)	Human (Alldata)	[33]
scJoint	Mouse: scRNA https://tabula-muris.ds.czbiohub.org/, sci-ATAC-seq Atlas https://atlas.gs.washington.edu/mouse-atac/	Human (scRNAGSE156793, sci-ATAC-seq3GSE149683)	[14]
BERMUDA	Human and Mouse GSE84133	Human (GSE85241, E-MTAB-5061, PBMC PBMC10 xGenomicssupport), Human and Mouse GSE84133,	[13]
trVAE	Human (Alldatasets)	Human (Authorspreprocesseddatasetsfromotherstudies)	[38]
XGSEA	Mouse (Embryonic development (GSE44183), Brain cancer (GSE38591), Ovarian cancer (GSE5987), Zebrafish (Melanomas GSE83399))	Human (Embryonic development (GSE44183), Brain cancer (GSE45874), Ovarian cancer (GSE6008), Melanomas (GSE83343))	[39]
PRECISE	Mouse (The cell lines dataset GDSC1000, ThePDXdataset)	Human (TCGA [53, 54])	[40]
AITL	Human (GDSCCelllines)	Human (GSE55145, GSE9782-GPL96, GSE18864, GSE23554, GSE25065,PDX [55] TCGA [56])	[41]
ssNN	Mouse (GSE7404, GSE7404, GSE26472) Human and Mouse (for augmented training set GSE5663)	Human (GSE37069, GSE36809, GSE3284, GSE13904)	[36]
TransComp-R	Mouse GSE95705 [57]	Human (GSE16879)	[42]
scDEAL	Human (GDSC, CCLEcelllineexpressionprofile)	Human (GSE117872, GSE112274, GSE140440, GSE140440 GSE149383), Mouse (GSE110894)	[43]
CaSee	Human (TCGA & GTEx)	Human (GSE116237, GSE150949, HumanCellLandscape(HCL))	[44]

*The authors used dataset pairs with source and target datasets, where they trained the model twice with same pair, once using dataset A and source data

Transductive transfer learning

A transfer variational autoencoder (trVAE) [38] is a transfer learning approach proposed for the transfer of conditions across different domains. The proposed method is motivated where the target domain of interest does not offer training data for a certain condition. It suggested using Maximum Mean Discrepancy (MMD) regularization to produce a more compact representation of a cross-condition distribution that would otherwise display high variance in the standard conditional variational autoencoder (CVAE), leading to more accurate out-of-distribution (OOD) prediction. The goal is to generate new gene expression profiles that are conditional on a categorical variable (such as cell type) and a latent vector and to handle OOD scenarios where the conditioning variable is not present in the training data. The trVAE was evaluated on several benchmark datasets (scRNA-seq), including a mouse brain dataset and a human bone marrow dataset. To evaluate the model performance, the authors benchmarked the proposed approach against some standards methods such as CycleGAN [45], CVAE [46], MMD-CVAE [47], MMD-regularized autoencoder [48], scVI [49] and scGen [50]. They used Pearson's correlation values for gene expression variance and mean and showed that trVAE showed better performance as compared with other published methods. However, training the model can be computationally expensive and it may not be scalable to extremely large datasets, which can be improved by adapting more robust model architecture. Furthermore, the model performance is highly dependent on high-quality annotations for the cell types, which may not always be available.

Gene set enrichment analysis (GSEA) study was done by Cai *et al.* [39], where the authors proposed XGSEA (cross-species GSEA); a domain adaption model to predict enrichment significant for phenotype analysis. They used four gene expression datasets including embryonic, brain, ovarian and melanomas, for both source and target species (mouse and human). The XGSEA method solves this issue in three steps: firstly, running GSEA over source gene sets; secondly, using pairwise similarities among gene sets based on MMD and domain adaptation to project gene sets from two species into a common latent space; thirdly, training a regression model using to predict enrichment scores and *P*-values for target gene sets. Overall, the proposed domain adaptation model used MMD to project source and target gene sets into a common latent space with affine mappings, and then a regression model was trained on the source gene sets to predict enrichment scores for the target genes set. They trained the XGBoost classification model using 80% of training data, with 20% held-out data. The authors compared the model performance

with multiple naïve methods using four datasets, including three mice to humans and one zebrafish to human. XGSEA outperformed all naïve methods in terms of AUROC. However, the model performance can be improved using more advanced classifiers. Furthermore, the proposed method can be improved in further directions, such as with robust domain adaptation model by developing a more comprehensive null hypothesis for the gene set enrichment score.

Following a domain adaption transductive transfer learning approach [40], the authors used pre-clinical data (such as cell lines and patient-derived xenografts) to transfer the predictions of drug responses for the human tumor data in the target task. The ability to predict the response of individual patients to anti-cancer drugs is a critical step toward personalized medicine. However, building reliable predictive models for drug response is challenging due to the heterogeneity of tumors and the limited availability of patient data. To overcome these challenges, the authors proposed a domain adaptation approach called PRECISE (Patient Response Estimation Corrected by Interpolation of Subspace Embeddings) that uses pre-clinical models (transcriptomic data) to transfer knowledge to tumors. First, they used PCA to find the common factors in pre-clinical models and human tumors data. An additional step was performed to find the consensus representation using principal vectors generated in the previous step. They adapted a Ridge regression to build the model and trained the model using the consensus representation to predict the drug responses in human tumor data. However, they only used the proposed study for transcriptomics data and it may not be applicable to other types of predictors or omics data. This can be further enhanced by integrating other types of omics data, such as epigenetic or proteomic data.

Inductive transfer learning

Cell labeling in scRNA-seq data is done by cell clustering or fluorescence-activated cell sorting, where both models have some limitations [37]. To address these issues, a transfer learning cell label classification model CaSTLe (Classification of single cells by transfer learning) is proposed by Lieberman *et al*. [37]. For transfer learning, six dataset pairs were created to train the model. They trained an XGBoost classifier after the selection of common genes across all single-cell datasets. For the multi-class scenario, CaSTLe outperformed both a simple benchmark of highest mean features and linear model classification and a more sophisticated benchmark, the beta-Poisson single cell differentially expressed genes and linear model classifier, in most cases. For the binary-class scenario, CaSTLe achieved high performance with AUC values above 95% for 16 cell types and a sensitivity higher than 97% for all 15 cell types that appeared only in the source dataset. The high accuracy levels achieved for larger and more imbalanced datasets demonstrated the method's strength and robustness. The CaSTLe has some limitations where it cannot detect novel cell types, and it requires that the source and target datasets are similar for the method to replace clustering effectively. Another limitation is the technical variability between datasets requires a more sophisticated approach for transfer learning, which can be improved in future research. CaSTLe has not been tested on transfer classification where the target dataset is only partially labeled; this is also a promising research direction that could potentially improve classification accuracy.

Lin *et al*. [14] proposed a domain adaptation transfer learning approach, scJoint, for the integration of atlas-scale single-cell RNA-seq and ATAC-seq (Assay for Transposase-Accessible Chromatin) data using a neural network architecture. This semisupervised transfer learning method learned the labels from the multiple source datasets and transfer these learned representations to ATAC-seq data in the target task. The proposed model was trained in three steps: (i) joint dimension reduction of scRNA and scATAC-seq data, (ii) cell label transfer using K-nearest neighbors and (iii) joint training with transferred cell labels from Step 2. The authors evaluated the scJoint from three different perspectives: (i) joint embedding evaluation, (ii) transferred label accuracy and (iii) evaluation with run time. For joint embedding evaluation, they calculated the Silhouette coefficient for each cell with two groups: modality silhouette coefficient and cell-type silhouette coefficient. They also trained the scJoint with complete scRNA-seq and scATAC-seq data, with 433 695 and 656 074 cells, respectively. Additionally, while scJoint is applicable to paired data, it has been designed for unpaired data, and adapting it to paired data during training could potentially enhance its performance on this type of data. Although the results are stable with respect to the choice of cosine similarity loss (main tuning parameter), other optimization details, such as the number of hidden nodes in the architecture, can also be considered tunable.

The batch effect can be defined as variations in the generated high-throughput data due to multiple factors [58], including technical variations and different experimental conditions, which can ultimately result in inaccurate and inconclusive findings. To address this issue, Wang *et al*. [13] proposed an unsupervised deep transfer learning approach called BERMUDA (Batch Effect ReMoval Using Deep Autoencoders) that utilizes scRNA-seq data from various batches and different cells. BERMUDA leverages an autoencoder to remove batch effects across different batches by identifying similar clusters in input data (multiple batches) and aligning cell populations. The authors used scRNA-seq data in the source task to find low-dimensional data representations. A graph-based clustering algorithm was then applied to different types of cells among multiple batches (figure 1 in [13]), followed by the application of a Spearman correlation-based method called MetaNeighbor to identify similar clusters among different batches. Once clusters were identified, an autoencoder was trained on unaligned cell clusters. UMAP visualization was used to cluster cell types from different batches. To evaluate the performance of BERMUDA, the authors compared it with various methods, including Seurat v2 [59], Seurat v3 [60], scVI [49], mnnCorrect [61] and BBKNN [62], using four scRNA datasets, each with two batches. However, some limitations of the proposed model include the use of more advanced clustering algorithms, as KNN may not handle large scRNA datasets with speed and accuracy, such as [63]. Future research may address this limitation.

Sharifi-Noghabi *et al*. [41] proposed a novel transfer learning method 'Adversarial Inductive Transfer Learning (AITL)' to address the output and input space discrepancies between pre-clinical and clinical datasets. AITL consists of four components: a deep neural network with a feature extractor, a multi-task learning sub-network and discriminators to reduce domain discrepancy using adversarial learning (further explained in Section 3). The authors evaluated the model performance using the area under the precision-recall curve (AUPR) and area under the receiver operating characteristic curve (AUROC) with state-of-the-art transfer learning methods (ProtoNet [64], ADDA [65]) and pharmacogenomics datasets (bladder, lung, kidney, breast and prostate cancer patients). The authors performed 3-fold cross-validation, where in source samples two folds were used as training and one fold as validation; similarly, in the target sample two folds were used as training and one fold as validation. Although the currently used cell lines datasets train machine learning models

for other cell lines or patient datasets [40, 66–68], they may not contain the identical distribution even with the same set of genes. Because of this data discrepancy in the source and target domain, an efficient computational model is difficult to train. Another limitation of AITL includes the small size of the patient datasets with drug response due to privacy and/or data sharing issues.

In a semi-supervised transfer learning model (transductive TL) [36], the authors proposed ssNN (semi-supervised neural network), for pathway and gene expression analysis using multiple mouse datasets in the source domain to human disease prediction (phenotype prediction) in the target domain. The source and target gene expression datasets were downloaded from GEO for inflammatory diseases. The source data have either 'sick' or 'healthy' labels, and they constructed 36 mouse-to-human pairs. They applied multiple machine learning methods such as SVM, KNN, Random Forest and neural networks. For the models' evaluation, the human label prediction performance was measured by precision and recall. The proposed method ssNN was first trained on labeled mouse data to predict the human labels from provided human gene expression data. After getting human labels, the authors get an augmented training dataset from mouse and human data, they used the human samples from the last step with the highest confidence. The authors adapted the retraining strategy on human data and get the new human labels with the highest confidence, then these samples were again used for augmented data with mouse expression data. They repeated the training loop until all human label data were used for augmented set generation. Overall, this strategy is only applicable to classification tasks, and it cannot be used for regression analysis.

After 1 year of the previous study in 2019, snNN, Brubaker *et al*. [36] proposed a new transfer learning method Trans Comp-R, for the prediction of treatment (with infliximab) resistance to IBD. The proposed study is based on the previous attempt by the authors for the transfer of cross-species knowledge. Additionally, this method added the novelty of the adaption of one data space to another, such as transcriptomics to proteomics. In the source domain, they selected labeled human transcriptomics data to learn the corresponding mouse proteomics data for responsive or non-responsive phenotypes in humans. During the first training step, the human gene expression data were provided as input to find the associated genes for the responder phenotype. After this training step, they provided the mouse proteomics data for homologous human genes found in the previous step. Then they performed the PCA on these genes. The next step was to project the human transcriptomic data into the PCA space to perform a regression task against the human responder phenotype. This prediction enabled finding new mouse proteins to help human phenotype prediction. After these experiments, the model predicted a collagen-binding integrin to be involved in resistance to treatment.

In a recent study by Chen *et al*. in 2022 [43], the authors developed scDEAL (single-cell Drug rEsponse AnaLysis) using deep transfer learning and adapting a Domain-adaptive Neural Network (DaNN) to predict drug response at the single-cell level. They used bulk RNA-seq data in source domain and used the trained model to predict responses in the target domain using single-cell data. The overall framework consists of five steps including (1) a denoising autoencoder (DAE) to extract the bulk features, (2) an encoder model to predict drug responses, (3) a second denoising autoencoder to extract single cells features, (4) DaNN used as a deep transfer learning model to derive the feature extractor at the single-cell level and (5) transfer the learned model and drug response prediction in single-cell data. The authors split the training data with hold-out strategy as 64%, 16% and 20% for training, validation and testing, respectively. They used precision, recall, F1-score, AUROC, AMI and ARI scores to evaluate the model performance. They trained scDEAL with nested 10-fold cross-validation; for each outer training fold, 90% data were used for training in inner 10-fold cross-validations. The authors adapted Ridge and ElasticNet regression in their model, and Pearson correlation was used to evaluate the predictor performance. The current domain adaptation model in scDEAL has been implemented using only transcriptomics data in the source and target domains, so the future direction of this study can be implemented with multi-omics data.

Another transfer learning application, pan-Cancer Seeker (CaSee), has been published in 2022 by Sh *et al*. [44]. The proposed model discriminates the normal and cancer cells using scRNA-seq as target domain data, while bulk RNA-seq as source domain data. They used a capsule network based on a 2D CNN for model training. The authors claimed this is the first transfer learning method for normal/cancer cell discrimination and has higher efficacy than other methods, such as copy number variation. First, they created a shared feature space from scRNA-seq and bulk RNA-seq data (candidate reference data) to generate output reference data to train the CaSee model. To train the model, the candidate reference data (bulk RNA-seq) were split into 80%, 10% and 10% for training, validation and testing, respectively. This candidate data count matrix was then fed to an encoder model with one fully connected layer with ReLU, and two Conv2d layers, then the output feature map was passed capsule encoder. To evaluate the CaSee performance, they applied the Wilcoxon test. The authors showed that the proposed model discriminated the normal and cancer cells with 96.69% accuracy. The one major shortcoming of this model is that candidate reference data genes are determined and unchangeable, so in case of low feature space in scRNA-seq data, the number of intersection genes remains low to create the output reference data.

We have summarized the source/domain data and high-throughput data type for transfer learning studies in Figure 3. Furthermore, the nature of transfer learning methods, strategy and evaluation methods have been shown in Figure 4. Moreover, we arranged the key information such as transfer learning task, source task and input, target task and output, machine learning methods and evaluation metrics in Table 6.

BENCHMARK RESULTS

We have benchmarked the selected transfer learning papers, including AITL [41], scJoint [14] and BERMUDA [13]. Firstly, we evaluated these papers with original model settings and then conducted experiments under different conditions, such as training data split, K-fold cross-validation, hyperparameter settings and various model architecture. We used the official source code for the selected studies and followed the instructions by the authors for each original experiment. The purpose of these experiments is to provide tutorials for transfer learning methods, as previous transfer learning reviews for high-throughput data have overlooked these analyses. The comparison source codes are given at the GitHub (https://github.com/mtoseef99/transfer-learning-for-geneExp) repository to follow.

In AITL[41], the authors proposed a space adaptation model for the pharmacogenomics adaptation from input to output space. We evaluated the performance of AITL with different sets of experiments; details have been shown in Supplementary Table

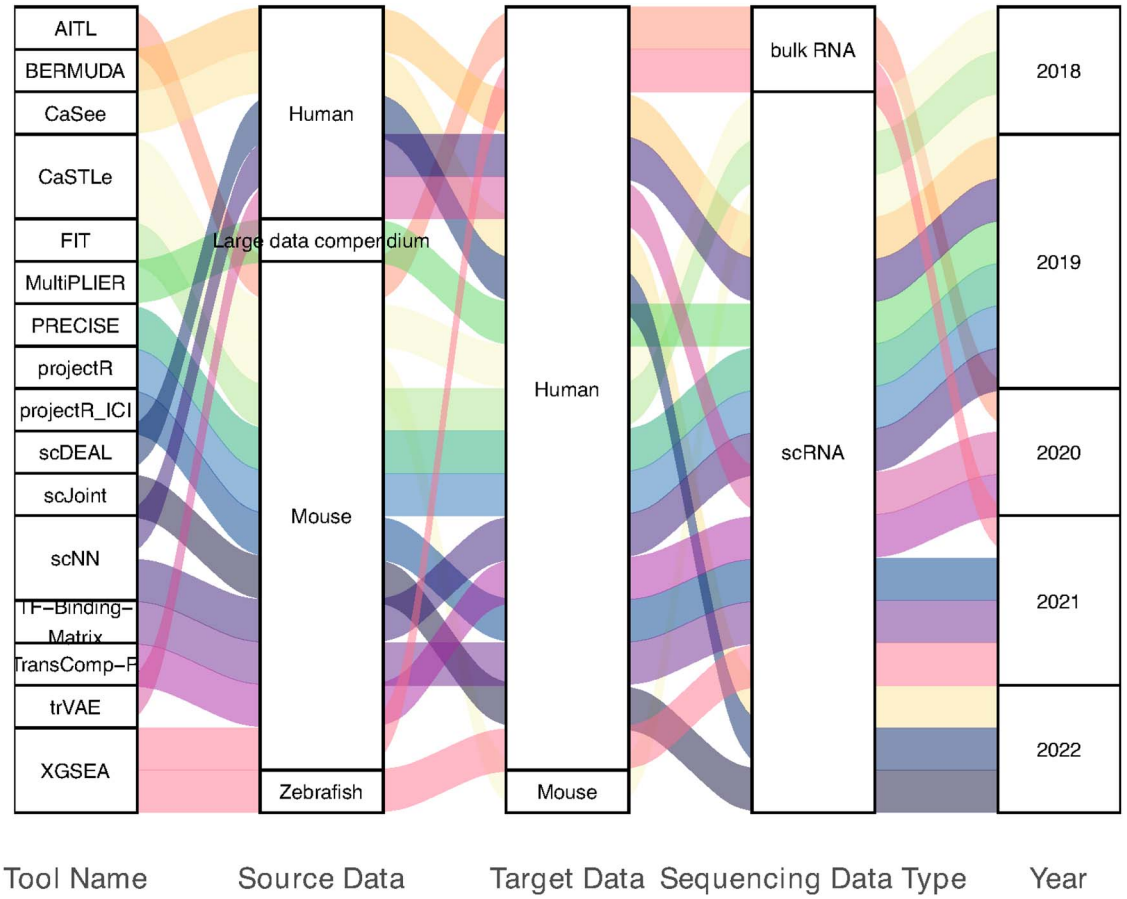

Figure 3. An alluvial plot showing the source and target data species for studies published in last 5 years, with the type of sequencing data: scRNA or bulk RNA (Large data compendium, recount2, contains data from multiple larger datasets, tissues and experiments from MultiPLIER. The original recount2 data compendium has RNA-seq and exon counts from 2041 different studies.)

S1. First, we downloaded the author's pre-processed data and created the new data splits for 5-fold and 10-fold for cross-validation experiments. Then we performed the experiments with the original model setting for both 5-fold and 10-fold splits, where authors only used 3-fold cross-validation in their experiments. The results for 3-fold, 5-fold and 10-fold for all four drugs have been shown in Table 3. While performing baseline experiments with new cross-validation settings, we observed the best results with different splits, such as Bortezomib with 10-fold, and Cisplatin and Docetaxel with 5-fold. After that, we benchmarked the original experiments with fine-tuned models, the details of the hyperparameters are given in Supplementary Table S1. Table 3 shows the average AUROC and ARP results for 15 epochs. We observed that the results with 5-fold and 10-fold cross-validation are better than baseline 3-fold cross-validation, even for the baseline model settings. It suggested that further better performance can be achieved with new cross-validation settings and changing the methodology.

For scJoint benchmarking experiments, we performed experiments for all three datasets, including 10xGenomics data, Mouse Primary Motor Cortex Data and CITE-seq and ASAP-seq PBMC datasets. The instruction for downloading and preparing datasets are given on tutorial GitHubpage of this paper. The results of scJoint benchmarking experiments have been shown in Table 4, where we have shown two experiments (baseline and fine-tuned) for each database. The details of the training configuration for each database are presented in Supplementary Table S2. For each database, the number of input and class size was set according to genes/proteins and cell types, such as common genes for input and cell types for classes. For 10x Genomics data, the input size was set 15463 (scRNA-seq and scATAC-seq common genes) and class size was set to 11 (cell types in scRNA_seq data), while for Mouse Primary Motor Cortex Data, the input size was 18 603 with 21 classes (RNA-seq data). In the same way, the input size for CITE-seq and ASAP-seq was set to 17 668 (17 441 genes and 227 proteins) with seven number of classes. We performed multiple experiments using different hyperparameter values, as shown in Supplementary Table S2; furthermore, the results for predicted cell types for the mouse primary motor cortex database using RNA-seq data have been shown in Supplementary Figure S3 (tSNE visualization) and S4 (UMAP visualization).

To evaluate the BERMUDA performance, we benchmarked the original experiments with the fine-tuned model, where we performed further experiments with different model architectures and hyperparameters settings. We performed experiments for the Human Pancreas dataset and PBMC dataset, as shown in Table 5. The authors used two types of model architectures in the baseline model, Encoder20 and Encoder2. Encoder20 has hidden units 200, 20, 200, and Encoder2 dimensions are 20, 2, 20. For each experiment, we trained both models for 2000 epochs, the run time for these experiments was 55 to 65 min. The reported results for the autoencoder and uncorrected model for both architectures are shown in Table 5; it is observed that Encoder20 has achieved better results for transfer loss in the fine-tuned model.

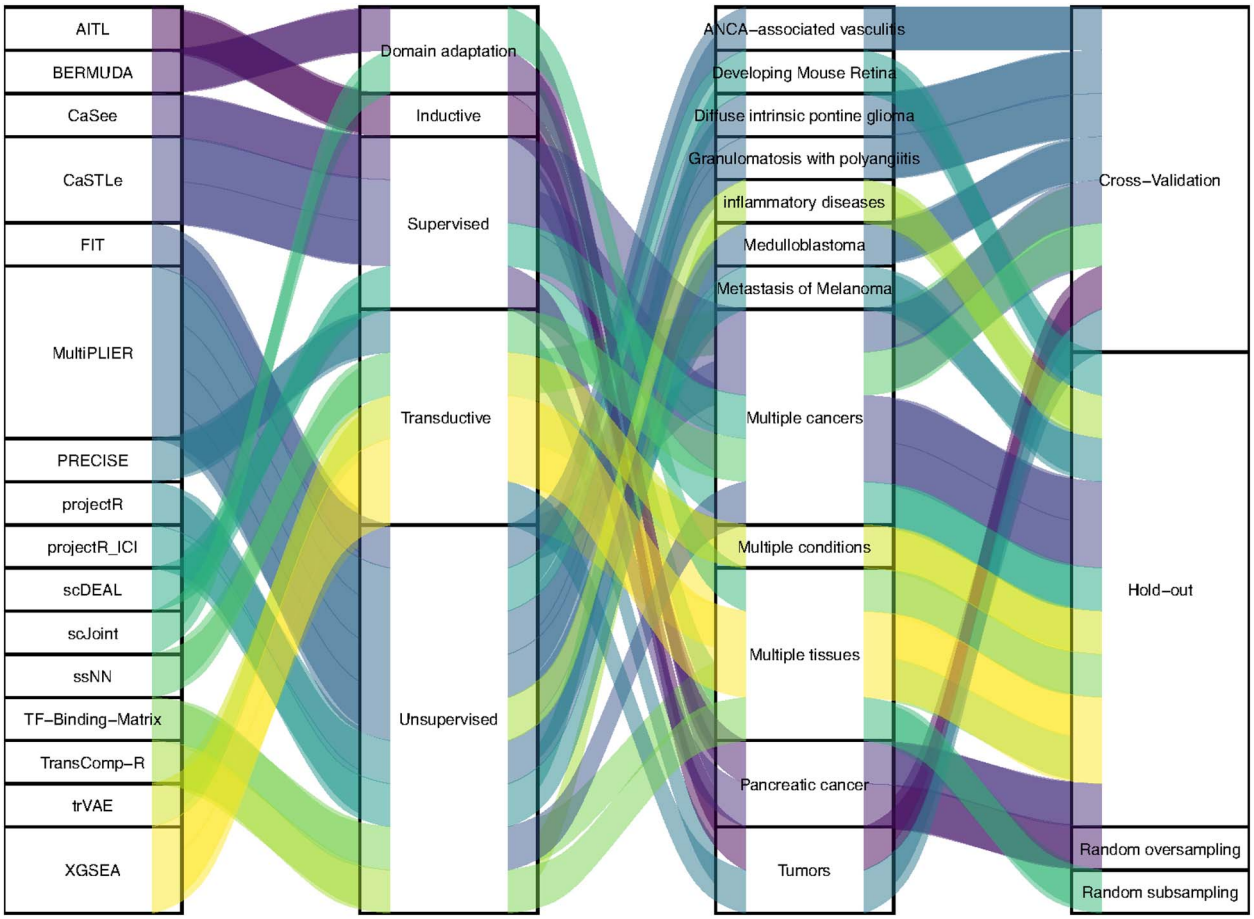

Figure 4. Reviewed methods based on applied transfer learning paradigm

Table 3. AITL experiments for 3-fold, 5-fold and 10-fold cross-validation (authors used only 3-fold for their experiments) for all four drugs Bortezomib, Cisplatin, Docetaxel and Paclitaxel with baseline model, best results in bold

Best Model for all drugs with 3-fold, 5-fold and 10-fold cross validation					
Drug	***Model***	***Cross-Validation***	***Epochs***	***Avg AUROC***	***Avg APR***
Bortezomib	Baseline	3-fold	15	0.7323	0.7424
		5-fold		0.7033	0.7272
		10-fold		**0.7376**	**0.76534**
Cisplatin		3-fold	15	0.6048	0.8571
		5-fold		**0.6220**	**0.8549**
		10-fold		0.6070	0.8738
Docetaxel		3-fold	15	0.4929	0.6212
		5-fold		**0.5318**	**0.6734**
		10-fold		0.4416	0.6478
Paclitaxel		**3-fold**	15	**0.5307**	**0.6120**
		5-fold		0.5156	0.5888
		10-fold		0.4966	0.5937
Bortezomib	Fine-tuned	3-fold	15	0.7210	0.7327
		5-fold		0.6966	0.7156
		10-fold		**0.7327**	**0.7514**
Cisplatin		3-fold	15	0.5567	0.8234
		5-fold		**0.6048**	**0.8442**
		10-fold		0.5781	0.8525
Docetaxel		3-fold	15	0.4989	0.6476
		5-fold		**0.5228**	**0.6669**
		10-fold		0.4482	0.6546
Paclitaxel		**3-fold**	15	**0.5508**	**0.6084**
		5-fold		0.5465	0.6205
		10-fold		0.5464	0.6182

Table 4. scJoint experiments for three databases, including 10x Genomics, mouse primary motor cortex data and CITE-seq & ASAP-seq PBMC data

Dataset	DB	Experiment	Stage 1 Accuracy	Stage 3 Accuracy
10x Genomics	10x	Baseline	**0.9125**	**0.9010**
10x Genomics	10x	Fine-tuned	0.9113	0.8992
Mouse Primary Motor Cortex Data	MOp	Baseline	0.8885	0.8976
Mouse Primary Motor Cortex Data	MOp	Fine-tuned	**0.8937**	**0.8982**
CITE-seq and ASAP-seq PBMC data	db4_control	Baseline	0.9225	0.9251
CITE-seq and ASAP-seq PBMC data	db4_control	Fine-tuned	**0.9266**	**0.9275**

Table 5. BERMUDA results for Human Pancreas and PBMC datasets for two autoencoder model: Encoder 2 and Encoder 200

For Human Pancrese Dataset							
Method	***Model***	***Model Setting***	***Epochs***	***Running time***	***Divergence Score***	***Silhoutee score***	
AE	Encoder 20	Baseline	2000	1 hr 5 mins	0.615	0.648	
Uncorrected					8.156	0.372	
AE	Encoder 20	Fine-tuned	2000	1 hr 5 mins	0.096	0.594	
Uncorrected					8.168	0.373	
For PBMC Dataset							
Method	***Model***	***Model Setting***	***Epochs***	***Running time***	***Total loss***	***Reconstrct loss***	***Transfer loss***
AE	Encoder 20	Fine-tuned	2000	58.7 mins	0.6373	0.6240	0.02900
AE	Encoder 20	Fine-tuned 1	2000	57.4 mins	0.8023	0.7748	0.0459
AE	Encoder 2	Fine-tuned	2000	57.9 mins	0.8756	0.8718	0.0083
AE	Encoder 2	Fine-tuned 1	2000	57.5 mins	0.9361	0.9285	0.0133

The most important in model training was similarity score S_{thr}, a threshold to identify similar clusters among different batches. We performed experiments mainly from 0.8 to 0.9 S_{thr} values as authors tried a wide range of S_{thr} from 0.6 to 1.0 and found best results at 0.8 to 0.9, provided in Supplementary material. For Pancreas experiments, Baron and Muraro scRNA-seq Seurat pre-processed datasets are used, while for PBMC experiments, we used 10X PBMC 8k scRNA-seq dataset. We performed experiments with both autoencoders with fine-tuned hyperparameters. We reported the results for both Pancreas and PBMC for autoencoder and uncorrected methods.

DISCUSSION

Transfer learning was widely demonstrated for its ability to generate feature representations for the prognosis, diagnosis and treatment of human diseases. However, it is noteworthy that the gap between published deep learning/classical machine learning and transfer learning methods in the last few years is significant. In 2022, for example, 4834 papers were published on combined deep learning and classical machine learning, while only 53 papers were published on transfer learning, indicating that transfer learning has not received adequate attention from the research community. More research and attention are needed to fully realize the potential of transfer learning in this field. Transfer learning, similar to homology modeling in bioinformatics, is always computationally tractable but may skip our necessities in understanding the underlying mechanisms and molecular processes behind decision-making. It has to be addressed and complemented by other interpretation approaches such as ablation studies, feature importance analysis, low-dimensional data visualization and model-agnostic interpretation methods (e.g. LIME and SHAP).

Challenges of data scarcity in health informatics

Transfer learning's ultimate goal is to transfer the knowledge to a target task with minimal or no amount of training data. With the advent of HTS technologies, a vast amount of sequencing data is available; however, the availability of annotated human data for precision medicine remains a challenge due to various ethical, financial and technical factors. Additionally, in the case of rare diseases, the unavailability of sufficient data makes it difficult to train machine learning models. The data scarcity issue can be addressed by adapting different measures including fine-tuning, using enough data from the same task from other domains (domain adaptation), and careful evaluation of source and target datasets leading to robust transfer learning model performance.

Challenges associated with the use of pre-clinical data

Although transfer learning has shown great potential for high-throughput data in clinical health informatics, it still faces multiple challenges, such as, but not limited to, genetic variations between pre-clinical and clinical data, different experimental settings and environmental variations. In light of those variations, transfer learning models' performance may fall short in terms of accurate predictions. Those issues pose serious concerns to transfer learning model training but we can try various customized approaches to solve these problems, which may include domain adaptation, batch normalization, fine-tuning and data fusion with a wide range of environmental conditions. Additionally, fine-tuning and domain adaptation could be required to address the limited transferability of pre-trained transfer learning models. Recently, researchers presented a multi-task pre-training and fine-tuning approach for limited phenotype data

Table 6. Transfer learning methods summary for source and target domain/tasks and input/output with adapted methods and performance metrics

Method	TL strategy	TL task	Source task and input	Target task and output	Methods and metrics	Evaluation strategy	TL performance
MultiPLIER [24]	Unsupervised	An unsupervised feature representation task	*Task*: Extract meaning patterns across samples *Input*: Large human gene expression data from multiple tissues and conditions	*Task*: transfer and identify learned patterns in target data *Output*: correlated gene expression patterns	*Methods*: PLIER [31], PCA *Metrics*: Benjamini-Hochberg correction (FDR) for LVs, AUC	Hold-out	More relevant processes were found in project human data from learned LV
projectR [15]	Unsupervised	Learn and project the single-cell patterns to identify shared biological factors	*Task:* Learn and explore latent spaces for multiple retina development (mice) *Input*: single-cell and bulk RNA data	*Task:* Project learned latent space for human brain development datasets *Output:* Identify shared biological factors among independent datasets	*Methods*: scCoGAPS, SVD, PCA *Metrics*: AUC	Hold-out	Correctly predicted the cell types for learned patterns using multiple scRNA datasets from published studies
projectR_ICI [16]	Unsupervised	Evaluation of preclinical to clinical pathways associated with anti-CTLA-4 treatment	*Task*: Detect conserved transcriptional signatures associated with anti-CTLA-4 treatment *Input*: scRNA gene expression data	*Task*: Use projectR to find NK cell activation using bulk-RNA, mass cytometry, and scRNA dataset Output:	*Methods*: projectR, CoGASP *Metrics*: ROC, Spearman correlation	Hold-out	Identify conserved and clinical transcriptional changes in response to ICIs
FIT [32]	Unsupervised	Regression (Lasso) and classification (SVM) for gene-level and FIT model prediction, respectively	*Task*: Compute, learn, and predict cross-species relationship *Input*: Mouse gene expression data and reference compendium	*Task*: Use prediction to learn relationships in human data *Output*: Genes with the high effect size for matching conditions	*Methods*: PCA, SVM, Lasso regression (FIT) *Metrics*: Mann–Whitney test	FIT: 10-fold cross-validation SVM: Hold-out	FIT predicted 20-50% more genes as compared with direct mouse-to-human genes exploration
CaSTLe [37]	Supervised	A supervised TL to classify and label previously unknown cell types in the target domain	*Task*: Label cell types *Input*: single-cell gene expression	*Task*: Single cell classification *Output*: Label for each cell in the target data	*Methods*: XGBoost PCA *Metrics*: AUC	Hold-out	Accuracy upto 90.9% for multi-class classification
TF-Binding Matrix [33]	Unsupervised	TF binding prediction by training a two-step model	*Task*: Learn common DNA features (TF binding motifs) *Input*: ChIP-seq peaks	*Task*: Exploit learned common features for TF prediction	*Methods*: CNN [70] *Metrics*: AUCPR	Hold-out	As compared with previous works, the proposed model improves TF binding prediction even when training data are limited
scJoint [14]	Domain Adaptation	Integrating single-cell data in the target domain	*Task*: Data Integrate *Input*: scRNA-seq	*Task*: Correctly classify cell types Output:	*Methods*: KNN, PCA *Metrics*: Accuracy, Silhouette coefficients	Random subsampling	scJoint outperforms the other methods with good performance
BERMUDA [13]	Domain Adaptation	Correct and remove batch effects using autoencoders	*Task*: Identify similar batch clusters and combine vastly different batches *Input*: gene expression data	*Task*: Train an autoencoder with different batches to remove batch effects *Output*: Removed cell types visualization with UMAP	*Methods*: Autoencoder *Metrics*: Silhouette score, Divergence score, Entropy score	Random oversampling	BERMUDA outperformed existing methods for batch correction under different cell populations compositions, simulated data, human pancreas data by preserving batch-specific biological signals

(continued)

Table 6. Continued

Method	TL strategy	TL task	Source task and input	Target task and output	Methods and metrics	Evaluation strategy	TL performance
trVAE [38]	Transductive	Transfer a condition from source domain to the target domain	*Task*: Learn relationship across domains *Input*: Human gene expression data	*Task*: Transfer the learned out of distribution condition to a new domain data *Output*: Transformation to second condition	*Methods*: Variational autoencoder *Metrics*: Pearson's correlation	Hold-out	trVAE shows better performance against the standard benchmarks, including scGen, scVI, MMD-CVAE
XGSEA [39]	Transductive	A regression task for p-values prediction of target gene sets	*Task*: gene set enrichment (gene ontology) *Input*: mouse or zebrafish	*Task*: gene set enrichment (gene ontology) *Output*: gene set associations	*Methods*: Logistic regression Metrics: MSE, MAE, CI, Pearson correlation, cosine similarity, and AUC	Hold-out	Tested on CD8+ T-cell ATACseq dataset, XGSEA predicted enriched pathways in human tumor data from mouse tumor data
PRECISE [40]	Transductive	Drug responses predictors from pre-clinical models to human tumors	*Task*: Find common factors between pre-clinical models and human data using principal vectors (PVs) *Input*: Mouse gene expression data	*Task*: Use human gene expression data to predict the drug responses *Output*: Drug response prediction	Methods: PCA, Ridge regression Metrics: Spearman correlation, Bregman matrix divergence	Nested 10-fold cross-validation	Showed better performance in finding biomarkers and their companion drugs, such as $BRAF^{V600E}$ in skin cancer
AITL [41]	Inductive (Domain Adaptation)	AITL first removed the discrepancies in both source and target domains, then predicted the drug responses in the target task from learned labels in the source task	*Task*: Feature extraction from source and target data *Input*: Gene expression (in vitro human) cell line, patient qualitative outcome data	*Task*: Drug response prediction for TCGA patients *Output*: *in vivo* predictions	Methods: Mutil-task subnetwork with regression (source task) and classification (target task) Metrics: AUROC, AUPR	3-fold cross-validation	Based of AUROC, AITL was tested on multiple state-of-the-art methods and performed best in all experiments
ssNN [36]	Transductive	Predict DEGs and enriched pathways (human *in vivo*)	*Task*: Supervised model on mouse data to find enriched phenotypes *Input*: Labeled mouse phenotype and molecular data	*Task*: Human phenotypes and DEGs prediction *Output*: Predicted pathways	Methods: KNN, SVM, RF, NN Metrics: AUROC, F1-score	10-fold cross-validation	Discovered relevant mouse features for human data
TransComp-R [42]	Unsupervised	Predict drug resistance to IBD disease	*Task*: Find human genes associated with responding phenotype *Input*: Human gene expression, and drug response data	*Task*: Find the homologous genes in the mouse proteomics data *Output*: Mouse proteins associated with drug response	Methods: PCA	Hold-out	Found a collagen-binding integrin, which was resisting to the treatment
scDEAL [43]	Supervised	Drug response prediction in source domain	Task: Feature extraction from bulk gene features and drug response prediction in each bulk cell line Input: Bulk RNA-seq data	Task: Deep transfer learning model training and tranferring the learned model to predict the drug responses to scRNA-seq data Output: Single cell drug responses prediction	Methods: Domain-adaptive Neural Network (DaNN) Metrics: F1-score, AUROC, AP score, Precision, Recall, Adjusted Mutual Information (AMI), and Adjusted Rand Index (ARI)	Hold-out	The authors benchmarked scDEAL with six drug treated scRNA-seq datasets, and it shows good performance for drug response label prediction
CaSee [44]	Supervised	Cancer/normal cell discrimination in scRNA-seq datasets	*Task*: Find shared feature space between bulk RNA-seq and snRNA-seq data for cancer/normal cells *Input*: Human gene expression data	Task: Predict cells from scRNA-seq data Output: Cancer vs normal cells	Methods: Capsule network (encoder) Metrics: Accuracy	Hold-out	Tested on multiple scRNA-seq data for different cancers, and the proposed method achieved 96.69% average accuracy

[69]. As most of the transfer learning studies relied upon pre-clinical data for pre-training, it is also important to note the ethical concerns related to pre-clinical data such as animal testing. It may limit the scope of research, and thus the availability of pre-clinical data. Therefore, cautions are needed to comply with the ethical standards and guidelines in different jurisdiction regions.

Transfer learning trends in clinical research

We observed a trend across different studies, where authors used pre-clinical data (such as mouse, zebrafish and large data compendium) in source domain tasks and clinical data (human) in target domain tasks. The main reason for this trend is the short-lived nature of animals, which enables *in vivo* experiments resulting in enough amount of data. Many transfer learning methods focus on transferring source domain knowledge to human data in the target domain, such as FIT [32], projectR and projectR_ICI [15, 16]. Furthermore, some methods use human control data with mouse models to identify similarities in the source domain and then train the model on these features to predict responses in the target domain (human data). As transfer learning continues to evolve, it holds the promise of addressing the data inequality challenge in precision medicine and clinical research, providing new insights into the diagnosis and treatment of human diseases.

Despite the breakthrough results that deep transfer learning has provided, transfer learning methods for gene expression analysis have not been explored at their full potential. There are relatively few studies compared with machine learning and deep learning methods, as shown in Figure 1. Nevertheless, transfer learning offers a promising avenue for the future of biomedical research and precision medicine, particularly in addressing the ethical and data scarcity issues associated with *in vivo* experiments for human. Similar to other deep learning model training, transfer learning models may also suffer from overfitting and generalization problems. One of the possible solutions in the case of high-throughput data is to integrate a large compendium of dataset for model training as in MultiPLIER [24]. It is also important to note that, because of ethical concerns, pre-clinical data may not always be readily available for the pre-training of transfer learning models.

FUTURE DIRECTIONS

Transfer learning can be a valuable tool in translational medicine in case of limited *in vivo* human data for clinical decision-making. By leveraging knowledge gained from related diseases or conditions, pre-clinical models and *in vitro* experiments, researchers can make informed predictions about human health outcomes, even when there are limited human data available. One promising area of future research is the development of more advanced transfer learning algorithms that can handle complex and diverse data types, such as multimodal data. With the rise of big data in healthcare, there is a growing need for transfer learning methods that can effectively integrate and learn from various data sources, such as genomics, imaging, electronic health records and patient-reported outcomes. Integration of multimodal data is a critical task to transfer knowledge across modalities and it provides a comprehensive understanding of the underlying mechanism of disease development.

Moreover, transfer learning could also be used to improve the interpretability and transparency of machine learning models in clinical research and health informatics. By leveraging knowledge from related domains or pre-trained models, transfer learning can help extract more meaningful and explainable features from the data, which could enhance the trust and acceptance of these models by clinicians and patients. Furthermore, the reviewed studies may provide a starting point for the future development of transfer learning methods in biomedical research and health informatics.

Key Points

- In bioinformatics, particularly with *in vivo* (human) experiments, obtaining sufficient data can be challenging due to ethical, financial and other factors.
- Transfer learning can enhance precision medicine by utilizing pre-clinical data (such as mouse or zebrafish data) in clinical decision-making.
- This comprehensive review sheds new light on the integration of transfer learning into clinical practices and precision medicine.
- This review also covered benchmarking of transfer learning methods and analysis of selected datasets.

FUNDING

This research was substantially sponsored by the research projects (Grant No. 32170654 and Grant No. 32000464) supported by the National Natural Science Foundation of China and was substantially supported by the Shenzhen Research Institute, City University of Hong Kong. The work described in this paper was substantially supported by the grant from the Research Grants Council of the Hong Kong Special Administrative Region [CityU 11203723]. This project was substantially funded by the Strategic Interdisciplinary Research Grant of City University of Hong Kong (Project No. 2021SIRG036). The work described in this paper was partially supported by the grant from City University of Hong Kong (CityU 9667265).

DATA AVAILABILITY

All datasets used for benchmarking and in transfer learning studies are publicly available. We have provided the data availability links in Table 2.

REFERENCES

1. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell rna-seq in the past decade. *Nat Protoc* 2018; **13**(4): 599–604.
2. D'Adamo GL, Widdop JT, Giles EM. The future is now? Clinical and translational aspects of omics technologies. *Immunol Cell Biol* 2021; **99**(2): 168–76.
3. Li R, Li L, Yungang X, Yang J. Machine learning meets omics: applications and perspectives. *Brief Bioinform* 2022; **23**(1): bbab460.
4. de Anda-Jáuregui G, Hernández-Lemus E. Computational oncology in the multi-omics era: state of the art. *Front Oncol* 2020; **10**:423.
5. Marx V. The big challenges of big data. *Nature* 2013; **498**(7453): 255–60.

6. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol* 2016; **12**(7): 878.
7. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform* 2017; **18**(5): 851–69.
8. Gao Y, Cui Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. *Nat Commun* 2020; **11**(1): 1–8.
9. Toseef M, Li X, Wong K-C. Reducing healthcare disparities using multiple multiethnic data distributions with fine-tuning of transfer learning. *Brief Bioinform* 2022; **23**(3): bbac078.
10. Kapp MB. Ethical and legal issues in research involving human subjects: do you want a piece of me? *J Clin Pathol* 2006; **59**(4): 335–9.
11. Honkala A, Malhotra SV, Kummar S, Junttila MR. Harnessing the predictive power of preclinical models for oncology drug development. *Nat Rev Drug Discov* 2022; **21**(2): 99–114.
12. Steger-Hartmann T, Raschke M. Translating in vitro to in vivo and animal to human. *Curr Opin Toxicol* 2020; **23**:6–10.
13. Wang T, Johnson TS, Shao W, *et al*. Bermuda: a novel deep transfer learning method for single-cell rna sequencing batch correction reveals hidden high-resolution cellular subtypes. *Genome Biol* 2019; **20**(1): 1–15.
14. Lin Y, Tung-Yu W, Wan S, *et al*. Scjoint integrates atlas-scale single-cell rna-seq and atac-seq data with transfer learning. *Nat Biotechnol* 2022; **40**(5): 703–10.
15. Stein-O'Brien GL, Clark BS, Sherman T, *et al*. Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. *Cell Syst* 2019; **8**(5): 395–411.
16. Davis-Marcisak EF, Fitzgerald AA, Kessler MD, *et al*. Transfer learning between preclinical models and human tumors identifies a conserved nk cell activation signature in anti-ctla-4 responsive tumors. *Genome Med* 2021; **13**(1): 1–22.
17. Thrun S, Pratt L. *Learning to learn*. Springer Science & Business Media, 2012.
18. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2009; **22**(10): 1345–59.
19. Scott T, Ridgeway K, Mozer MC. Adapted deep embeddings: a synthesis of methods for k-shot inductive transfer learning. *Adv Neural Inf Process Syst* 2018; **31**.
20. Ravì D, Wong C, Deligianni F, *et al*. Deep learning for health informatics. *IEEE J Biomed Health Inform* 2016; **21**(1): 4–21.
21. Butte AJ. Translational bioinformatics: coming of age. *J Am Med Inform Assoc* 2008; **15**(6): 709–14.
22. Translational bioinformatics. https://www.sciencedirect.com/topics/medicine-and-dentistry/translational-bioinformatics.
23. Ebbehoj A, Thunbo MØ, Andersen OE, *et al*. Transfer learning for non-image data in clinical research: a scoping review. *PLOS digital. Health* 2022; **1**(2): e0000014.
24. Taroni JN, Grayson PC, Qiwen H, *et al*. Multiplier: a transfer learning framework for transcriptomics reveals systemic features of rare disease. *Cell Syst* 2019; **8**(5): 380–94.
25. Aslan MF, Unlersen MF, Sabanci K, Durdu A. Cnn-based transfer learning–bilstm network: a novel approach for covid-19 infection detection. *Appl Soft Comput* 2021; **98**:106912.
26. Gautam Y. Transfer learning for covid-19 cases and deaths forecast using lstm network. *ISA Trans* 2022; **124**:41–56.
27. Arora V, Ng EY-K, Leekha RS, *et al*. Transfer learning-based approach for detecting covid-19 ailment in lung ct scan. *Comput Biol Med* 2021; **135**:104575.
28. Ahuja S, Panigrahi BK, Dey N, *et al*. Deep transfer learning-based automated detection of covid-19 from lung ct scan slices. *Applied Intelligence* 2021; **51**(1): 571–85.
29. Maqsood M, Nazir F, Khan U, *et al*. Transfer learning assisted classification and detection of alzheimer's disease stages using 3d mri scans. *Sensors* 2019; **19**(11): 2645.
30. Collado-Torres L, Nellore A, Kammers K, *et al*. Reproducible rna-seq analysis using recount2. *Nat Biotechnol* 2017; **35**(4): 319–21.
31. Mao W, Zaslavsky E, Hartmann BM, *et al*. Pathway-level information extractor (plier) for gene expression data. *Nat Methods* 2019; **16**(7): 607–10.
32. Normand R, Wenfei D, Briller M, *et al*. Found in translation: a machine learning model for mouse-to-human inference. *Nat Methods* 2018; **15**(12): 1067–73.
33. Novakovsky G, Saraswat M, Fornes O, *et al*. Biologically relevant transfer learning improves transcription factor binding prediction. *Genome Biol* 2021; **22**(1): 1–25.
34. Chèneby J, Gheorghe M, Artufel M, *et al*. Remap 2018: an updated atlas of regulatory regions from an integrative analysis of dna-binding chip-seq experiments. *Nucleic Acids Res* 2018; **46**(D1): D267–75.
35. Le Dily F, Bau D, Pohl A, *et al*. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* 2014; **28**(19): 2151–62.
36. Brubaker DK, Proctor EA, Haigis KM, Lauffenburger DA. Computational translation of genomic responses from experimental model systems to humans. *PLoS Comput Biol* 2019; **15**(1): e1006286.
37. Lieberman Y, Rokach L, Shay T. Castle - classification of single cells by transfer learning: harnessing the power of publicly available single cell rna sequencing experiments to annotate new experiments. *PloS One* 2018; **13**(10): e0205499.
38. Lotfollahi M, Naghipourfar M, Theis FJ, *et al*. Conditional out-of-distribution generation for unpaired data using transfer vae. *Bioinformatics* 2020; **36**(Supplement_2): i610–7.
39. Cai M, Nguyen CH, Mamitsuka H, Li L. Xgsea: cross-species gene set enrichment analysis via domain adaptation. *Brief Bioinform* 2021; **22**(5): bbaa406.
40. Mourragui S, Loog M, Van De Wiel MA, *et al*. Precise: a domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *Bioinformatics* 2019; **35**(14): i510–9.
41. Sharifi-Noghabi H, Peng S, Zolotareva O, *et al*. Aitl: adversarial inductive transfer learning with input and output space adaptation for pharmacogenomics. *Bioinformatics* 2020; **36**(Supplement_1): i380–8.
42. Brubaker DK, Kumar MP, Chiswick EL, *et al*. An interspecies translation model implicates integrin signaling in infliximab-resistant inflammatory bowel disease. *Sci Signal* 2020; **13**(643): eaay3258.
43. Chen J, Wang X, Ma A, *et al*. Deep transfer learning of cancer drug responses by integrating bulk and single-cell rna-seq data. *Nat Commun* 2022; **13**(1): 1–13.
44. Yuan S, Zhang X, Guo C, *et al*. Casee: a lightning transfer-learning model directly used to discriminate cancer/normal cells from scrna-seq. *bioRxiv* 2022.
45. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–32, 2017.
46. Lee DD, Pham P, Largman Y, Ng A. Advances in neural information processing systems 22 Technical report, Tech. Rep.*Tech Rep* 2009.

47. Louizos C, Swersky K, Li Y, *et al*. The variational fair autoencoder. *arXiv preprint arXiv:151100830* 2015.
48. Amodio M, Van Dijk D, Srinivasan K, *et al*. Exploring single-cell data with deep multitasking neural networks. *Nat Methods* 2019; **16**(11): 1139–45.
49. Lopez R, Regier J, Cole MB, *et al*. Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018; **15**(12): 1053–8.
50. Mohammad Lotfollahi F, Wolf A, Theis FJ. Scgen predicts single-cell perturbation responses. *Nat Methods* 2019; **16**(8): 715–21.
51. Clark BS, Stein-O'Brien GL, Shiau F, *et al*. Single-cell rna-seq analysis of retinal development identifies nfi factors as regulating mitotic exit and late-born cell specification. *Neuron* 2019; **102**(6): 1111–26.
52. Sade-Feldman M, Yizhak K, Bjorgaard SL, *et al*. Defining t cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* 2018; **175**(4): 998–1013.
53. Gao J, Aksoy BA, Dogrusoz U, *et al*. Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal. *Sci Signal* 2013; **6**(269): pl1–1.
54. Xin H, Wang Q, Tang M, *et al*. Tumorfusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res* 2018; **46**(D1): D1144–9.
55. Gao H, Korn JM, Ferretti S, *et al*. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* 2015; **21**(11): 1318–25.
56. Ding Z, Songpeng Z, Jin G. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics* 2016; **32**(19): 2891–5.
57. Lyons J, Brubaker DK, Ghazi PC, *et al*. Integrated in vivo multiomics analysis identifies p21-activated kinase signaling as a driver of colitis. *Sci Signal* 2018; **11**(519): eaan3580.
58. Tran HTN, Ang KS, Chevrier M, *et al*. A benchmark of batch-effect correction methods for single-cell rna sequencing data. *Genome Biol* 2020; **21**(1): 1–32.
59. Butler A, Hoffman P, Smibert P, *et al*. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018; **36**(5): 411–20.
60. Stuart T, Butler A, Hoffman P, *et al*. Comprehensive integration of single-cell data. *Cell* 2019; **177**(7): 1888–902.
61. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* 2018; **36**(5): 421–7.
62. Polański K, Young MD, Miao Z, *et al*. Bbknn: fast batch alignment of single cell transcriptomes. *Bioinformatics* 2020; **36**(3): 964–5.
63. Ren X, Zheng L, Zhang Z. Sscc: a novel computational framework for rapid and accurate clustering large-scale single cell rna-seq data. *Genomics Proteomics Bioinformatics* 2019; **17**(2): 201–10.
64. Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. *Adv Neural Inf Process Syst* 2017; **30**.
65. Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–76, 2017.
66. Ding MQ, Chen L, Cooper GF, *et al*. Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeuticsassigning cancers to effective drugs with big data. *Mol Cancer Res* 2018; **16**(2): 269–78.
67. Paltun BG, Mamitsuka H, Kaski S. Improving drug response prediction by integrating multiple data sources: matrix factorization, kernel and network-based approaches. *Brief Bioinform* 2021; **22**(1): 346–59.
68. Rampášek L, Hidru D, Smirnov P, *et al*. Dr. vae: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics* 2019; **35**(19): 3743–51.
69. Si Y, Bernstam EV, Roberts K. Generalized and transferable patient language representation for phenotyping with limited data. *J Biomed Inform* 2021; **116**:103726.
70. Kelley DR, Snoek J, Rinn JL. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res* 2016; **26**(7): 990–9.