


Review

# Transcriptome-Powered Pluripotent Stem Cell Differentiation for Regenerative Medicine

Derek A. Ogi<sup>1</sup> and Sha Jin<sup>1,2,\*</sup> 

<sup>1</sup> Department of Biomedical Engineering, Thomas J. Watson College of Engineering and Applied Sciences, State University of New York at Binghamton, Binghamton, NY 13902, USA; dogi1@binghamton.edu

<sup>2</sup> Center of Biomanufacturing for Regenerative Medicine, State University of New York at Binghamton, Binghamton, NY 13902, USA

\* Correspondence: sjin@binghamton.edu; Tel.: +1-607-777-5718; Fax: +1-607-777-5780

**Abstract:** Pluripotent stem cells are endless sources for in vitro engineering human tissues for regenerative medicine. Extensive studies have demonstrated that transcription factors are the key to stem cell lineage commitment and differentiation efficacy. As the transcription factor profile varies depending on the cell type, global transcriptome analysis through RNA sequencing (RNAseq) has been a powerful tool for measuring and characterizing the success of stem cell differentiation. RNAseq has been utilized to comprehend how gene expression changes as cells differentiate and provide a guide to inducing cellular differentiation based on promoting the expression of specific genes. It has also been utilized to determine the specific cell type. This review highlights RNAseq techniques, tools for RNAseq data interpretation, RNAseq data analytic methods and their utilities, and transcriptomics-enabled human stem cell differentiation. In addition, the review outlines the potential benefits of the transcriptomics-aided discovery of intrinsic factors influencing stem cell lineage commitment, transcriptomics applied to disease physiology studies using patients' induced pluripotent stem cell (iPSC)-derived cells for regenerative medicine, and the future outlook on the technology and its implementation.

**Keywords:** transcriptomics; human stem cell differentiation; RNA sequencing and data analysis; differential gene expression



**Citation:** Ogi, D.A.; Jin, S.

Transcriptome-Powered Pluripotent Stem Cell Differentiation for Regenerative Medicine. *Cells* **2023**, *12*, 1442. <https://doi.org/10.3390/cells12101442>

Academic Editors: Albert Rizvanov and Ayşegül Doğan

Received: 1 April 2023

Revised: 15 May 2023

Accepted: 18 May 2023

Published: 22 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Stem cells are able to differentiate into different cell types depending on their potency. The two most utilized sources of human pluripotent stem cells (hPSCs) are human embryonic stem cells (hESCs) and induced pluripotent stem cells (iPSCs). They are utilized in organ regeneration, cell and tissue transplantation, drug testing, tissue repair, wound healing, and diverse disease treatments after induced lineage specification. Over the past two decades, tremendous efforts have been made to generate favorable tissues or cell types through in vitro stem cell differentiation with the proper extracellular conditions. Cell–cell and cell–microenvironment interactions have a large impact on stem cell differentiation [1]. Extracellular matrix (ECM) compositions vary depending on the tissue type, which may create better microenvironmental conditions to guide hPSC differentiation into specific cell types and organoids [1–4]. In addition, the regulation of cellular transcriptome expression plays an important role in stem cell lineage commitment [5,6]. Studies in this regard have focused on understanding the complex interplay between gene expression and cell fate as cells specify distinct lineages. Microfabricated substrates provided optimal topographical placement for cell–cell interactions, promoting cell signaling for the lineage commitment of stem cells [7–9]. Growth factor nanoparticle delivery has been implicated in osteogenic differentiation, and laminin expression promotes differentiation into many cell types, including cardiomyocytes, keratinocytes, and retinal cells [9,10]. However, the generation of biologically mature cells/tissues from hPSCs has been challenging due to the complexity

of stepwise developmental trajectories during differentiation and the multiple extrinsic and intrinsic regulatory components. In the process of hPSC differentiation, transcription factors are the most critical intrinsic regulators in cell fate decisions [11]. For example, *FLK2* and *THY1.1* promote the lineage specification of hematopoietic stem cells (HSCs) toward either short-term HSCs, long-term HSCs, or a multipotent progenitor population, depending on the expression levels of the two genes [12]. Short- and long-term HSCs are retained in the bone marrow and are characterized by the duration that they will remain as HSCs. The characterized multipotent progenitor populations are related to either myeloid progenitor cells or the more specified myeloid-committed oligopotent progenitor cells [12]. *THY1.1*<sup>lo</sup>− and *FLK2*− cells committed to a long-term HSC cell fate, *THY1.1*<sup>lo</sup>+ and *FLK2*+ cells differentiated into short-term HSCs, and the multipotent progenitor population was characterized by *THY1.1*− and *FLK2*+ expression [12].

One approach to analyzing global transcriptome expression, including transcription factors, is through RNA sequencing (RNAseq) and subsequent data analysis. RNAseq technologies generate the transcriptomic data of individual cells and allow the determination of how gene expression varies through the comparison of differential gene expression (DGE) among biological samples [13,14]. They generate high-quality genomic information for single cells or groups of cells. Hence, they have been widely applied to characterize the heterogeneity of hPSC-derived cells and cellular specification. Determining marker genes for each cell type allows for the more accurate cell typing of individual cells. Promoting the expression of these genes could also lead to increased differentiation toward that specific fate. Furthermore, DGE analysis offers an indicative assessment of the status of hPSC differentiation toward a specific cell or tissue type and the maturity of the hPSC-derived cells and organoids. Due to high heterogeneity within tissues, it is important to screen differentiated cells individually. Hence, single-cell RNA sequencing (scRNAseq) has been utilized for high-throughput RNAseq data analysis to assess each individual cell's biological identity [15]. This review focuses on RNAseq techniques, tools for RNAseq data interpretation, RNAseq data analytic methods and their utilities, and transcriptomics-enabled human stem cell differentiation. In addition, the review outlines the potential benefits of the transcriptomics-aided discovery of intrinsic factors influencing stem cell lineage commitment, transcriptomics for disease physiology studies using patients' iPSC-derived cells, and the future outlook on the technology and its implementation.

## 2. RNAseq Techniques

RNAseq provides the transcriptomic information that gives a cell its own identity. scRNAseq allows for gene expression analysis and interpretation at as low as the cellular level [16]. It is performed by first extracting the RNA, reverse transcribing RNA to a cDNA library, scaling up the generated cDNA library, and then sequencing the library [13,17–19]. Although the reverse transcription of RNA into cDNA might introduce some errors that limit the accuracy of sequence quantification [20], the accuracy of RNAseq of bulk cells is comparable to that of traditional quantitative reverse transcription PCR (qRT-PCR) in both the gene expression level and transcriptome complexity [21]. Furthermore, single-nucleus RNA sequencing (sNuc-Seq) that uses isolated nuclei is available with comparable transcriptome analysis sensitivity and cell type classification ability [22]. Several different single-cell and sNuc-Seq methods are utilized for sequencing. Table 1 highlights the methods of different RNAseq techniques, their advantages, and their limitations. The RNAseq methods can be either low throughput, such as Smart-seq2 and CEL-Seq2, or high throughput, such as Drop-seq, Seq-Well, or sci-RNA-seq [22]. RNAseq has also been accomplished by focusing specifically on the 5'-end with different methods, such as CAGE, RAMPAGE, STRT, NanoCAGE XL, and Oligo-capping [23]. All methods provide useful data but vary from one another in data consistency, cost, sensitivity, and duration [22]. sNuc-Seq is utilized for the sequencing of cells that are not easily separated into single cells, such as those in the brain, muscle, and adipose tissue [22]. There are associated limitations with sNuc-Seq since it does not consider transcriptomic data outside of the

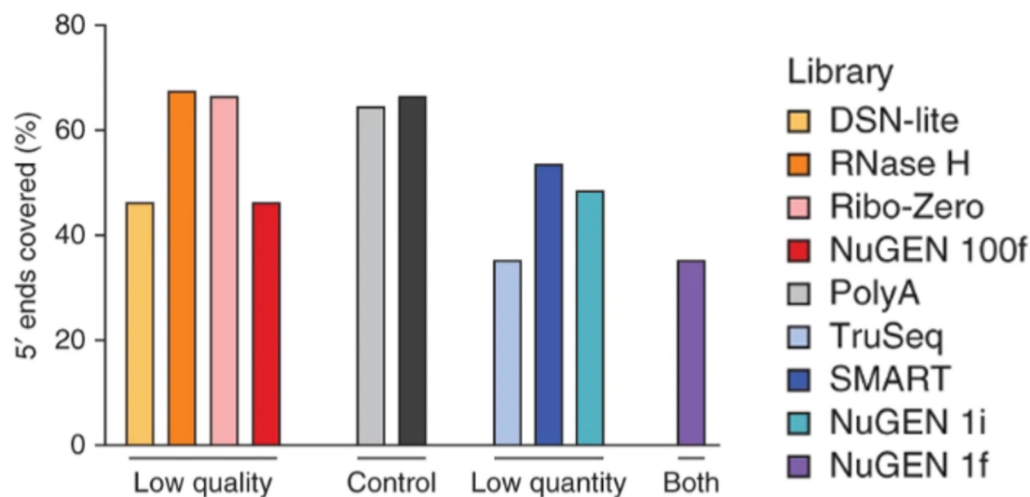
nucleus like scRNAseq or bulk RNAseq [22]. Bulk RNAseq differs from scRNAseq in that it measures the average gene expression of many cells rather than an individual cell's genome expression. Thus, bulk RNAseq provides less data variability between samples compared to scRNAseq or sNuc-Seq sequencing [22] and is unable to distinguish unique cell populations in biological samples [24]. This limitation results from the inability to separate cell types in heterogeneous tissues so that the average genome expression from bulk sequencing cannot determine each cell type's unique genome expression.

**Table 1.** RNAseq techniques applied in current research.

RNAseq Technique	Description	Advantages	Limitations	Methods	References
Bulk RNAseq	Measures genome expression of bulk sample containing many cells and averages expression	Lower data variability between samples, cost-effective	Cannot distinguish unique cell populations	SMARTer SuperScript II reverse transcriptase from Invitrogen Prime-seq method	[21,22,24,25]
scRNAseq	Measures genome expression of individual cells	Distinguishes unique cell populations	Cannot accurately distinguish gene expression from multinucleated cells	Tube-based sequencing methods: SMARTer Ultra Low RNA Kit, TransPlex kit Microfluidic-based sequencing methods: SMARTer cDNA synthesis utilizing the C1 microfluidic system	[21,22]
sNuc-Seq	Measures genome expression of individual nuclei	Best method for analyzing multinucleated cells	Does not consider genomic information from outside of the nucleus	Tube-based sequencing methods: SMARTer Ultra Low RNA Kit, TransPlex kit Microfluidic-based sequencing methods: SMARTer cDNA synthesis utilizing the C1 microfluidic system	[21,22]
Temporal RNAseq	scRNAseq method for analyzing genome expression of individual cells at different time points	More complete data acquisition compared to scRNAseq, tracks gene expression of differentiating and maturing cells	Very expensive; variable differentiation rate can lead to transcriptome inconsistencies between cells by assuming they are at the same differentiation step	Next-maSigPro Modeling the Poisson distribution's rate parameter	[26,27]
Spatial RNAseq	scRNAseq method for analyzing genome expression of individual cells within a measured cluster to find location of specific gene expression	Provides cell transcriptomic information at the spatial level within histological sections, measures heterogeneity of genomic information and tissue architecture at a cellular level	Very limited current applications in research, no applications in stem cell differentiation	DBiT-seq for spatial omics sequencing using formaldehyde-fixed tissues Spatial sequencing performed with 2100 Bioanalyzer from Agilent with Qubit dsDNAHS Assay Kit from Life Technologies and sequenced with Illumina NextSeq sequencer	[28–31]
Spatiotemporal RNAseq	scRNAseq method for analyzing genome expression that combines spatial and temporal RNAseq	Combines advantages of spatial and temporal sequencing to obtain genomic information and tissue architecture of a histological section over time	Very expensive, very limited current applications in research, no applications in stem cell differentiation	10x RNAseq Visium spatial transcriptomic platform utilizing PHATE trajectory analysis	[28,29,31–33]

Interestingly, certain RNAseq methods, such as Ribo-Zero, RNase H, duplex-specific nuclease, NuGEN, and SMART, have been developed for low-input and degraded samples with varying success [34]. The RNase H sequencing method was the most successful out of the observed methods in obtaining high values for the number of genes detected, the percentages of 5' and 3' covered, and Pearson correlation coefficients, with the lowest levels of rRNA and coefficient of variation [34]. The comparison of 5'-to-3' sequence coverage for the observed low-input methods is shown in Figure 1. The Ribo-Zero sequencing method was also effective in transcribing moderately degraded samples with comparable 5'-to-3' coverage [34,35]. However, Ribo-Zero loses its effectiveness if the RNA input is extremely degraded. In contrast, another method known as RNA Access RNAseq remains effective even for severely degraded samples as long as the input sample is larger than 5 nanograms [35]. However, many sequencing methods are inadequate to use for these samples due to poor observed performance and technical drawbacks [34]. These inadequate RNAseq methods utilize oligo(dT) to isolate RNA, particularly poly(A)<sup>+</sup> RNA [34]. These methods could not generate significant results given low-quality or low-quantity readings due to a failure to observe poly(A)– or ribosomal RNA transcripts. Nevertheless, oligo(dT)

is utilized as a primer for the reverse transcription of RNA into cDNA for most high-throughput scRNAseq methods, including 5' capture, Drop-seq, Seq-well, sci-RNA-seq, 10x, and inDrop [36,37]. It is also utilized for a few low-throughput scRNAseq methods, such as Smart-seq and CEL-seq, and bulk sequencing methods, such as TagSeq, QuantSeq, and 3' Pool-seq [37]. While oligo(dT)-related RNAseq methods are not optimal for degraded sample sequencing, methods such as RNase H, Ribo-Zero, NuGEN, and SMART can more effectively transcribe limited transcriptomic information [34].



**Figure 1.** The 5'-to-3' sequence coverage of low-quantity and low-quality samples utilizing different RNAseq analysis techniques. Reproduced with permission from Adiconis, Nature Methods; published by Springer Nature, 2013 [34].

Additionally, paired-end sequencing was suggested to provide substantial benefits to the sequencing yield compared to single-end sequencing [38]. Single-end sequencing involves only sequencing a single side of the cDNA generated during RNAseq, whereas paired-end sequencing involves sequencing both ends of the cDNA [38]. Inclusion and exclusion isoforms can be better detected in paired-end sequencing compared to single-end sequencing due to the increased ability to detect alternative exons [38]. Paired-end sequencing is also more cost-effective compared to single-end sequencing, especially when using short paired-end reads compared to long single-end reads [39]. Further, paired-end reads provide better gene expression results for each isoform and more characterized DGE results compared to single-end reads [39].

### 3. Tools for RNAseq Data Interpretation

There are several data analysis tools for interpreting RNAseq results applied to hPSC differentiation into diverse cell types. Each software tool interprets the data in a slightly different way, but each generates meaningful data. These different analysis packages and their uses, alongside the software that runs each program, are displayed in Table 2. These data include information on the DGE of the cells and can also include more specified information on cell clustering, cell subpopulation identification, network reconstruction, pseudotime analysis, and how transcriptomic expression changes given varying environmental or cellular conditions [40]. Pseudotime analysis is the process of developing a continuous trajectory for the genome expression of the cell over a specified time course [41]. Saturation, or the determination that sufficient sequencing has occurred for a specific sample, can be analyzed through software tools such as *vegan* [42]. Three analysis methods utilized for DGE are *limma*, *DESeq2*, and *EdgeR* [43]. They have differences in the approach to processing the data, for which *limma* utilizes a linear model, while *DESeq2* and *EdgeR* use a negative binomial distribution [43]. The normalization of RNA data is not necessary for *DESeq2*, while the other two methods require it. *DESeq2* also utilizes local regression to create bounds for the mean and variance of the data, as well as utilizes the Bayes theorem

to generate gene movement and apply thresholds for gene expression based on available information [44]. EdgeR and the similar baySeq analysis method use the Bayesian empirical method to moderate transcript overdispersion [44]. The limma package is recommended to be used alongside other analysis methods, such as the EdgeR-associated package voom, which is used to normalize the data [44]. Myrna is another software analysis tool that determines the DGE from RNAseq data and is particularly useful for larger datasets [45]. Myrna utilizes parametric and non-parametric permutation testing to determine DGE after gene or exon coverage is calculated post-alignment [45]. Not all sequencing analysis methods are focused on DGE. MAP-RSeq is software focused on the generation of transcriptome alignment, gene and exon counts, fusion transcript information, and single-nucleotide variant information [46]. This software is intended to generate information on transcript quantification and should be paired with other downstream software such as EdgeR to determine DGE [46]. Variations in the datasets make certain analysis methods significantly more accurate depending on the environmental conditions of the experiment. EBSeq, which is similar in function to the baySeq analysis method, is beneficial for identifying differentially expressed isoforms, whereas EdgeR is better for determining biological variation [44]. NOIseq is a method that determines the model's noise but is limited in that replication is not possible. Other analysis software such as sleuth is more beneficial for filtering and observing transcripts that occur in low abundance [44]. Based on the information needed for a specific dataset, different software packages can be used to obtain the relevant information. Software packages can also be utilized alongside one another, similar to how MAP-RSeq can be utilized with EdgeR to obtain more information from the transcriptome data. Analysis software may vary in effectiveness based on the cell type. For example, Rajkumar et al. showed that EdgeR was superior compared to Cuffdiff2, DESeq2, and the two-stage Poisson model software analysis methods in determining DGE in mouse brain samples [47]. Software analysis tools can compensate for the limitations of current RNAseq methods [13]. The analysis software allows for normalizing the transcriptomic data, correcting errors associated with factors that are not biological, and smoothing the data by predicting missing values [41,48]. Some data analysis tools will also visualize the data at the cellular level through either cluster or trajectory analysis charts or at the genetic level through differential expression, gene set, or gene regulatory network analysis [48]. Regardless of the analysis method used, it is expected that the transcriptomic analysis of the RNAseq data can provide information that helps to uniquely identify different cell types.

**Table 2.** RNAseq analysis software and associated software packages.

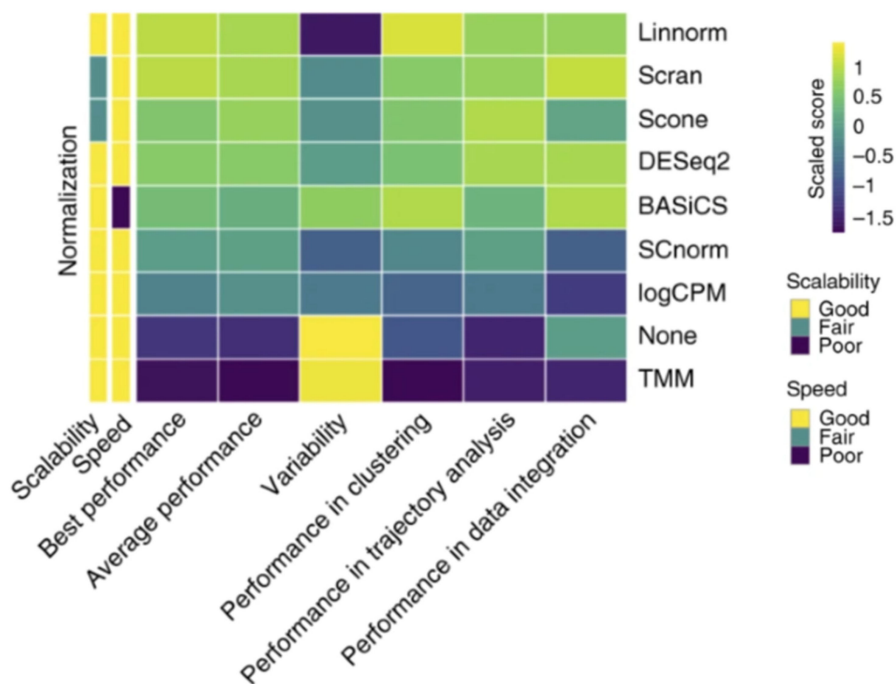
Software	Analysis Packages	Description and Common Uses	References
R	vegan	Determines saturation of a sample	[42]
R	EdgeR v3.24, limma, DESeq, DESeq2 v.1.0.19, baySeq, voom, Myrna, EBSeq, two-stage Poisson model, MAST v.1.0.5, monocle, SCDE	DGE analysis of datasets	[43–45,47,49,50]
Python v.2.7.6	D <sup>3</sup> E	DGE analysis of datasets	[49,50]
C++	Cuffdiff2 v.2.1.1	DGE analysis of datasets	[47]
R, Python v.2.7, Java v.1.6.0_17, PERL v.5.10.0	Map-RSeq	Transcriptome alignment, gene/exon count, fusion transcripts, single-nucleotide variant information	[46]
R	NOIseq	Determines noise within a dataset	[44]
R	sleuth	Observes low-abundance transcripts	[44]
R	DESeq2 v.1.20.0, count-per-million (EdgeR v.3.24.2), Linnorm v.2.6.0, SCnorm v.1.4.2, BASiCs v.1.4.0, scran v.1.8.2, TMM (EdgeR v.3.24.2)	Normalization of a dataset	[51]
R	Seurat v.2.3.4, Seurat3.0, RaceID3 v.0.1.3, RCA v.1.0, SC3 v.1.10.0, clusterExperiment v.2.2.0, MNN (scran v.1.8.2), CCA, scanorama v.1.0	Clustering analysis of a dataset	[51,52]
Python	BERMUDA, scVI	Clustering analysis of a dataset	[52]
R	DPT, Monocle2, Slingshot, SLICER, TSCAN	Trajectory analysis of a dataset	[51]
Python	NS-Forest v2.0, DESC	Machine learning of DEGs	[52,53]

#### 4. RNAseq Data Analytic Methods and Their Utilities

There are different methods to observe DGE and perform normalization depending on the RNAseq analysis method used. Bulk RNAseq normalization methods include TMM, DESeq2, and count-per-million, and bulk RNAseq differential expression methods include DESeq and EdgeR [49,51]. Linnorm, SCnorm, BASiCS, and scran are scRNAseq normalization methods, and MAST, Monocle, SCDE, and D<sup>3</sup>E are differential expression analysis methods mainly used for scRNAseq [50,51]. Linnorm and scran are the normalization methods that most consistently produced beneficial normalization results [51]. A comparison between normalization methods is shown in Figure 2. The differential expression analysis of scRNAseq data can utilize bulk analysis methods in some scenarios but requires a new characterization of the difference in expression to be established beyond having the average expression be a non-zero difference [49]. While normalization and differential expression are more effective when utilizing different methods for bulk and scRNAseq, clustering analysis software can be used on datasets from both sequencing methods. Clustering analysis methods include Seurat, RaceID3, RCA, SC3, and clusterExperiment [51]. Trajectory analysis methods such as DPT, Monocle2, Slingshot, SLICER, and TSCAN can be utilized for both bulk and scRNAseq as well but require distinct pseudotime paths from one cell type to another to generate each cell's trajectory and overall lineage commitment [51]. While there are a few popular RNAseq analysis methods, such as EdgeR and DESeq2, for DGE, there is not a universal method. This is due to assumptions that must be made when using each analysis tool. Both EdgeR and DESeq2 assume that differentially expressed genes (DEGs) are rare and that the variance and signal intensity are independent of one another [54]. The transformation of the data must also occur in both methods when combining normalization and differential expression testing to minimize inconsistencies in results [54]. These assumptions may introduce errors in the data analysis. Therefore, each analysis method requires testing to find the optimal experimental conditions to minimize the error observed due to the assumptions made. The removal of outlier data points is one major source of error for some analysis methods, such as DESeq and DESeq2, due to the data manipulation of these methods [55]. DESeq utilizes a dispersion-mean trend to remove outlier data points, whereas DESeq2 uses Cook's distance metric. In both instances, the removal of data points that fall outside of the distribution range occurs, which limits the discovery of new differentially expressed genes but still prevents false discovery [55]. Down-weighting the outlier data points rather than removing them from the dataset has the potential to reduce false discovery rates while still allowing for new DEGs to be discovered. Reducing false discovery errors not only improves the accuracy of the data but also allows for better comparison of the data between samples [55]. User error is another limitation of RNAseq analysis software. While DESeq2 and EdgeR are commonly used due to the strong results obtained, as well as the relative ease of use, novice users of the software packages could miss steps, such as checking the hypothesis of the methods or controlling data quality [54]. Tools such as SARTools have been created to streamline the use of DESeq2 or EdgeR to prevent user error [54]. This tool is used to ensure that all steps necessary for the proper use of the software are followed to prevent errors. Reducing error is beneficial to the software's accuracy in producing normalization, cluster and trajectory analyses, and DGE results from RNAseq datasets.

RNAseq data analysis facilitates the identification of cell identity and the discovery of new cell subtypes by observing the DEGs and comparing them to previous results [56]. For instance, Clark et al. utilized the Smart-Seq2 data analysis method and 10x Genomics Chromium 3' v2 platform software to characterize retinal cell types through the DGE analysis of scRNAseq transcriptomic data [57]. They observed gene expression changes throughout retinal neurogenesis while specifying different retinal cell types throughout development [57]. *CCND1*, *CDK4*, and *PAX6* were the gene markers used to specify retinal progenitor cells, while further specified cells, such as retinal ganglion cells and amacrine cells, were identified with *POU4F2/ISL1* and *TFAP2B* expression, respectively [57]. In Villani and co-workers' work on characterizing blood cell types [58], the dendritic cells in

their scRNAseq study were sequenced by utilizing Smart-Seq2 and analyzed with Seurat software. Two new subtypes of dendritic cells were discovered through sequencing analysis, finding the unique expression of *CD1C/BDCA-1<sup>+</sup>* cDC2 cells [58]. The first subpopulation expressed *CD1CA*, while the other expressed *CD1CB*, and neither subtype was related to any previously characterized dendritic cell type [58].



**Figure 2.** Normalization capabilities of several different RNAseq methods. Each method was scaled in comparison to the other tested methods, creating a scaled score for each method. Reproduced with permission from Tian, Nature Methods; published by Springer Nature, 2019 [51].

Machine learning uses algorithms and statistical computational models to analyze megadata. It has recently been utilized to observe the number of unique transcripts necessary to identify the cell type of a tested cell [53]. Aevermann et al. utilized NS-Forest v2.0 machine learning software to characterize cell type in a sNuc-Seq study of cells from each cortical layer of the middle temporal gyrus [53]. Machine learning utilizes known marker genes to classify cell clusters into specific cell types and further uses the genome expression information of these classified cell types to identify cell types in future datasets. In another study, Li et al. compared machine learning techniques for the cluster analysis of pancreatic cells from scRNAseq datasets [52]. The DESC deep neural network machine learning software was able to perform clustering stability and accuracy analyses comparable to other standard analysis methods, such as scVI, MNN, Seurat3.0, CCA, BERMUDA, and scanorama [52]. Furthermore, DESC was able to preserve the biological variation, remove the batch effect from the datasets, and allow for the biological interpretation of pseudo-temporal and discrete cellular structures [52]. Notably, cluster analysis and batch effect removal are usually accomplished with the simultaneous use of multiple different analysis software methods, while DESC can accomplish it with a single method [52]. Machine learning can present baseline information as to the number of unique genes displayed by a cell to be accurately characterized [59]. Peng et al. found that unique gene expression can be the result of RNA editing and not of unique cell types [59]. This creates the same issue as when the sequencing analysis does not properly filter outliers in that there will be false positives in determining DGE for that particular cell type [59]. Hence, predicted gene expression using machine learning-based methods after RNAseq should be validated by means of other biological characterizations.

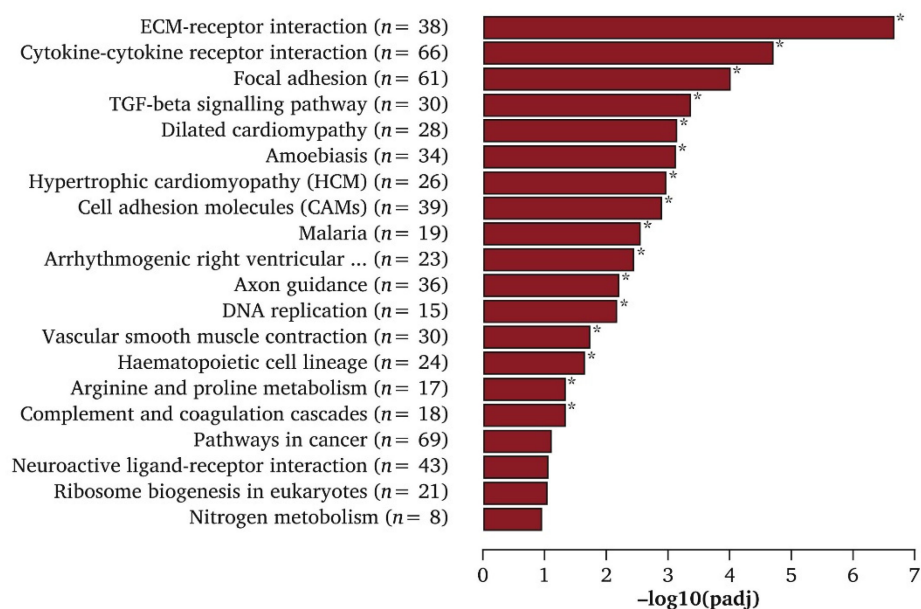
## 5. Transcriptomics-Enabled Human Stem Cell Differentiation

RNAseq technologies have been extensively applied to the generation of a variety of cells and tissues from hPSCs in the past decade since differential expression characterization is vital for determining how cells develop and mature during the course of the stepwise differentiation of stem cells. The best way to analyze stem cell differentiation through RNAseq is to perform RNAseq at different differentiation stages to acquire important information regarding the specific genes crucial for the differentiation process [26,60,61]. Nair et al. used a 20-day stepwise differentiation protocol to differentiate hESCs into endocrine  $\beta$ -like cells and purified insulin-GFP-expressing  $\beta$ -like cells by flow cytometry, followed by the aggregation of these immature insulin-GFP+ cells with extended culture. They compared the transcriptome differences between the hESC-derived immature  $\beta$ -cells after 20 days without aggregation, cells after aggregation and extended culture, and native  $\beta$ -cells isolated from adult human islets using RNAseq analysis. They found that the isolation of hESC-derived immature insulin-GFP-expressing cells and the aggregation of these cells into islet sizes with extended culture allowed the generation of mature  $\beta$ -cells with functions similar to those of human islet  $\beta$  cells [62]. These stem-cell-derived cells showed dynamic insulin secretion and increased calcium signaling in response to insulin secretagogues. Wang et al. utilized EdgeR to determine DEGs in human mesenchymal stem cell (MSC) differentiation into endothelial cells. Differentially expressed genes in endothelial cells compared to MSCs include *HIPK2*, *GREM1*, *LEF1*, and *EFNB2* [63]. The upregulations of these genes, along with seven other genes, namely, *ADGRA2*, *CHRNA7*, *LRG1*, *NTRK1*, *S100A9*, *MMRN2*, and *RAPGEF3*, in MSC differentiation toward endothelial cells is an indication of endothelial specification by RNAseq analysis [63]. Through RNAseq analysis at different stages of MSC differentiation, specifically toward endothelial cells, the study found that ECM organization, angiogenesis, blood vessel morphogenesis, and growth-factor- and ECM-binding pathways all had enriched expression at the later stage of differentiation [63].

The process of analyzing RNAseq results at varying time points is known as temporal RNAseq. DyNB and DESeq analysis methods were utilized by Åijö et al. to assess temporal RNAseq data [60]. Both analysis methods are run with an assumption that ignores correlations between time points, which allows for the testing of each time point for DEGs independently of time. This reduces the need for the regularization of the data and provides estimations of gene expression throughout differentiation, as well as shows transcriptome responses to differentiation [60]. Th17 cells are a subclass of T helper lymphocyte cells that express interleukin 17 (IL17) [60]. *ISG20* and *IL17A* are known to have roles in Th17 cell function. Their expression was initiated after 48 h post-TH17 differentiation and showed increased expression thereafter, as quantified by temporal RNAseq analysis [60]. By analyzing stem cells at different time points, one can determine how gene expression changes between each time point. This can show not only the differentiation pathway of the cell throughout its lineage commitment but also reveal new cell subpopulations or intermediates during differentiation. Wang et al. demonstrated that signaling pathways involved in ECM–receptor interactions, TGF- $\beta$ , and cytokine–cytokine-receptor interactions had the highest expression for MSC differentiation toward endothelial cells [63]. The top Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways for MSC-to-endothelial-cell differentiation are shown in Figure 3. The TGF- $\beta$  pathway is consistently downregulated across associated genes, such as *NOG*, *ID4*, *INHBA*, *INHBB*, *INHBE*, *THBS2*, *BMP8B*, *COMP*, and *CHRD*, during endothelial specification [63]. However, ECM organization genes such as *MMP1*, *ELANE*, *CTSG*, *MMP10*, *KDR*, *MYF5*, *VWF*, *PDGFB*, and *COL17A1* are all significantly upregulated for MSC-to-endothelial-cell differentiation [63]. These results indicate that ECM interaction and organization-related gene upregulation and TGF- $\beta$  inhibition are important for endothelial specification. RNAseq can also increase the knowledge of the differentiation mechanisms involved in the specification of cells derived from stem cells [63]. Such sequencing could give insight into the regulatory networks that are utilized to promote cell specification and lineage commitment [26]. This can be



useful in determining the balance between the proliferation and differentiation of stem cells [26]. Certain data analysis techniques are also able to discern cell cycle processes that are closely aligned to cellular differentiation [61]. Temporal RNAseq allows for more complete data acquisition compared to normal sequencing analysis. However, cell differentiation still presents issues even with temporal sequencing analysis. For instance, single-cell temporal RNAseq can reveal the differentiation of a single cell, but there is no guarantee that analyzing at the same time points during the differentiation of another cell of the same type will yield the same results. The second cell could differentiate at a much different rate, suggesting that comparing time points is not completely accurate. As an alternative approach, same-time temporal modeling eliminates this issue, as there is no need for replicates to observe each different differentiation step [60]. Same-time temporal sequencing is the process of sequencing cells simultaneously, each at a different stage of differentiation into the desired cell type. This removes variable differentiation rate errors since specific cells are selected at specific stages of differentiation for sequencing and data analysis [60].



**Figure 3.** Top KEGG enrichment pathways of differentially expressed genes (DEGs) in endothelial cells generated from MSCs. \* Indicated significantly enriched KEGG pathway terms found in induced endothelial cells compared with MSCs. Reproduced with permission from Wang, European Journal of Vascular and Endovascular Surgery; published by Elsevier, 2020 [63].

Table 3 exhibits diverse RNAseq methods that aid in hPSC differentiation into varied cell types. Collin et al. focused on the development of retinal organoids from hESCs and further characterized the cell populations within the organoid through temporal scRNAseq using integrated fluidic circuits [64]. Transcriptomic data for multiple time points during differentiation were collected to characterize the orderly appearance of retinal cell types as organoid development progressed [64]. Additionally, scRNAseq has been applied to hESC differentiation into pancreatic  $\beta$ -cells [65] and heart organoid tissue [66] (Table 3). hESC-to- $\beta$ -cell differentiation was tracked at multiple points throughout differentiation to characterize genomic expression changes during  $\beta$ -cell development [65]. Spatiotemporal scRNAseq was utilized to map the cell type distribution and genome expression in the human heart during embryonic development to better understand cardiac morphogenesis [66]. Spatiotemporal scRNAseq has also been applied to understanding how hypoxia plays a role in iPSC differentiation toward liver bud cells throughout organoid development [67]. Kidwai et al. assessed the progress of hESC and iPSC differentiation into osteogenic progenitor cells through bulk RNAseq and focused on transcriptomic differences between osteogenic progenitors derived from the paraxial mesoderm, lateral plate

mesoderm, and neural crest [68]. Tsujimoto et al. applied both bulk and scRNAseq to characterize iPSC differentiation toward kidney-lineage-specific cells, such as metanephric and mesonephric nephron progenitors, as well as ureteric bud cells, to gain a better understanding of the mechanisms that are behind kidney development [69]. They focused on early renal development to characterize the origin of embryonic kidney cell types derived from the intermediate mesoderm to better characterize early kidney development in kidney organoids. Furthermore, RNAseq and transcriptome characterization can be utilized for disease treatment. Wang et al. assessed the transcriptome data of endothelial cell differentiation from MSCs by utilizing bulk RNAseq, aiming to understand the mechanism underlying differentiation for potential cell-based treatment that can reduce ischemic symptoms in patients with peripheral artery disease [63] (Table 3).

**Table 3.** RNAseq methods applied to hPEC differentiation.

Cell Line	RNA Sequencing Method	Terminally Differentiated Cell Type	Reference
hESC H9	Integrated fluidic circuit scRNAseq	Retinal cells	[64]
Human bone-marrow-derived MSCs	Bulk RNAseq	Endothelial cells	[63]
iPSCs	Bulk RNAseq: KAPA Stranded mRNAseq Kit with BCL2FASTQ v1.8.4 software	Metanephric nephron progenitors, mesonephric nephron progenitors, ureteric bud	[69]
iPSCs	scRNAseq: ddSEQ with BCL2FASTQ v1.8.4 software	Metanephric nephron progenitors, mesonephric nephron progenitors, ureteric bud	[69]
iPSC	Genome-wide scRNAseq	Liver bud	[65,67]
hESC H9	Genome-wide scRNAseq	Pancreatic $\beta$ -cells	[65]
Human embryo-derived developmental heart tissue	scRNAseq with CellRanger analysis with Seurat v2.3.4 software	Heart-derived cells: cardiomyocytes, endothelium, immune, epicardium, fibroblasts, smooth muscle, cardiac neural crest	[66]
NCRM-5 human iPSCs from male CD34 <sup>+</sup> cord blood	Bulk RNAseq with FASTQC, GENCODE v25, and STAR v2.5.2a software	Osteogenic progenitors	[68]
hESC H9	Bulk RNAseq with FASTQC, GENCODE v25, and STAR v2.5.2a software	Osteogenic progenitors	[68]

Huang et al. focused on analyzing the transcriptomes of pancreatic  $\beta$ -cells derived from iPSCs, looking specifically for DEGs [70]. iPSC generation of  $\beta$ -cells was shown to be possible in both healthy patients and diabetic patients. The main process of iPSC-to- $\beta$ -cell differentiation involved the generation of embryoid bodies from the initial iPSCs, the differentiation of embryoid bodies into multipotent progenitors, and the induction of  $\beta$ -cells from the multipotent progenitors. Differentiation was aided by increased signaling via the TGF- $\beta$ , PI3K-AKT, and MAPK signaling pathways in later pancreatic differentiation toward a  $\beta$ -cell fate [70]. Specific gene expression favoring  $\beta$ -cell differentiation includes genes associated with each of the highly expressed signaling pathways. The  $\beta$ -cell fate is tied to the high expression of signaling pathway genes, such as TGF- $\beta$  genes (*TNF*, *TGFB2*, and *THSB1*), PI3K genes (*ITGB4*, *FGF3*, *LAMA1*, *JAK2*, and *FLT1*), and MAPK genes (*TNF*, *FGF3*, *TGFB2*, and *FGFR2*) [70]. Liang et al. also focused on how transcriptomic expression can benefit stem cell differentiation but focused on dopaminergic progenitor generation from hESCs [71]. Dopaminergic progenitors are located in the midbrain and are part of the differentiation pathway that progresses from hESCs to radial glial cells to early floor plate progenitor cells to dopaminergic progenitor cells. Favorable transcriptomic signatures have been discovered for each step in the differentiation process from hESCs to dopaminergic progenitors. The initial high expression of *SOX2*, *FABP7*, *FEZF1*, *TCF7L2*, and *HMGB2* promotes radial glial specification and eventually gives way to the high expression of floor plate markers such as *LMX1A*, *CORIN*, *OTX2*, and *FOXA2* [71]. The final high expression of dopaminergic progenitor markers, including *TH*, *NR4A2*, and *PITX3*, demonstrates effective differentiation into dopaminergic progenitor cells from hESCs [71].

Aside from characterizing known cell types and discovering new unique cell subtypes, RNAseq analysis has been used to observe how substrates alter cellular gene expression and in cell differentiation studies [26,72]. Carrow et al. reported how the presence of synthetic clay nanosilicates can alter the cellular transcription profile in human MSCs and prompt differentiation toward an osteochondral cell fate [72]. Gene expression analysis indicates that nanosilicate introduction negatively impacts the expression of the multipotent and motility markers *WNT5A*, *AFAP1*, *SOCS5*, and *INHBA* in MSCs while increasing the expression of the bone and cartilage development genes *COMP*, *COL1A1*, *ACAN*, and *COL11A* [72]. The data support the theory that nanosilicates alter the transcriptome expression in human MSCs toward an osteochondral cell fate. Lange et al. developed a toolkit, named CellRank, for single-cell fate mapping after scRNAseq. The toolkit helps to predict the cell fate trajectory during cellular differentiation [73]. It predicted both the terminal state and fate potential of the analyzed cell. This is accomplished by looking at the genome expression in the analyzed cell and comparing it to all of the possible cell types that the progenitor could terminally differentiate into [73]. Based on DGE, the analysis software predicts the most likely cell fate. While this is not an exact prediction, the software is capable of determining the level of uncertainty in the lineage commitment prediction [73].

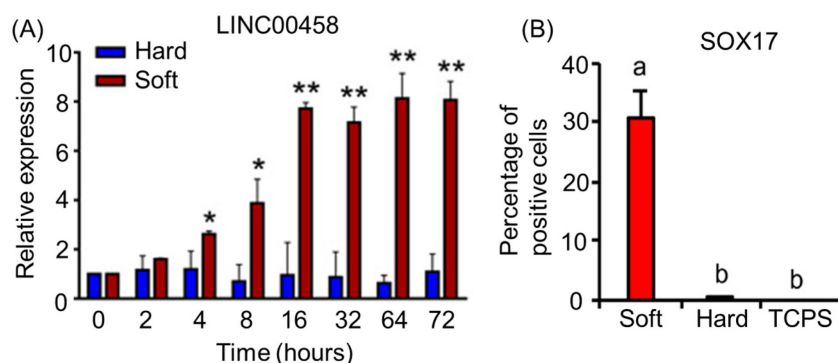
## 6. Transcriptomics-Aided Discovery of Intrinsic Factors Influencing hPSC Lineage Commitment

Transcription factors bind to DNA and directly influence target genes' transcription and expression in the cells [74]. Hence, transcription factors play a pivotal role in regulating cellular differentiation. Cells cannot further specify the desired lineage without expressing specific transcription factors [75]. Reimold reported that the transcription factor XBP-1 is selectively and specifically required for the generation of plasma cells. Heller et al. analyzed RNAseq and chromatin immunoprecipitation-sequencing (ChIPseq) data to understand transcriptional and regulatory mechanisms during hPSC pancreatic development [76]. They discovered that *GATA6* and *FOXA2* are responsible for inducing hESC differentiation toward a definitive endoderm cell fate, which will later differentiate into pancreatic and endocrine progenitors alongside the expression of other transcription factors, including *ONECUT1*, *PDX1*, *SOX9*, and *NKX6.1* [76]. Using the RNAseq approach, Jennings et al. identified transcription factors responsible for initiating the development of either the pancreas or liver [77]. *PDX1*, *GATA6*, *FOXA2*, *HNF1B*, and *ONECUT1* control pancreatic endoderm commitment from hPSCs, while *HNF4A* and *FOXA* are key transcription factors regulating the onset of liver specification from hPSCs [77].

Transcription factor enhancement or repression could make a difference in late-stage differentiation into two separate committed cell lines. Yagi et al. demonstrated how *GATA3* expression influences the late-stage differentiation of CD4+ T cells into either Th1 or Th2 cells [78]. As transcription factors are capable of binding to DNA to directly influence genes that are transcribed and expressed in a cell, gene expression is directly influenced and controlled by transcription factors. Ge et al. confirmed that *ATOH7*, *OTX2*, *POU4F2*, and *ISL1* all influence the differentiation of retinal progenitors toward a retinal ganglion cell fate by attaching to the genome and influencing gene expression through RNAseq data analysis [79]. Due to the significant relationship between transcription factor presence and specific gene expression, transcription factors are vital for cell lineage commitment.

Notably, RNAseq technologies allow the sequencing of not only transcription factors but also long non-coding RNA (lncRNA) and microRNA (miRNA). lncRNA and miRNA have been implicated in cellular differentiation and are able to regulate gene expression [80–83]. miRNAs are small RNA transcripts, and lncRNAs make up the majority of the transcriptome [82,83]. However, lncRNAs have a complex overlap of transcripts that makes the discovery of their function difficult [82]. Through RNAseq analysis, Chen et al. discovered that the expression of *LINC00458*, a lncRNA, induces hPSCs toward an endodermal cell fate, provided that the cells are cultured on soft substrates [84]. The influence of substrate stiffness on *LINC00458* expression in cultured hPSCs is shown in Figure 4A.

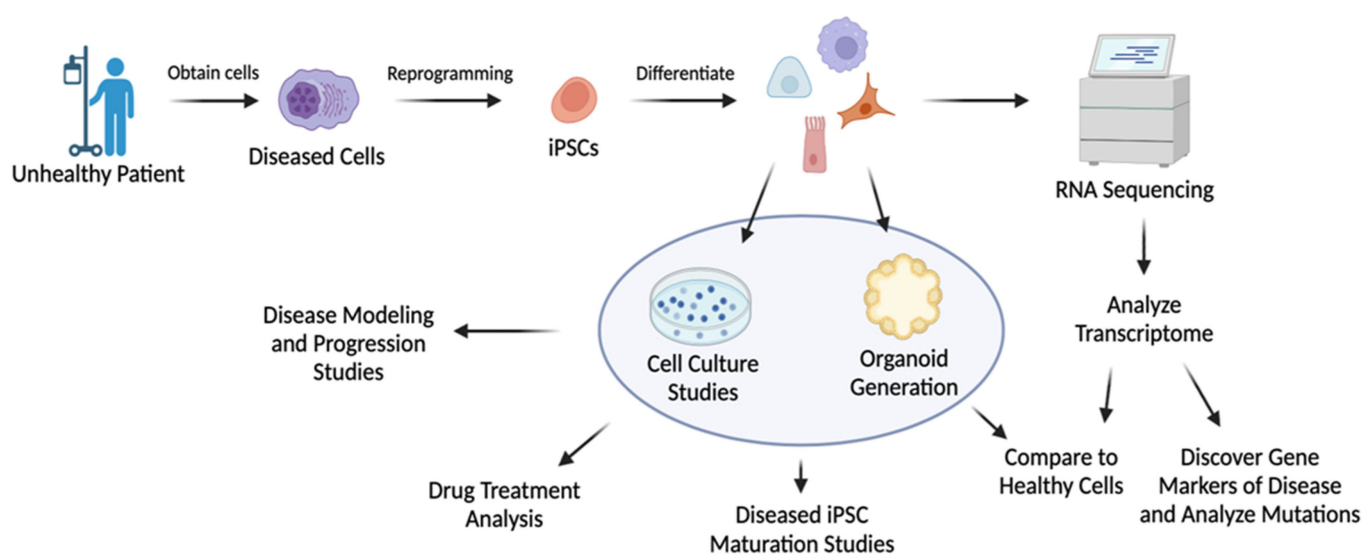
Different expression levels of this lncRNA resulted in varied endodermal cell commitment, as quantified by SOX17 protein expression levels (Figure 4B) [84]. Increasing the stiffness of the substrate has been shown to negatively influence *LINC00458* expression, stunting differentiation into endodermal lineages. Soft substrates are favorable for inducing *LINC00458* expression, which in turn interacts with *SMAD2/3* within the nucleus. The knockout of *LINC00458* also resulted in the loss of endodermal marker genes such as *FOXA2* and *SOX17*, indicating the importance of this lncRNA in hPSC-to-endoderm differentiation [84]. Jha et al. showed the lncRNA *GATA6-AS1* is necessary for proper differentiation from hPSCs to cardiomyocytes. This lncRNA expression is directly related to the expression level of the *GATA6* gene [80]. It was shown that the knockout of *GATA6-AS1* restricted cardiomyocyte lineage specification through the direct loss of *GATA6* expression, which is essential for cardiomyocyte development from hPSCs [80]. On the other hand, *miR-17~92* has been shown to repress *CXCR5* expression in developing T cells, which results in inhibited T-cell generation and migration. However, *miR-17~92* is also necessary for the repression of *TH17*, *Th22*, and other associated incorrect lineages' gene expression during T-cell differentiation into T follicular helper cells [83]. These repressed genes include *RORA*, *CCR6*, *ILLR2*, *ILLR1*, and *IL22* [83]. By means of miRNAseq and RNAseq, miRNA has also been discovered to impact human placental development studies using the early stages of placental samples [81]. These studies strongly suggest that RNAseq techniques facilitate the discovery of miRNAs that play a role in the differentiation and development of cells.



**Figure 4.** (A) Expression of lncRNA *LINC00458* in hPSCs cultured on hard (165 kPa) or soft (3 kPa) substrates. Cells were cultured on soft and hard substrates for 3 days. RNA was isolated at the indicated time points, and *LINC00458* expression was detected by qRT-PCR. The results are presented as means  $\pm$  SD of triplicates. \*\*  $p < 0.005$ , \*  $p < 0.05$ , two-sided Student's *t* test. (B) Percentage of SOX17+ cells in hPSCs grown on substrates with varied stiffnesses. The results are presented as means  $\pm$  SD of triplicates. One-way ANOVA ( $n = 3$  independent experiments). Different letters indicate significant differences, and the same letters indicate no significant differences. The data are adapted from Chen et al. [84].

## 7. Transcriptomics Applied to Disease Physiology Studies Using Patients' iPSC-Derived Cells for Regenerative Medicine

Transcriptome analysis is beneficial for understanding disease and providing insight into potential regenerative medicines capable of mitigating or treating disease. This commonly encompasses utilizing iPSCs derived from patients' tissues to generate diseased organoids for research (Table 4) [85–91]. Figure 5 summarizes the key steps in disease modeling through a combination of iPSC technology and transcriptomics for regenerative medicine. For instance, iPSC-derived brain cells have been utilized to mimic the progression of Alzheimer's disease (AD) to study the transcriptomic effect on cells. These patient-derived brain cells obtained from iPSC differentiation showed genotypes and phenotypes associated with AD [85]. Zhao et al. generated cerebral organoids from AD dementia patient-derived iPSCs. Transcriptomic profiling revealed that the cerebral organoids originating from AD patients showed abnormal cytoplasmic RNA granules and deranged RNA metabolism [92]. The studies provide a more in-depth understanding of the disease.



**Figure 5.** Model of diseased iPSC generation and its use for regenerative studies and disease characterization. Created using Biorender.com.

Another disease in which transcriptomic analysis has been utilized for potential regenerative medicine is Parkinson's disease (PD). PD studies currently lack dopaminergic neuron samples, as these neurons die off at an increasing rate with the progression of the disease, leaving few neuronal samples post-mortem [86]. iPSCs derived from fibroblasts have been generated with the *PINK1* gene mutation *ILE368ASN* to imitate PD dopaminergic neurons [86]. These mutated neurons were utilized to analyze the dysregulation of neuronal function as a result of PD (Table 4). Novak et al. theorized that PD is not the result of a single mutation but rather the effect of transcriptomic dysregulation from several interconnected genes and signaling pathways [86]. Azevedo et al. also analyzed PD through iPSCs but instead analyzed how myelination was prevented during oligodendrocyte specification and maturation [93]. This loss of maturation in PD patients leads to a lower transcriptional contribution of oligodendrocytes compared to healthy controls [87]. The utilization of iPSCs developed from PD patients led to the discovery that alterations in the cell's transcriptome are associated with the loss of myelination in oligodendrocytes [93]. For example, mutations in the *SNCA* gene are known to play a role in the loss of oligodendrocyte myelination in PD [93]. *A53T  $\alpha$ -synuclein* and *G2019S LRRK2* mutations have also been implicated in the formation of PD [87]. *LMX1B* and *OTX2* are suggested to influence PD pathogenesis [87]. Numerous genes were found to be enriched or depleted in PD-patient-derived oligodendrocytes through transcriptomic analysis, suggesting the involvement of these genes in PD. iPSC-derived oligodendrocytes also showed the involvement of the inflammatory components C4b and HLA proteins, promoting oligodendrocyte immune reactivity rather than myelination [93]. Moreover, studies in iPSC-derived PD models are able to determine familial-PD- and sporadic-PD-related gene expression, including the expression levels of miRNA and piRNA, for understanding PD gene deregulation. miRNA and piRNA transcriptome expression is altered as a result of PD presence (Table 4) [87]. It was shown that 26, 34, and 40 different miRNAs were dysregulated in the PD group compared to the healthy control group for fibroblasts, iPSCs, and neurons, respectively. The increased expression of *LET7*-family miRNAs occurs consistently in PD patient samples. *SINE*- and *LINE*-family piRNA overexpression is associated with PD, as diseased neurons cannot silence these genes properly [87]. Temporal RNAseq analysis has been utilized to analyze DEGs for mutated PD dopaminergic neurons [86]. Dopamine metabolism genes *TH* and *DCC* were significantly expressed in mutated dopaminergic neurons alongside *NES* and *VIM*, which are cytoskeletal protein genes. Both *NES* and *VIM* are associated with cytoskeletal transport, a process commonly impacted by PD [86]. Other common

PD-related genes include *LG11* and *CNTNAP2*, which are both dysregulated throughout disease progression, although they do not have direct links to the disease [86].

Recently, transcriptome analysis became a powerful tool for regenerative medicine applications in diabetes research thanks to the continual improvement of protocols for hPSC–endocrine cell differentiation [2–4,94–97]. Technologies for iPSC differentiation into insulin-secreting  $\beta$ -cells have been applied to generate  $\beta$ -cells or endocrine tissue from patients suffering from different subtypes of diabetes, such as type I diabetes (T1D), type II diabetes (T2D), and monogenic diabetes (MD) (Table 4) [88,98–100]. Memon et al. focused on utilizing T2D-derived iPSCs to study the mechanism behind insulin resistance in T2D [89]. RNAseq analysis of insulin-resistant iPSCs discovered the downregulation of *L1TD1*, *RIF1*, *MYSM1*, *ZNF195*, *ZNF208*, and *ZNF770* compared to healthy controls [89]. Upregulated genes in the insulin-resistant group compared to healthy controls include *AGRN*, *SLFN13*, *CD74*, *SEMA6B*, *SLC22A17*, and *TMEM151B* [89]. The T2D iPSC group had increased lactate secretion and increased oxidative stress in the cells, suggesting lowered oxygen availability to the diseased islet cells. Oxidative stress also induces mitochondrial dysfunction, which compounds the impact of insulin resistance on affected islet cells [89]. A combination of diseased iPSC-derived cells and transcriptome analysis has been utilized to understand the mechanism and transcriptomic changes causing diabetes. For example, it helped to reveal the genetic mechanism underlying transcription factor *PDX1*-mutation-associated impaired glucose tolerance in patients with an increased risk for diabetes [98]. Using a *PDX1*-mutant-derived iPSC line and subsequent differentiation, the study deduced that amino acid mutations in *PDX1* impair pancreatic endocrine formation and  $\beta$ -cell function, contributing to a predisposition for diabetes [98]. The downregulation of the *PDX1*-bound genes *MNX1*, *MEG3*, and *CES1* is attributed to the lowered  $\beta$ -cell differentiation ability [98]. Augsornworawat et al. discovered that  $\beta$ -cell maturation occurs in both iPSC- and hESC-derived islets post-implantation [101]. Maturation was assessed based on the comparison of  $\beta$ -cell maturation gene expression levels before and after implantation through RNAseq data analysis. The  $\beta$ -cell maturation genes *INS*, *MAFA*, *MNX1*, *SIX2*, and *G6PC2* were all expressed in higher quantities in the post-transplantation group. The increased maturation and functionality of the transplanted islets suggest that native tissue gene expression can be acquired once the iPSC-derived cells are transplanted in vivo. Both  $\alpha$ - and  $\beta$ -cells undergo increased maturation once they are transplanted in vivo, further suggesting that the in vivo environment and native tissue influence genome expression toward more mature cells [101]. Furthermore, RNAseq data analysis allowed the identification of transcriptomic abnormalities. Mutated transcription factors can be corrected through genome editing using tools such as CRISPR/Cas9 to restore the proper functionality in the diseased islets [91,102]. These genes include but are not limited to *HNF1A* [103], *INS* [99,104], and *GATA6* [105].

In addition, hypertrophic cardiomyopathy (HCM) is another disease in which transcriptomic information can be uncovered from diseased iPSC-derived cells. Patient tissue was utilized to generate iPSCs, which were then developed into cardiomyocytes for transcriptomic analysis studies focused on discovering genes related to HCM functionality (Table 4) [90]. Approximately 50% of HCM patients carry mutations in sarcomere genes, such as *MYH7* and *MYBPC3* [106]. The *MYBPC3* *c.1928-569G > T* mutation was discovered as one of the causes of HCM through iPSC-derived cardiomyocyte transcriptome analysis. This mutation results in aberrant splicing, which is commonly undetected in genetic testing [90]. Not only does this method uncover a new cause for HCM, but it also opens the door for research into the therapeutic inhibition of aberrant splicing, which could treat the disease [90]. iPSC-derived cardiomyocytes have also been used for disease modeling and RNAseq analysis of congenital heart disease (CHD). Xu et al. utilized iPSC-derived cardiomyocytes to model hypoplastic left heart syndrome by analyzing the impact of the defects caused by the disease [107]. scRNAseq showed a correlation between the number of DEGs and disease severity compared to the healthy control group. An increased number of DEGs indicated the increased severity of disease progression [107]. The iPSC-derived

cardiomyocytes that were from patients with less severe disease progression had gene expression more similar to the healthy control iPSC-derived cardiomyocytes when compared to patients with more severe disease progression. Many of the DEGs correspond to signaling pathways in mitochondria, hypoxia, cell death, or apoptosis. These pathways all corresponded to increased mitochondrial dysfunction in the diseased groups that caused severe oxidative stress and increased cell death [107].

**Table 4.** Patients' iPSC-derived cells for disease modeling by means of transcriptomics.

Disease Type	Patient Cells Utilized	Final Cell Types	Purpose	References
Alzheimer's disease	iPSCs	Neural progenitors, astrocytes, microglia, and oligodendrocytes	Brain cells generated from AD patients' iPSCs to mimic disease progression by transcriptomic analysis	[85]
Parkinson's disease	iPSCs	Dopaminergic neurons	A PD patient's dopaminergic neurons generated from iPSCs for scRNAseq analysis	[86]
	iPSCs	Oligodendrocytes	Transcriptome analysis of PD patient iPSC-derived oligodendrocytes to determine cause of reduced myelination	[93]
	iPSCs	Midbrain neurons	iPSC-derived model to study the effect of mRNA, miRNA, and piRNA on PD progression	[87]
Diabetes	iPSCs	Insulin-resistant iPSCs	T2D-patient-derived iPSCs to study insulin resistance mechanisms such as increased oxidative stress and lactate secretion	[89]
	iPSCs	Pancreatic $\beta$ cells	Patient iPSCs to study the effect of mutant PDX1 expression on impaired glucose tolerance	[98]
	hESCs and iPSCs	Pancreatic $\alpha$ , $\beta$ , and $\delta$ cells	iPSC-derived islets further mature after implantation into the body based on RNAseq data analysis	[101]
Hypertrophic cardiomyopathy	iPSCs	Cardiomyocytes	Transcriptomic analysis of iPSC-derived cardiomyocytes from diseased patients	[90]
Congenital heart disease	iPSCs	Cardiomyocytes	Modeling hypoplastic left heart syndrome through iPSC-derived cardiomyocytes	[107]

## 8. Challenges and Limitations

A limitation of current RNAseq is the necessity of cDNA reverse transcription for the sufficient scale-up of genomic information [13]. The challenge with utilizing cDNA comes with many manipulation steps necessary for RNAseq. Manipulation is necessary to remove second-strand cDNA from strand-specific sequencing, account for cDNA library generation-related dissociation, and synthesize cDNA without independent primer synthesis. Some genes are uncooperative with RNAseq and will not provide accurate results. Uncooperative genes are typically small in base-pair number and produce a significant batch effect when sequenced utilizing different library preparation techniques [108]. When using unique versus multiple alignment, some genes seemed to be differentially expressed in one of the methods rather than both [108]. This occurrence suggests that the differentially expressed gene could be a false-positive value, which throws off the accuracy of the data interpretation. On top of uncooperative genes, some RNAseq methods have limited reproducibility [109]. NOIseq is one analysis method that cannot be reproduced, as the noise from one dataset to another is not consistent [44]. Another major challenge to RNAseq analysis is analyzing low-input samples or samples that have become degraded [34]. The majority of analysis methods are limited to analyzing full transcripts, and incomplete transcripts lead to a highly inaccurate interpretation of the transcriptome. For more accurate results, studies must be completed using productive RNAseq and analysis methods, together with experimental confirmation. For degraded samples, more effective analysis methods include RNase H and Ribo-Zero [34]. Due to the availability of different types of analysis software, only a limited number of methods have been significantly investigated in a research context [109]. More studies will have to be completed to validate the accuracy of the software programs and the resulting transcriptomic data.

Another major limitation of RNAseq and the associated data analysis is the variability in the sequencing results [108,110]. RNAseq has had a limited number of sample replicates for many cell types, resulting in the lack of sufficient data to compare the new samples

to [55]. Efforts must be made to increase the number of samples sequenced for each known cell type to increase the accuracy of the analysis and comparison to previous methods. Increasing the sample size will decrease the deviation of gene expression for each cell type and allow for the comparison of new samples to a more uniform dataset of genome libraries. The false discovery of DEGs or new cell subtypes is another source of error [55,108]. This can partially be caused by RNA editing of a cell's transcriptome [59]. No two cells are identical due to biological variations regarding how RNA is transcribed, which could lead to increased error for the analyzed dataset. Moreover, environmental changes of any magnitude can potentially alter the identification of DEGs during RNAseq analysis [111]. This could be from using different batches for sequencing or having the batches collected at different time points within the experiment. Environmental changes could also occur during RNA isolation from contamination or slight variations in how the RNA libraries are obtained and prepared [111]. Temporal RNAseq analysis is limited by the variability between replicates due to differentiation efficiency [60]. This can throw off the data observed during stem cell differentiation analysis. Some analysis software, such as DyNB, has implemented filtering that accounts for the variable timing of differentiation, which significantly reduces the error [60].

On the other hand, RNAseq analysis for genes that have low copy numbers or are expressed at low levels presents challenges. Filters must be applied to separate the low-copy-number genes from the outlier genes and other associated noise [112]. Noise is the result of biological variation within each cell as well as transcript loss during the preparation of the cDNA library [113]. Genes from regions of low density can sometimes be falsely classified as noise due to the assumptions necessary to run RNAseq analysis [113]. Hence, properly eliminating noise and distinguishing noise from a low copy number of genes are pivotal to obtaining highly accurate DGE and cluster formation analysis results. Furthermore, the sequencing depth and the number of genes detected in each biological sample influence RNAseq analysis accuracy [112,114–116]. The sequencing depth, otherwise known as the size of the sequenced library, varies from sample to sample due to biological variation between samples [114]. Therefore, normalization must be performed to accurately compare different genes and datasets. Zyprych-Walczak et al. investigated multiple data normalization methods, including the EdgeR packages TMM and Upper Quartile, the DESeq package Median, the EBSeq package Quantile, and the PoissonSeq package PoissonSeq normalization [114]. They noted that TMM performed poorly, but there was not one method that performed significantly better than the others out of the remaining methods [114]. It was also noted that bias and variance should be analyzed as a part of normalization in order to find a method that is best suited to analyze the dataset and prevent differential expression errors [114]. Furthermore, current stem cell differentiation methods cannot obtain homogeneity of the final derived cellular product. A portion of subpopulations of cells still differentiate into undesired cell types. By determining the genomic expression of the cells, the goal is to gain an understanding of how the cells differentiate and why subpopulations form. Once this is known, the undesired cell clusters can be eliminated to achieve the desired differentiation pathways.

## 9. Future Direction of RNAseq-Aided Stem Cell Differentiation

The future of RNAseq should look toward addressing the limitations of the current methods. Sample multiplexing for scRNAseq is one method that would enable the super-loading of single cells and the unbiased analysis and discovery of cell types [117,118]. Multiplexing scRNAseq has advanced to the point where over  $10^5$  cells can be sequenced at once [117]. RNAseq has a variation in read depths when using different sequencing methods, which results in a different number of transcripts generated by sequencing. This variability is an issue for comparison and accuracy. This can increase the error in the analysis between that dataset and another that utilized a different analysis method. Normalization methods such as those described by Zyprych-Walczak et al. must be perfected to generate accurate and comparable RNAseq data [114]. There should be



a focus on finding a universal method for RNAseq and analysis that provides accurate results regardless of experimental conditions [119]. This necessitates the generation of a method that has little to no assumptions necessary to properly characterize the dataset. Assumptions limit the accuracy of the data analysis and further restrict comparisons between datasets analyzed by different methods [54]. Finding a universal method will not only standardize the experimental procedure necessary for proper sequencing but also eliminate the variability and error in the results generated and compared by using different sequencing and analysis methods.

Spatial single-cell transcriptomics is another aspect of RNAseq that can be focused on more in future stem cell research. It provides cell type information at the spatial level in histological sections. Analyzing the RNAseq data in this manner allows for heterogeneity and tissue architecture to be observed in a spatial context at the cellular level [28–31]. Moncada et al. were able to map multi-cell-type locations and identified that populations of ductal cells, macrophages, dendritic cells, and cancer cells are spatially restricted, along with distinctly co-localized with other cell types, using a combination of spatial transcriptomics and scRNAseq techniques [29]. The combination of these pseudo-spatial techniques, called multimodal intersection analysis, was used to detect cell type and subtype enrichment across different locations within a specific tissue microenvironment, such as pancreatic ductal adenocarcinoma [29]. Spatial transcriptomics has been applied to embryos to characterize and identify the location and makeup of the developing organs [28]. Liu et al. utilized the Seurat3.2 module SCTransform for data normalization and stabilization, while spatial variability and differential expression were determined by SpatialDE and ToppGene, respectively, for their “deterministic barcoding in tissue for spatial omics sequencing” [28]. The method was utilized to characterize the spatial clustering of major cell types, including the heart, telencephalon, neural tube, hindlimb bud, and branchial arches, in a developing embryo.

Spatial transcriptomics can also be paired with temporal transcriptomics in RNAseq analysis to give an even more complete picture of gene expression over both space and time [28,31]. This approach allows for the analysis of cell interactions that drive differentiation and the transcriptomic profile associated with those interactions at specific locations over the time period of development. Hence, the lineage commitment of specific parts of the organ, as well as each cell cluster’s progression into specific cell types, can be tracked and mapped [28]. The prospect of analyzing RNA expression and cellular interactions over space and time allows for a better understanding of cellular interactions and development. Spatiotemporal RNAseq technology has been applied to viewing organ development and differentiation in chicken models [28], mouse organogenesis [32], human cancer research [29], thymus organogenesis in human embryos [33], and human fetal digestive tract development [31]. However, its application to stem cell differentiation is currently limited due to the high complexity of differentiated cells and the cost of spatiotemporal RNAseq [28]. This technology should be further improved to analyze developing organs and differentiating cells. The next steps should be toward implementing organogenesis observation with RNAseq to further characterize development in a spatiotemporal manner.

Taken together, stem cell differentiation has yet to accomplish pure lineage commitment toward the desired mature cell types. Global transcriptome analysis by RNAseq provides a powerful characterization tool to ensure the systematic assessment of the differentiated cells’ identities and identify key signaling pathways critical for the improvement of lineage commitment. Spatial and temporal transcriptomics alongside RNAseq would give more complex data involving cell–cell interactions, tissue architecture, and the regulatory networks that control stem cell differentiation and maturation. RNAseq would assist in the full characterization of human organoids to benefit in vitro organoid development for potential implantation into the body in the near future.

**Author Contributions:** D.A.O. wrote the manuscript. S.J. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is partially supported by the National Institutes of Health (EB027391-01) and the National Science Foundation (CBET1928855 and CBET1919830).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data supporting this review article are available in individual citations.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Chen, S.S.; Fitzgerald, W.; Zimmerberg, J.; Kleinman, H.K.; Margolis, L. Cell-cell and cell-extracellular matrix interactions regulate embryonic stem cell differentiation. *Stem Cells* **2007**, *25*, 553–561. [[CrossRef](#)] [[PubMed](#)]
2. Bi, H.; Karanth, S.S.; Ye, K.; Stein, R.; Jin, S. Decellularized Tissue Matrix Enhances Self-Assembly of Islet Organoids from Pluripotent Stem Cell Differentiation. *ACS Biomater. Sci. Eng.* **2020**, *6*, 4155–4165. [[CrossRef](#)] [[PubMed](#)]
3. Bi, H.; Ye, K.; Jin, S. Proteomic analysis of decellularized pancreatic matrix identifies collagen V as a critical regulator for islet organogenesis from human pluripotent stem cells. *Biomaterials* **2020**, *233*, 119673. [[CrossRef](#)] [[PubMed](#)]
4. Heaton, E.S.; Jin, S. Importance of multiple endocrine cell types in islet organoids for type 1 diabetes treatment. *Transl. Res.* **2022**, *250*, 68–83. [[CrossRef](#)] [[PubMed](#)]
5. Pajeroski, J.D.; Dahl, K.N.; Zhong, F.L.; Sammak, P.J.; Discher, D.E. Physical plasticity of the nucleus in stem cell differentiation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15619–15624. [[CrossRef](#)]
6. Huang, H.; Bader, T.N.; Jin, S. Signaling molecules regulating pancreatic endocrine development from pluripotent stem cell differentiation. *Int. J. Mol. Sci.* **2020**, *21*, 5867. [[CrossRef](#)]
7. McNamara, L.E.; McMurray, R.J.; Biggs, M.J.; Kantawong, F.; Oreffo, R.O.; Dalby, M.J. Nanotopographical control of stem cell differentiation. *J. Tissue Eng.* **2010**, *1*, 120623. [[CrossRef](#)]
8. Park, J.; Cho, C.H.; Parashurama, N.; Li, Y.; Berthiaume, F.; Toner, M.; Tilles, A.W.; Yarmush, M.L. Microfabrication-based modulation of embryonic stem cell differentiation. *Lab Chip* **2007**, *7*, 1018–1028. [[CrossRef](#)]
9. Dayem, A.A.; Choi, H.Y.; Yang, G.M.; Kim, K.; Saha, S.K.; Kim, J.H.; Cho, S.G. The potential of nanoparticles in stem cell differentiation and further therapeutic applications. *Biotechnol. J.* **2016**, *11*, 1550–1560. [[CrossRef](#)]
10. Yap, L.; Tay, H.G.; Nguyen, M.T.; Tjin, M.S.; Tryggvason, K. Laminins in cellular differentiation. *Trends Cell Biol.* **2019**, *29*, 987–1000. [[CrossRef](#)]
11. Busskamp, V.; Lewis, N.E.; Guye, P.; Ng, A.H.; Shipman, S.L.; Byrne, S.M.; Sanjana, N.E.; Murn, J.; Li, Y.; Li, S. Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **2014**, *10*, 760. [[CrossRef](#)] [[PubMed](#)]
12. Christensen, J.L.; Weissman, I.L. Flk-2 is a marker in hematopoietic stem cell differentiation: A simple method to isolate long-term stem cells. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14541–14546. [[CrossRef](#)] [[PubMed](#)]
13. Oszolak, F.; Milos, P.M. RNA sequencing: Advances, challenges and opportunities. *Nat. Rev. Genet.* **2011**, *12*, 87–98. [[CrossRef](#)] [[PubMed](#)]
14. Han, X.; Chen, H.; Huang, D.; Chen, H.; Fei, L.; Cheng, C.; Huang, H.; Yuan, G.-C.; Guo, G. Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. *Genome Biol.* **2018**, *19*, 47. [[CrossRef](#)] [[PubMed](#)]
15. Kolodziejczyk, A.A.; Kim, J.K.; Svensson, V.; Marioni, J.C.; Teichmann, S.A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **2015**, *58*, 610–620. [[CrossRef](#)]
16. Jovic, D.; Liang, X.; Zeng, H.; Lin, L.; Xu, F.; Luo, Y. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin. Transl. Med.* **2022**, *12*, e694. [[CrossRef](#)]
17. Stark, R.; Grzelak, M.; Hadfield, J. RNA sequencing: The teenage years. *Nat. Rev. Genet.* **2019**, *20*, 631–656. [[CrossRef](#)]
18. Kukurba, K.R.; Montgomery, S.B. RNA sequencing and analysis. *Cold Spring Harb. Protoc.* **2015**, *2015*, 951–969. [[CrossRef](#)]
19. Ziegenhain, C.; Vieth, B.; Parekh, S.; Reinius, B.; Guillaumet-Adkins, A.; Smets, M.; Leonhardt, H.; Heyn, H.; Hellmann, I.; Enard, W. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **2017**, *65*, 631–643.e634. [[CrossRef](#)]
20. Oszolak, F.; Platt, A.R.; Jones, D.R.; Reifemberger, J.G.; Sass, L.E.; McInerney, P.; Thompson, J.F.; Bowers, J.; Jarosz, M.; Milos, P.M. Direct RNA sequencing. *Nature* **2009**, *461*, 814–818. [[CrossRef](#)]
21. Wu, A.R.; Neff, N.F.; Kalisky, T.; Dalerba, P.; Treutlein, B.; Rothenberg, M.E.; Mburu, F.M.; Mantalas, G.L.; Sim, S.; Clarke, M.F. Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **2014**, *11*, 41–46. [[CrossRef](#)]
22. Ding, J.; Adiconis, X.; Simmons, S.K.; Kowalczyk, M.S.; Hession, C.C.; Marjanovic, N.D.; Hughes, T.K.; Wadsworth, M.H.; Burks, T.; Nguyen, L.T. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **2020**, *38*, 737–746. [[CrossRef](#)] [[PubMed](#)]
23. Adiconis, X.; Haber, A.L.; Simmons, S.K.; Levy Moonshine, A.; Ji, Z.; Busby, M.A.; Shi, X.; Jacques, J.; Lancaster, M.A.; Pan, J.Q. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat. Methods* **2018**, *15*, 505–511. [[CrossRef](#)] [[PubMed](#)]

24. Sun, X.; Sun, S.; Yang, S. An efficient and flexible method for deconvoluting bulk RNA-seq data with single-cell RNA-seq data. *Cells* **2019**, *8*, 1161. [[CrossRef](#)] [[PubMed](#)]
25. Janjic, A.; Wange, L.E.; Bagnoli, J.W.; Geuder, J.; Nguyen, P.; Richter, D.; Vieth, B.; Vick, B.; Jeremias, I.; Ziegenhain, C. Prime-seq, efficient and powerful bulk RNA sequencing. *Genome Biol.* **2022**, *23*, 88. [[CrossRef](#)]
26. Trapnell, C.; Cacchiarelli, D.; Grimsby, J.; Pokharel, P.; Li, S.; Morse, M.; Lennon, N.J.; Livak, K.J.; Mikkelsen, T.S.; Rinn, J.L. Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat. Biotechnol.* **2014**, *32*, 381. [[CrossRef](#)]
27. Conesa, A.; Nueda, M.J. Next-masigpro: Dealing with RNA-seq time series. *EMBnet J.* **2013**, *19*, 42–43. [[CrossRef](#)]
28. Liu, Y.; Yang, M.; Deng, Y.; Su, G.; Enniful, A.; Guo, C.C.; Tebaldi, T.; Zhang, D.; Kim, D.; Bai, Z. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **2020**, *183*, 1665–1681.e1618. [[CrossRef](#)]
29. Moncada, R.; Barkley, D.; Wagner, F.; Chiodin, M.; Devlin, J.C.; Baron, M.; Hajdu, C.H.; Simeone, D.M.; Yanai, I. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **2020**, *38*, 333–342. [[CrossRef](#)]
30. Mantri, M.; Scuderi, G.J.; Abedini-Nassab, R.; Wang, M.F.; McKellar, D.; Shi, H.; Grodner, B.; Butcher, J.T.; De Vlamincq, I. Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis. *Nat. Commun.* **2021**, *12*, 1771. [[CrossRef](#)]
31. Gao, S.; Yan, L.; Wang, R.; Li, J.; Yong, J.; Zhou, X.; Wei, Y.; Wu, X.; Wang, X.; Fan, X. Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat. Cell Biol.* **2018**, *20*, 721–734. [[CrossRef](#)]
32. Chen, A.; Liao, S.; Cheng, M.; Ma, K.; Wu, L.; Lai, Y.; Qiu, X.; Yang, J.; Xu, J.; Hao, S. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **2022**, *185*, 1777–1792.e1721. [[CrossRef](#)] [[PubMed](#)]
33. Zeng, Y.; Liu, C.; Gong, Y.; Bai, Z.; Hou, S.; He, J.; Bian, Z.; Li, Z.; Ni, Y.; Yan, J. Single-cell RNA sequencing resolves spatiotemporal development of pre-thymic lymphoid progenitors and thymus organogenesis in human embryos. *Immunity* **2019**, *51*, 930–948.e936. [[CrossRef](#)] [[PubMed](#)]
34. Adiconis, X.; Borges-Rivera, D.; Satija, R.; DeLuca, D.S.; Busby, M.A.; Berlin, A.M.; Sivachenko, A.; Thompson, D.A.; Wysocker, A.; Fennell, T. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* **2013**, *10*, 623–629. [[CrossRef](#)] [[PubMed](#)]
35. Schuierer, S.; Carbone, W.; Knehr, J.; Petitjean, V.; Fernandez, A.; Sultan, M.; Roma, G. A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. *BMC Genom.* **2017**, *18*, 442. [[CrossRef](#)] [[PubMed](#)]
36. Sanborn, M.A.; Li, T.; Victor, K.; Siegfried, H.; Fung, C.; Rothman, A.L.; Srikiatkhachorn, A.; Fernandez, S.; Ellison, D.; Jarman, R.G. Analysis of cell-associated DENV RNA by oligo (dT) primed 5' capture scRNAseq. *Sci. Rep.* **2020**, *10*, 9047. [[CrossRef](#)] [[PubMed](#)]
37. Svoboda, M.; Frost, H.R.; Bosco, G. Internal oligo (dT) priming introduces systematic bias in bulk and single-cell RNA sequencing count data. *NAR Genom. Bioinform.* **2022**, *4*, lqac035. [[CrossRef](#)]
38. Katz, Y.; Wang, E.T.; Airoidi, E.M.; Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **2010**, *7*, 1009–1015. [[CrossRef](#)]
39. Freedman, A.H.; Gaspar, J.M.; Sackton, T.B. Short paired-end reads trump long single-end reads for expression analysis. *BMC Bioinform.* **2020**, *21*, 149. [[CrossRef](#)]
40. Bacher, R.; Kendziorowski, C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* **2016**, *17*, 63. [[CrossRef](#)]
41. Andrews, T.S.; Kiselev, V.Y.; McCarthy, D.; Hemberg, M. Tutorial: Guidelines for the computational analysis of single-cell RNA sequencing data. *Nat. Protoc.* **2021**, *16*, 1–9. [[CrossRef](#)]
42. Chung, M.; Bruno, V.M.; Rasko, D.A.; Cuomo, C.A.; Muñoz, J.F.; Livny, J.; Shetty, A.C.; Mahurkar, A.; Dunning Hotopp, J.C. Best practices on the differential expression analysis of multi-species RNA-seq. *Genome Biol.* **2021**, *22*, 121. [[CrossRef](#)] [[PubMed](#)]
43. Liu, S.; Wang, Z.; Zhu, R.; Wang, F.; Cheng, Y.; Liu, Y. Three differential expression analysis methods for RNA sequencing: Limma, EdgeR, DESeq2. *JoVE (J. Vis. Exp.)* **2021**, *175*, e62528. [[CrossRef](#)]
44. Costa-Silva, J.; Domingues, D.; Lopes, F.M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **2017**, *12*, e0190152. [[CrossRef](#)] [[PubMed](#)]
45. Langmead, B.; Hansen, K.D.; Leek, J.T. Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biol.* **2010**, *11*, R83. [[CrossRef](#)]
46. Kalari, K.R.; Nair, A.A.; Bhavsar, J.D.; O'Brien, D.R.; Davila, J.I.; Bockol, M.A.; Nie, J.; Tang, X.; Baheti, S.; Doughty, J.B. MAP-RSeq: Mayo analysis pipeline for RNA sequencing. *BMC Bioinform.* **2014**, *15*, 224. [[CrossRef](#)]
47. Rajkumar, A.P.; Qvist, P.; Lazarus, R.; Lescai, F.; Ju, J.; Nyegaard, M.; Mors, O.; Børghlum, A.D.; Li, Q.; Christensen, J.H. Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. *BMC Genom.* **2015**, *16*, 548. [[CrossRef](#)]
48. Luecken, M.D.; Theis, F.J. Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* **2019**, *15*, e8746. [[CrossRef](#)]
49. Wang, T.; Li, B.; Nelson, C.E.; Nabavi, S. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. *BMC Bioinform.* **2019**, *20*, 40. [[CrossRef](#)]

50. Dal Molin, A.; Baruzzo, G.; Di Camillo, B. Single-cell RNA-sequencing: Assessment of differential expression analysis methods. *Front. Genet.* **2017**, *8*, 62. [[CrossRef](#)]
51. Tian, L.; Dong, X.; Freytag, S.; Lê Cao, K.-A.; Su, S.; JalalAbadi, A.; Amann-Zalcenstein, D.; Weber, T.S.; Seidi, A.; Jabbari, J.S. Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nat. Methods* **2019**, *16*, 479–487. [[CrossRef](#)] [[PubMed](#)]
52. Li, X.; Wang, K.; Lyu, Y.; Pan, H.; Zhang, J.; Stambolian, D.; Susztak, K.; Reilly, M.P.; Hu, G.; Li, M. Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nat. Commun.* **2020**, *11*, 2338. [[CrossRef](#)] [[PubMed](#)]
53. Aevermann, B.; Zhang, Y.; Novotny, M.; Keshk, M.; Bakken, T.; Miller, J.; Hodge, R.; Lelieveldt, B.; Lein, E.; Scheuermann, R.H. A machine learning method for the discovery of minimum marker gene combinations for cell type identification from single-cell RNA sequencing. *Genome Res.* **2021**, *31*, 1767–1780. [[CrossRef](#)] [[PubMed](#)]
54. Varet, H.; Brillet-Guéguen, L.; Coppée, J.-Y.; Dillies, M.-A. SARTools: A DESeq2-and EdgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. *PLoS ONE* **2016**, *11*, e0157022. [[CrossRef](#)] [[PubMed](#)]
55. Zhou, X.; Lindsay, H.; Robinson, M.D. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Res.* **2014**, *42*, e91. [[CrossRef](#)]
56. Buettner, F.; Natarajan, K.N.; Casale, F.P.; Proserpio, V.; Scialdone, A.; Theis, F.J.; Teichmann, S.A.; Marioni, J.C.; Stegle, O. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **2015**, *33*, 155–160. [[CrossRef](#)]
57. Clark, B.S.; Stein-O'Brien, G.L.; Shiau, F.; Cannon, G.H.; Davis-Marcisak, E.; Sherman, T.; Santiago, C.P.; Hoang, T.V.; Rajaii, F.; James-Esposito, R.E. Single-cell RNA-seq analysis of retinal development identifies NFI factors as regulating mitotic exit and late-born cell specification. *Neuron* **2019**, *102*, 1111–1126.e1115. [[CrossRef](#)]
58. Villani, A.-C.; Satija, R.; Reynolds, G.; Sarkizova, S.; Shekhar, K.; Fletcher, J.; Griesbeck, M.; Butler, A.; Zheng, S.; Lazo, S. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **2017**, *356*, eaah4573. [[CrossRef](#)]
59. Peng, Z.; Cheng, Y.; Tan, B.C.-M.; Kang, L.; Tian, Z.; Zhu, Y.; Zhang, W.; Liang, Y.; Hu, X.; Tan, X. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* **2012**, *30*, 253–260. [[CrossRef](#)]
60. Åijö, T.; Butty, V.; Chen, Z.; Salo, V.; Tripathi, S.; Burge, C.B.; Lahesmaa, R.; Lähdesmäki, H. Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. *Bioinformatics* **2014**, *30*, i113–i120. [[CrossRef](#)]
61. Rizvi, A.H.; Camara, P.G.; Kandrór, E.K.; Roberts, T.J.; Schieren, I.; Maniatis, T.; Rabadan, R. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* **2017**, *35*, 551–560. [[CrossRef](#)] [[PubMed](#)]
62. Nair, G.G.; Liu, J.S.; Russ, H.A.; Tran, S.; Saxton, M.S.; Chen, R.; Juang, C.; Li, M.-I.; Nguyen, V.Q.; Giacometti, S. Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived  $\beta$  cells. *Nat. Cell Biol.* **2019**, *21*, 263–274. [[CrossRef](#)] [[PubMed](#)]
63. Wang, C.; Liu, H.; Yang, M.; Bai, Y.; Ren, H.; Zou, Y.; Yao, Z.; Zhang, B.; Li, Y. RNA-seq based transcriptome analysis of endothelial differentiation of bone marrow mesenchymal stem cells. *Eur. J. Vasc. Endovasc. Surg.* **2020**, *59*, 834–842. [[CrossRef](#)] [[PubMed](#)]
64. Collin, J.; Queen, R.; Zerti, D.; Dorgau, B.; Hussain, R.; Coxhead, J.; Cockell, S.; Lako, M. Deconstructing retinal organoids: Single cell RNA-Seq reveals the cellular components of human pluripotent stem cell-derived retina. *Stem Cells* **2019**, *37*, 593–598. [[CrossRef](#)]
65. Li, X.; Yang, K.Y.; Chan, V.W.; Leung, K.T.; Zhang, X.-B.; Wong, A.S.; Chong, C.C.; Wang, C.C.; Ku, M.; Lui, K.O. Single-cell RNA-seq reveals that CD9 is a negative marker of glucose-responsive pancreatic  $\beta$ -like cells derived from human pluripotent stem cells. *Stem Cell Rep.* **2020**, *15*, 1111–1126. [[CrossRef](#)]
66. Asp, M.; Giacomello, S.; Larsson, L.; Wu, C.; Fürth, D.; Qian, X.; Wårdell, E.; Custodio, J.; Reimegård, J.; Salmén, F. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **2019**, *179*, 1647–1660.e1619. [[CrossRef](#)]
67. Ayabe, H.; Anada, T.; Kamoya, T.; Sato, T.; Kimura, M.; Yoshizawa, E.; Kikuchi, S.; Ueno, Y.; Sekine, K.; Camp, J.G. Optimal hypoxia regulates human iPSC-derived liver bud differentiation through intercellular TGF $\beta$  signaling. *Stem Cell Rep.* **2018**, *11*, 306–316. [[CrossRef](#)]
68. Kidwai, F.; Mui, B.W.; Arora, D.; Iqbal, K.; Hockaday, M.; Castro Diaz, L.F.; Cherman, N.; Martin, D.; Myneni, V.D.; Ahmad, M. Lineage-specific differentiation of osteogenic progenitors from pluripotent stem cells reveals the FGF1-RUNX2 association in neural crest-derived osteoprogenitors. *Stem Cells* **2020**, *38*, 1107–1123. [[CrossRef](#)]
69. Tsujimoto, H.; Kasahara, T.; Sueta, S.-i.; Araoka, T.; Sakamoto, S.; Okada, C.; Mae, S.-i.; Nakajima, T.; Okamoto, N.; Taura, D. A modular differentiation system maps multiple human kidney lineages from pluripotent stem cells. *Cell Rep.* **2020**, *31*, 107476. [[CrossRef](#)]
70. Huang, Y.; Wan, J.; Guo, Y.; Zhu, S.; Wang, Y.; Wang, L.; Guo, Q.; Lu, Y.; Wang, Z. Transcriptome Analysis of Induced Pluripotent Stem Cell (iPSC)-derived Pancreatic  $\beta$ -like Cell Differentiation. *Cell Transplant.* **2017**, *26*, 1380–1391. [[CrossRef](#)]
71. Liang, L.; Tian, Y.; Feng, L.; Wang, C.; Feng, G.; Stacey, G.N.; Shyh-Chang, N.; Wu, J.; Hu, B.; Li, W. Single-cell transcriptomics reveals the cell fate transitions of human dopaminergic progenitors derived from hESCs. *Stem Cell Res. Ther.* **2022**, *13*, 412. [[CrossRef](#)] [[PubMed](#)]
72. Carrow, J.K.; Cross, L.M.; Reese, R.W.; Jaiswal, M.K.; Gregory, C.A.; Kaunas, R.; Singh, I.; Gaharwar, A.K. Widespread changes in transcriptome profile of human mesenchymal stem cells induced by two-dimensional nanosilicates. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3905–E3913. [[CrossRef](#)] [[PubMed](#)]

73. Lange, M.; Bergen, V.; Klein, M.; Setty, M.; Reuter, B.; Bakhti, M.; Lickert, H.; Ansari, M.; Schniering, J.; Schiller, H.B. CellRank for directed single-cell fate mapping. *Nat. Methods* **2022**, *19*, 159–170. [[CrossRef](#)] [[PubMed](#)]
74. Mitsis, T.; Efthimiadou, A.; Bacopoulou, F.; Vlachakis, D.; Chrousos, G.P.; Eliopoulos, E. Transcription factors and evolution: An integral part of gene expression. *World Acad. Sci. J.* **2020**, *2*, 3–8. [[CrossRef](#)]
75. Reimold, A.M.; Iwakoshi, N.N.; Manis, J.; Vallabhajosyula, P.; Szomolanyi-Tsuda, E.; Gravalles, E.M.; Friend, D.; Grusby, M.J.; Alt, F.; Glimcher, L.H. Plasma cell differentiation requires the transcription factor XBP-1. *Nature* **2001**, *412*, 300–307. [[CrossRef](#)] [[PubMed](#)]
76. Heller, S.; Li, Z.; Lin, Q.; Geusz, R.; Breunig, M.; Hohwieler, M.; Zhang, X.; Nair, G.G.; Seufferlein, T.; Hebrok, M. Transcriptional changes and the role of ONECUT1 in hPSC pancreatic differentiation. *Commun. Biol.* **2021**, *4*, 1298. [[CrossRef](#)] [[PubMed](#)]
77. Jennings, R.E.; Berry, A.A.; Gerrard, D.T.; Wearne, S.J.; Strutt, J.; Withey, S.; Chhatriwala, M.; Hanley, K.P.; Vallier, L.; Bobola, N. Laser capture and deep sequencing reveals the transcriptomic programmes regulating the onset of pancreas and liver differentiation in human embryos. *Stem Cell Rep.* **2017**, *9*, 1387–1394. [[CrossRef](#)]
78. Yagi, R.; Zhu, J.; Paul, W.E. An updated view on transcription factor GATA3-mediated regulation of Th1 and Th2 cell differentiation. *Int. Immunol.* **2011**, *23*, 415–420. [[CrossRef](#)]
79. Ge, Y.; Chen, X.; Nan, N.; Bard, J.; Wu, F.; Yergeau, D.; Liu, T.; Wang, J.; Mu, X. Key transcription factors influence the epigenetic landscape to regulate retinal cell differentiation. *Nucleic Acids Res.* **2023**, *51*, 2151–2176. [[CrossRef](#)]
80. Jha, R.; Li, D.; Wu, Q.; Ferguson, K.E.; Forghani, P.; Gibson, G.C.; Xu, C. A long non-coding RNA GATA6-AS1 adjacent to GATA6 is required for cardiomyocyte differentiation from human pluripotent stem cells. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2020**, *34*, 14336. [[CrossRef](#)]
81. Awamleh, Z.; Gloor, G.B.; Han, V.K. Placental microRNAs in pregnancies with early onset intrauterine growth restriction and preeclampsia: Potential impact on gene expression and pathophysiology. *BMC Med. Genom.* **2019**, *12*, 91. [[CrossRef](#)] [[PubMed](#)]
82. Fatica, A.; Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* **2014**, *15*, 7–21. [[CrossRef](#)] [[PubMed](#)]
83. Baumjohann, D.; Kageyama, R.; Clingan, J.M.; Morar, M.M.; Patel, S.; de Kouchkovsky, D.; Bannard, O.; Bluestone, J.A.; Matloubian, M.; Ansel, K.M. miR-17~ 92 promotes T follicular helper cell differentiation and represses subset-inappropriate gene expression. *Nat. Immunol.* **2013**, *14*, 840. [[CrossRef](#)] [[PubMed](#)]
84. Chen, Y.-F.; Li, Y.-S.J.; Chou, C.-H.; Chiew, M.Y.; Huang, H.-D.; Ho, J.H.-C.; Chien, S.; Lee, O.K. Control of matrix stiffness promotes endodermal lineage specification by regulating SMAD2/3 via lncRNA LINC00458. *Sci. Adv.* **2020**, *6*, eaay0264. [[CrossRef](#)]
85. Penney, J.; Ralvenius, W.T.; Tsai, L.-H. Modeling Alzheimer’s disease with iPSC-derived brain cells. *Mol. Psychiatry* **2020**, *25*, 148–167. [[CrossRef](#)]
86. Novak, G.; Kyriakis, D.; Grzyb, K.; Bernini, M.; Rodius, S.; Dittmar, G.; Finkbeiner, S.; Skupin, A. Single-cell transcriptomics of human iPSC differentiation dynamics reveal a core molecular network of Parkinson’s disease. *Commun. Biol.* **2022**, *5*, 49. [[CrossRef](#)]
87. Schulze, M.; Sommer, A.; Plötz, S.; Farrell, M.; Winner, B.; Grosch, J.; Winkler, J.; Riemenschneider, M.J. Sporadic Parkinson’s disease derived neuronal cells show disease-specific mRNA and small RNA signatures with abundant deregulation of piRNAs. *Acta Neuropathol. Commun.* **2018**, *6*, 58. [[CrossRef](#)]
88. Maxwell, K.G.; Millman, J.R. Applications of iPSC-derived beta cells from patients with diabetes. *Cell Rep. Med.* **2021**, *2*, 100238. [[CrossRef](#)]
89. Memon, B.; Elsayed, A.K.; Bettahi, I.; Suleiman, N.; Younis, I.; Wehedy, E.; Abou-Samra, A.B.; Abdelalim, E.M. iPSCs derived from insulin resistant offspring of type 2 diabetic patients show increased oxidative stress and lactate secretion. *Stem Cell Res. Ther.* **2022**, *13*, 428. [[CrossRef](#)]
90. Holliday, M.; Singer, E.S.; Ross, S.B.; Lim, S.; Lal, S.; Ingles, J.; Semsarian, C.; Bagnall, R.D. Transcriptome sequencing of patients with hypertrophic cardiomyopathy reveals novel splice-altering variants in MYBPC3. *Circ. Genom. Precis. Med.* **2021**, *14*, e003202. [[CrossRef](#)]
91. Maxwell, K.G.; Augsornworawat, P.; Velazco-Cruz, L.; Kim, M.H.; Asada, R.; Hoglebe, N.J.; Morikawa, S.; Urano, F.; Millman, J.R. Gene-edited human stem cell-derived  $\beta$  cells from a patient with monogenic diabetes reverse preexisting diabetes in mice. *Sci. Transl. Med.* **2020**, *12*, eaax9106. [[CrossRef](#)] [[PubMed](#)]
92. Zhao, J.; Fu, Y.; Yamazaki, Y.; Ren, Y.; Davis, M.; Liu, C.; Lu, W.; Wang, X.; Chen, K.; Cherukuri, Y. APOE4 exacerbates synapse loss and neurodegeneration in Alzheimer’s disease patient iPSC-derived cerebral organoids. *Nat. Commun.* **2020**, *11*, 5540. [[CrossRef](#)]
93. Azevedo, C.; Teku, G.; Pomeschchik, Y.; Reyes, J.F.; Chumarina, M.; Russ, K.; Savchenko, E.; Hammarberg, A.; Lamas, N.J.; Collin, A. Parkinson’s disease and multiple system atrophy patient iPSC-derived oligodendrocytes exhibit alpha-synuclein-induced changes in maturation and immune reactive properties. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2111405119. [[CrossRef](#)]
94. Rezanian, A.; Bruin, J.E.; Arora, P.; Rubin, A.; Batushansky, I.; Asadi, A.; O’dwyer, S.; Quiskamp, N.; Mojobian, M.; Albrecht, T. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **2014**, *32*, 1121–1133. [[CrossRef](#)]
95. Wesolowska-Andersen, A.; Jensen, R.R.; Alcántara, M.P.; Beer, N.L.; Duff, C.; Nylander, V.; Gosden, M.; Witty, L.; Bowden, R.; McCarthy, M.I. Analysis of differentiation protocols defines a common pancreatic progenitor molecular signature and guides refinement of endocrine differentiation. *Stem Cell Rep.* **2020**, *14*, 138–153. [[CrossRef](#)] [[PubMed](#)]

96. Singh, R.; Cottle, L.; Loudovaris, T.; Xiao, D.; Yang, P.; Thomas, H.E.; Kebede, M.A.; Thorn, P. Enhanced structure and function of human pluripotent stem cell-derived beta-cells cultured on extracellular matrix. *Stem Cells Transl. Med.* **2021**, *10*, 492–505. [[CrossRef](#)] [[PubMed](#)]
97. Karanth, S.S.; Sun, S.; Bi, H.; Ye, K.; Jin, S. Angiopoietins stimulate pancreatic islet development from stem cells. *Sci. Rep.* **2021**, *11*, 13558. [[CrossRef](#)] [[PubMed](#)]
98. Wang, X.; Sterr, M.; Burtscher, I.; Böttcher, A.; Beckenbauer, J.; Siehler, J.; Meitinger, T.; Häring, H.-U.; Staiger, H.; Cernilogar, F.M. Point mutations in the PDX1 transactivation domain impair human  $\beta$ -cell development and function. *Mol. Metab.* **2019**, *24*, 80–97. [[CrossRef](#)]
99. Ma, S.; Viola, R.; Sui, L.; Cherubini, V.; Barbetti, F.; Egli, D.  $\beta$  cell replacement after gene editing of a neonatal diabetes-causing mutation at the insulin locus. *Stem Cell Rep.* **2018**, *11*, 1407–1415. [[CrossRef](#)]
100. Millman, J.R.; Xie, C.; Van Dervort, A.; Gürtler, M.; Pagliuca, F.W.; Melton, D.A. Generation of stem cell-derived  $\beta$ -cells from patients with type 1 diabetes. *Nat. Commun.* **2016**, *7*, 11463. [[CrossRef](#)]
101. Augsornworawat, P.; Maxwell, K.G.; Velazco-Cruz, L.; Millman, J.R. Single-cell transcriptome profiling reveals  $\beta$  cell maturation in stem cell-derived islets after transplantation. *Cell Rep.* **2020**, *32*, 108067. [[CrossRef](#)] [[PubMed](#)]
102. Lithovius, V.; Saarimäki-Vire, J.; Balboa, D.; Ibrahim, H.; Montaser, H.; Barsby, T.; Otonkoski, T. SUR1-mutant iPSC cell-derived islets recapitulate the pathophysiology of congenital hyperinsulinism. *Diabetologia* **2021**, *64*, 630–640. [[CrossRef](#)] [[PubMed](#)]
103. Cardenas-Diaz, F.L.; Osorio-Quintero, C.; Diaz-Miranda, M.A.; Kishore, S.; Leavens, K.; Jobaliya, C.; Stanescu, D.; Ortiz-Gonzalez, X.; Yoon, C.; Chen, C.S. Modeling monogenic diabetes using human ESCs reveals developmental and metabolic deficiencies caused by mutations in HNF1A. *Cell Stem Cell* **2019**, *25*, 273–289.e275. [[CrossRef](#)] [[PubMed](#)]
104. Balboa, D.; Saarimäki-Vire, J.; Borshagovski, D.; Survila, M.; Lindholm, P.; Galli, E.; Euroola, S.; Ustinov, J.; Grym, H.; Huopio, H. Insulin mutations impair beta-cell development in a patient-derived iPSC model of neonatal diabetes. *eLife* **2018**, *7*, e38519. [[CrossRef](#)]
105. Kishore, S.; De Franco, E.; Cardenas-Diaz, F.L.; Letourneau-Freiberg, L.R.; Sanyoura, M.; Osorio-Quintero, C.; French, D.L.; Greeley, S.A.W.; Hattersley, A.T.; Gadue, P. A non-coding disease modifier of pancreatic agenesis identified by genetic correction in a patient-derived iPSC line. *Cell Stem Cell* **2020**, *27*, 137–146.e136. [[CrossRef](#)]
106. Marian, A.J.; Braunwald, E. Hypertrophic cardiomyopathy: Genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. *Circ. Res.* **2017**, *121*, 749–770. [[CrossRef](#)]
107. Xu, X.; Jin, K.; Bais, A.S.; Zhu, W.; Yagi, H.; Feinstein, T.N.; Nguyen, P.K.; Criscione, J.D.; Liu, X.; Beutner, G. Uncompensated mitochondrial oxidative stress underlies heart failure in an iPSC-derived model of congenital heart disease. *Cell Stem Cell* **2022**, *29*, 840–855.e847. [[CrossRef](#)]
108. Hirsch, C.D.; Springer, N.M.; Hirsch, C.N. Genomic limitations to RNA sequencing expression profiling. *Plant J.* **2015**, *84*, 491–503. [[CrossRef](#)]
109. Simoneau, J.; Dumontier, S.; Gosselin, R.; Scott, M.S. Current RNA-seq methodology reporting limits reproducibility. *Brief. Bioinform.* **2021**, *22*, 140–145. [[CrossRef](#)]
110. Fonseca, N.A.; Marioni, J.; Brazma, A. RNA-seq gene profiling—a systematic empirical comparison. *PLoS ONE* **2014**, *9*, e107026. [[CrossRef](#)]
111. Koch, C.M.; Chiu, S.F.; Akbarpour, M.; Bharat, A.; Ridge, K.M.; Bartom, E.T.; Winter, D.R. A beginner’s guide to analysis of RNA sequencing data. *Am. J. Respir. Cell Mol. Biol.* **2018**, *59*, 145–157. [[CrossRef](#)] [[PubMed](#)]
112. Griffith, M.; Griffith, O.L.; Mwenifumbo, J.; Goya, R.; Morrissy, A.S.; Morin, R.D.; Corbett, R.; Tang, M.J.; Hou, Y.-C.; Pugh, T.J. Alternative expression analysis by RNA sequencing. *Nat. Methods* **2010**, *7*, 843–847. [[CrossRef](#)] [[PubMed](#)]
113. Xu, C.; Su, Z. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* **2015**, *31*, 1974–1980. [[CrossRef](#)] [[PubMed](#)]
114. Zypych-Walczak, J.; Szabelska, A.; Handschuh, L.; Górczak, K.; Klamecka, K.; Figlerowicz, M.; Siatkowski, I. The impact of normalization methods on RNA-Seq data analysis. *BioMed Res. Int.* **2015**, *2015*, 621690. [[CrossRef](#)] [[PubMed](#)]
115. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 13. [[CrossRef](#)]
116. Rapaport, F.; Khanin, R.; Liang, Y.; Pirun, M.; Krek, A.; Zumbo, P.; Mason, C.E.; Socci, N.D.; Betel, D. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.* **2013**, *14*, R95. [[CrossRef](#)]
117. Zhang, Y.; Xu, S.; Wen, Z.; Gao, J.; Li, S.; Weissman, S.M.; Pan, X. Sample-multiplexing approaches for single-cell sequencing. *Cell. Mol. Life Sci.* **2022**, *79*, 466. [[CrossRef](#)]
118. Islam, S.; Kjällquist, U.; Moliner, A.; Zajac, P.; Fan, J.-B.; Lönnerberg, P.; Linnarsson, S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **2011**, *21*, 1160–1167. [[CrossRef](#)]
119. Whitley, S.K.; Horne, W.T.; Kolls, J.K. Research techniques made simple: Methodology and clinical applications of RNA sequencing. *J. Investig. Dermatol.* **2016**, *136*, e77–e82. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.