OXFORD

Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data

Shun H. Yip, Pak Chung Sham and Junwen Wang

Corresponding author: Junwen Wang, 13400 E Shea Blvd., Scottsdale, AZ 85259, USA. Tel.: +1-480-301-4644; Fax: +1-480-301-8387; E-mail: wang.junwen@mayo.edu

Abstract

Traditional RNA sequencing (RNA-seq) allows the detection of gene expression variations between two or more cell populations through differentially expressed gene (DEG) analysis. However, genes that contribute to cell-to-cell differences are not discoverable with RNA-seq because RNA-seq samples are obtained from a mixture of cells. Single-cell RNA-seq (scRNA-seq) allows the detection of gene expression in each cell. With scRNA-seq, highly variable gene (HVG) discovery allows the detection of genes that contribute strongly to cell-to-cell variation within a homogeneous cell population, such as a population of embryonic stem cells. This analysis is implemented in many software packages. In this study, we compare seven HVG methods from six software packages, including BASiCS, Brennecke, scLVM, scran, scVEGs and Seurat. Our results demonstrate that reproducibility in HVG analysis requires a larger sample size than DEG analysis. Discrepancies between methods and potential issues in these tools are discussed and recommendations are made.

Key words: single-cell RNA seq; scRNA-seq; software; DEG analysis; highly variable gene

Background

Single-cell RNA sequencing (scRNA-seq) technologies allow RNA-seq to be performed on single cells and thus can investigate RNA expression differences on a cell-by-cell basis [1–3]. Hence, scRNA-seq enables statistical analyses that can yield more biological insights than traditional RNA-seq. For example, cell-to-cell variations are often observed within cancerous and embryonic cell samples. However, these variations cannot be detected by bulk RNA-seq. Highly variable gene (HVG) detection is made possible with scRNA-seq data. It allows researchers to detect genes that contribute to cell-to-cell differences in a mixed cell population, such as a group of embryonic stem (ES) cells and cancer cells. Therefore, while existing scRNA-seq analysis packages often focus on various purposes, a function for HVG analysis is often included. BASiCS [4], Brennecke [5], scLVM [6], scran [7], scVEGs [8] and Seurat [9] are commonly used tools that can perform HVG analysis.

HVG detection methods are often composed of two main components: normalization and the analysis of variation. Because batch effects can effectively affect the number of detected HVGs [10], normalization is a crucial step in HVG analysis. Normalization is often accomplished by DESeq's normalization method [5] or the conversion of raw counts into relative expressions. Both methods use scaling factors to normalize expression values among cells. In DESeq's method, it first calculates the geometric mean of all genes' expression. Then, for each cell, it calculates the ratios between each gene's expressions with the geometric means. For each cell, the scaling factor is the median of the ratios. If spike-in references are provided, such as the External RNA Controls Consortium (ERCC) spike-in, DESeq's method can calculate scaling factors by using the spike-ins. For the relative expression normalization method, the scaling factor is the inverse of a cell's total sum of expression values, which can be raw count, reads per kilobase

of gene per million mapped reads or fragments per kilobase of gene per million mapped reads. Count/transcript per million are units that multiply the relative expression by a million. With the relative expression normalization, each cell is normalized independently, which eliminates the risk of erroneous normalization caused by outlying cells or cells with low quality. In contrast, DESeq considers information across all cells and allows a higher level of technical noise reduction.

After normalization, the methods identify genes with high biological variations. Because heteroscedasticity is observed in expression data [11, 12], variance cannot be used as a direct indicator of HVGs. Existing methods use the relationship between variance, or its variations, and the mean as an indicator. Each method then fits their respective mean–variance relationships onto their respective models. Based on the fitted models, they perform statistical tests for high biological variations. Many differences among methods were observed in this step. For example, in scLVM's LogVar algorithm and scran, variance is calculated with a logarithmic transformed expression matrix. Different adaptations of standard deviation (SD), such as variance (in BASiCS [4]), coefficient of variation (CV) (in scVEGs [8]) and squared coefficient of variation (CV2) (in Brennecke [5]), are used by different methods.

We compared seven HVG analysis methods from six different packages, BASiCS, Brennecke, scLVM [6], scran [7], scVEGs and Seurat [9]. Real scRNA-seq data sets from Gierahn, Klein, Deng, Islam and Yan were tested. The reliability of their HVG lists was investigated using clustering. A simulated data set generated by scDD [13] was used to examine their accuracy. Results between and within methods were compared and their similarities and reproducibility were investigated. As heteroscedasticity is an important issue in method development, each result's dependency on the expression mean was also examined. Finally, each method's runtime was investigated; and the best tool is recommended.

Methods

Overview of software packages

In this section, a brief overview of the seven HVG discovery methods is delineated.

BASiCS
BASiCS [4] uses the Bayesian hierarchical model. It first fits the spike-ins to a hierarchical model that can explain biological baseline variance and any technical noise. Subsequently, using all genes in the data set, it extends the hierarchical model to account for biological cell-to-cell variability. Using the fitted hierarchical model, it decomposes variance into biological background, technical variability and biological variability. The biological variability component is used to test for HVGs. One of the advantages of BASiCS' method is that lowly variable genes can also be discovered.

Brennecke
Brennecke [5] uses DESeq's method for normalization. This method uses the CV2 on normalized count data to analyze biological variation. Technical noise in the data set is modeled by fitting a generalized linear model onto the mean versus the CV2 plot. Using this model, a coefficient of biological variation is obtained. Testing for high variance is performed on the coefficient of biological variation using the chi-square distribution. The reasoning being that extraordinarily high variance indicates high uncertainty of variation in a gene when the expression

mean is low. Therefore, one advantage of their method is that genes with high uncertainty can be filtered using a CV2 threshold, which allows better control of false-positive rates.

scVEGs
scVEGs [8] normalizes expression data by using the relative expression method, which is subsequently multiplied by the mean of all the sample's total counts. The authors propose using the CV to model variations in genes. They model the mean versus CV relationship in the data set by assuming a negative binomial distribution. The relationship between mean and CV is fitted with a modified local regression. After obtaining the parameters from the fitted model, *P*-values are obtained by assuming the normal distribution in the difference between each gene and the model curve.

Seurat
Seurat [9] performs normalization with the relative expression multiplied by 10 000. It uses variance divided by mean (VDM). It assigns the VDMs into 20 bins based on their expression means. Then, within each bin, Seurat normalizes the VDMs into *z*-scores. As VDMs are normalized among genes that share similar expression levels, this strategy has an advantage in decreasing its results' dependency on the mean. Finally, it uses a threshold to identify HVGs.

scLVM
scLVM [6] normalizes the data set using DESeq's method. It detects HVG analysis with two different approaches. It either uses log-mean to log-CV2 relationship with the log-linear fit (scLVM_Log) or uses locally weighted scatterplot smoothing (LOESS) with the mean–variance relationship after logarithmic transformation (scLVM_LogVar). Both of these approaches will be analyzed discreetly in this evaluation as individual methods.

scran
scran [7, 14] has a specialized scaling factor calculation algorithm for normalization [14] similar to DESeq. However, in scran's calculation, multiple scaling factors are calculated for multiple pools of cells, which are subsets of the data set. Thereby, a more accurate estimate of the scaling factor for each cell is obtained through a system of linear equations, which are calculated from each pool's scaling factors. This strategy strives to obtain a more accurate estimate of the scaling factor for each cell. To perform HVG, it uses LOESS with the mean–variance relationship of log-transformed expression values. After obtaining a LOESS fit as the model, it estimates technical variance in each gene. Finally, it subtracts from each variance its inferred technical component to obtain biological variation.

Summary
In summary, Brennecke, scVEGs and Seurat divided SD or variance by the mean. Then, the tools fit their respective models onto the mean versus CV/CV2/VDM plots to eliminate the remaining effects of the mean on variance. Alternatively, scLVM and scran performed log-transformation and used regression models on the mean–variance plots to eliminate the effects of the mean on variance. Unlike the other methods, BASiCS uses a Bayesian hierarchical model to obtain biological variations from the data.

Data sets and their processing

The Gierahn data set was downloaded from Gene Expression Omnibus (GEO) under accession number GSE92495 [15], where the

tuberculosis-exposed human peripheral blood mononuclear cell (PBMC) sample was used. It contains 4296 cells, and the mean total count is 1861. The Klein data set was downloaded from GSE65525 [16], which included a 0-day mouse ES cell sample (File: GSM1599494_ES_d0_main.csv with 933 samples) and three samples following leukemia inhibitory factor (LIF) withdrawal for 2, 4 and 7 days (Files: GSM1599497_ES_d2_LIFminus.csv with 303 samples, GSM1599498_ES_d4_LIFminus.csv with 683 samples, GSM1599499_ES_d7_LIFminus.csv with 798 samples). This data set contains 2717 cells, and the mean total count is 20 033.

The Deng, Yan and Islam data sets were downloaded from GEO under accession number GSE45719[17], GSE36552 [18] and GSE29087 [1], respectively. Cell types with >10 cells were used. In the Deng data set, cell types labeled 16cell, 4cell, 8cell, C57twocell, early2cell, earlyblast, late2cell, lateblast, mid2cell and midblast were downloaded. In the Yan data set, cell types labeled 4-cell, 8-cell, Morulae, Late blastocyst and hESC were downloaded. The Islam data set contains 48 ES cells and 44 are embryonic fibroblasts. Their average total counts are 15.5 million, 25.2 million and 0.58 million, and their total number of cells are 247, 110 and 92, respectively.

The Deng and Yan data sets's fastq files were first trimmed using Trimmomatics [19] with default parameters and were quantified by Kallisto [20] using Ensembl [21] mm10 and hg38 assemblies, respectively, also with default parameters. The Deng and Yan data sets are transcript expression data. The gene raw counts of the other data sets were downloaded directly.

The simulated data set was obtained using the scDD software.[13] The simulateSet function was used according to the user manual, nDE, nDP, nDM and nDB were set to 100, nEE and nEP were set to 200 and numSamples was set to 200. Hence, there was a total 400 samples, where 400 genes were differentially expressed and 400 were nondifferentially expressed.

Technical details

BASiCS, Brennecke, scLVM_Log, scLVM_LogVar, scran, scVEGs and Seurat were run with default parameters as described in their user manuals. In BASiCS' BASiCS_MCMC function, N, Thin and Burn were set to 1000, 10 and 500, respectively.

To obtain *P*-values from Seurat, we converted the *z*-scores from its outputs into *P*-values by using the Z table. The *P*-value threshold of 0.05 is used by Brennecke, scran and Seurat in the results. scVEGs had an error with the Gierahn data set and detected no HVGs with the Yan and Deng data sets. Seurat had an error with the simulated data set. BASiCS requires spike-in genes, and it was not used to analyze the Gierahn and Klein data sets. Hence, these results were not shown in the figures.

Rediscovery rate

To investigated the rediscovery rate [22] of each method, two independent experiments were simulated by randomly choosing cells from the Klein and Islam data sets without replacements and putting them into two groups, Group1 and Group2; then, the percentage of HVGs from Group1 that were rediscovered in Group2 was calculated. This analysis was repeated 30 times for the Klein data set and 100 times for Islam data set.

Clustering

Each of the data sets were normalized into the relative expression and multiplied by the median total count. Log plus one transformation was used. Each method's detected HVGs were extracted from the normalized data sets and t-distributed
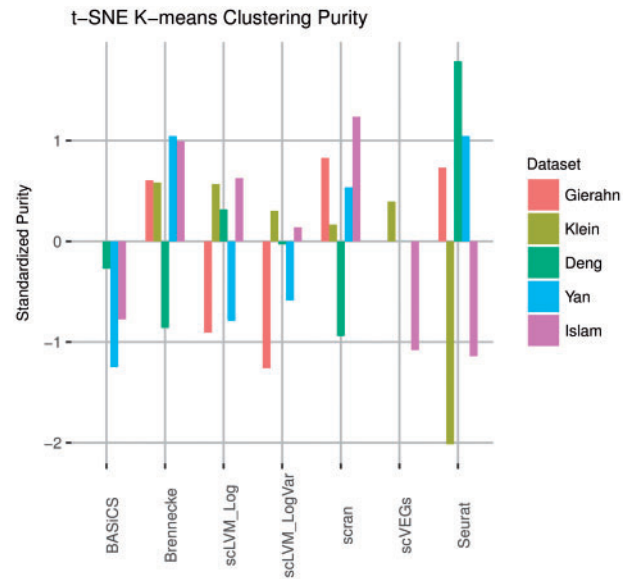
Figure 1. Standardized clustering purity of t-SNE *k*-means clustering results by using HVGs only. Table 1 contains the raw purity results. Standardized purity >0 indicates better than average (mean) performance. Purities from each data set are standardized separately. BASiCS was not used to analyze the Gierahn and Klein data sets because of the lack of spike-ins. scVEGs had an error with the Gierahn data set and called zero HVGs with the Yan and Deng data sets.

stochastic neighbor embedding (t-SNE) was performed. The R library Rtsne was used to perform t-SNE. Principal components (PCs) 2 to 6 were used for a more accurate assessment, because, except scLVM_Log, all tools showed better performances at higher dimensions. *K*-means clustering was performed, and purity was calculated using the known cell-type information. Therefore, given one method and one data set, five purities would be obtained because 2 to 6 PCs were used. The average purity (from PC 2 to 6) is reported.

Results and discussion

Accuracy of the methods

t-SNE k-means clustering
In a group of cells that contain multiple cell types, HVGs contain the variability among cells. Therefore, different cells can be clustered into their specific cell types by using the HVGs. Using significant HVGs from each method, t-SNE *k*-means clustering is performed on Gierahn, Klein, Deng and Yan data sets. Purity is used to assess the accuracy of the list of HVGs of each method. Table 1 contains the purity of each method across all four data sets. In Figure 1, purities from each data set are normalized into standardized scores to allow comparison across data sets. Standardized purity >0 indicates better than average performance. None of the tools performed better than average in all five data sets. Brennecke performed better than average in four of the data sets. Seurat performed notably well with the Gierahn, Deng and Yan data sets, and scran ranked top with Gierahn and Islam data sets. scLVM and scVEGs showed average performances. BASiCS performed lower than average in all three data sets that it analyzed.

Accuracy, precision and sensitivity using simulated data set
Using the scDD R library, an scRNA-seq data set with 800 genes and 400 cells was generated, where 400 genes were known to be

Table 1. Purity of *t*-SNE *k*-means clustering results by using HVGs only

| Method	Dataset				
	Gierahn	Klein	Deng	Yan	Islam
BASiCS	NA	NA	0.66	0.88	0.92
Brennecke	0.86	0.99	0.64	0.96	0.98
scLVM_Log	0.74	0.99	0.67	0.87	0.97
scLVM_LogVar	0.71	0.96	0.66	0.91	0.95
scran	0.88	0.95	0.64	0.95	0.99
scVEGs	NA	0.97	NA	NA	0.90
Seurat	0.88	0.75	0.72	0.96	0.90
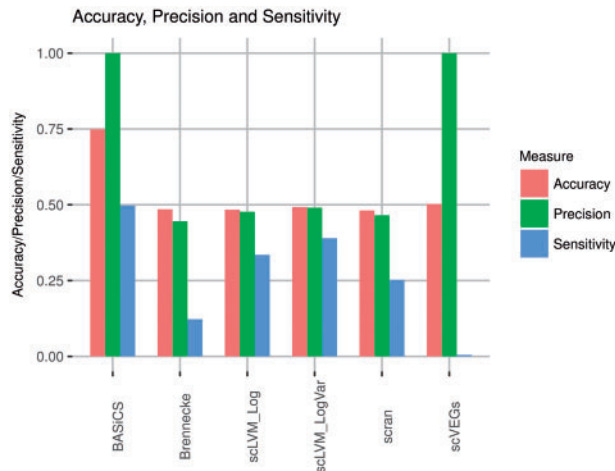
Figure 2. HVG analysis with the simulated data set. Accuracy, precision and sensitivity of each method.

highly variable. Accuracy, precision and sensitivity of each method are shown in Figure 2 (Supplementary Table S1). In contrary to Figure 1, BASiCS showed perfect precision and had the best performance in all three aspects. scVEGs showed perfect precision but low sensitivity. Brennecke, scLVM and scran showed similar precision, but scLVM had better sensitivity.

Number of detected HVGs

Figure 3 shows the number of detected HVGs of each method (Supplementary Table S2). The Gierahn, Klein and Islam data sets have 6713, 24 047 and 14 909 genes totally. The Deng and Yan data sets have 91 518 and 157 954 genes totally. The differences between tools are notable. The tool that detected the most HVGs can have 15–139 times more genes than the tool that detected the least HVGs. Relative to the other methods, Brennecke detected more HVGs with the Gierahn and Klein data sets, but less HVGs with the other data sets. In contrast, scLVM detected less HVGs with the Gierahn and Klein data sets, but more HVGs with the other data sets. Overall, scran and Seurat showed stable performance compared with the other methods.

Figure 4 shows the number of the detected HVGs with respect to the number of cells. With the Klein data set, a larger sample size would increase the number of detected HVGs with BASiCS, scran and scVEGs. With the Islam data set, BASiCS, Brennecke, scran and Seurat's numbers of detected HVGs would increase with larger sample sizes. Overall, more HVGs were detected in the Klein data set because it had more genes and more cells than the Islam data set.
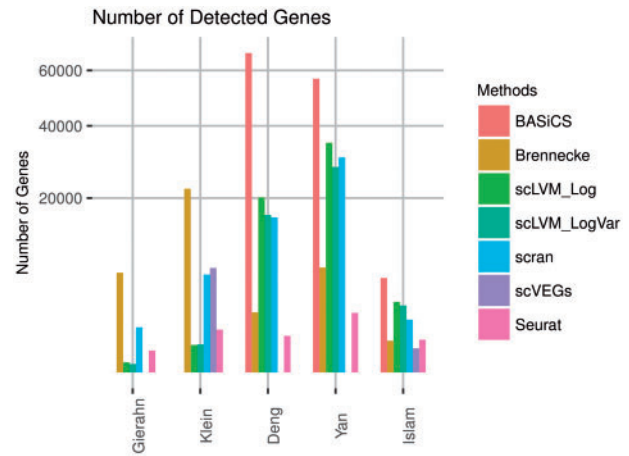
Figure 3. Number of detected HVGs. BASiCS was not used to analyze the Gierahn and Klein data sets because of the lack of spike-ins. scVEGs had an error with the Gierahn data set and called zero HVGs with the Yan and Deng data sets.

Overlap within methods

In Figure 5, we examine the reproducibility of the results of each method using the rediscovery rate. [22] Supplementary Figure S1 shows the result of the same analysis using the Deng data set's mid stage blastocysts. All methods show higher rediscovery rates with higher sample sizes. This suggests that a larger sample size can improve reproducibility in these methods. Because the Klein data set has more cells than the Islam data set, it results in a higher reproducibility in each method.

Overlap between methods

Figure 6A shows the Venn diagram of the most significant 900 HVGs called by Brennecke, scran, scVEGs and Seurat with the Islam data set. Figure 6B shows the Venn diagram of the detected HVGs called by all tools with the Islam data set. Overall, the HVG analysis methods do not show good overlaps with each other. Few HVGs are concurrently detected as significant by all methods. Overlap analyses with the Deng and Yan data sets also showed similar results (Supplementary Tables S3 and S4).

In summary, if the same data set is analyzed by two different tools, the resulting list of HVGs can be different. This can lead to conflicting conclusions if downstream analyses are performed using these differing lists of results.

Dependence with the mean

In scRNA-seq count data, expression mean and variance are positively correlated, and all of the tested methods have developed strategies to tackle this heteroscedasticity issue. In Table 2, the percentage overlap between each method's result and the same number of top highly expressing genes is calculated. For example, if a method detects 500 genes as highly variable, the amount of overlap between its list of HVGs and the 500 highest expressing genes in the data set would be obtained. In this test, some overlaps should be expected because some highly expressing genes can also be HVGs.

Brennecke and scran show similar results (Table 2), and their average percent overlaps are 21.5 and 25.99%, respectively. Seurat is more conservative in declaring a highly expressed gene as significant, and its average percent overlap with the
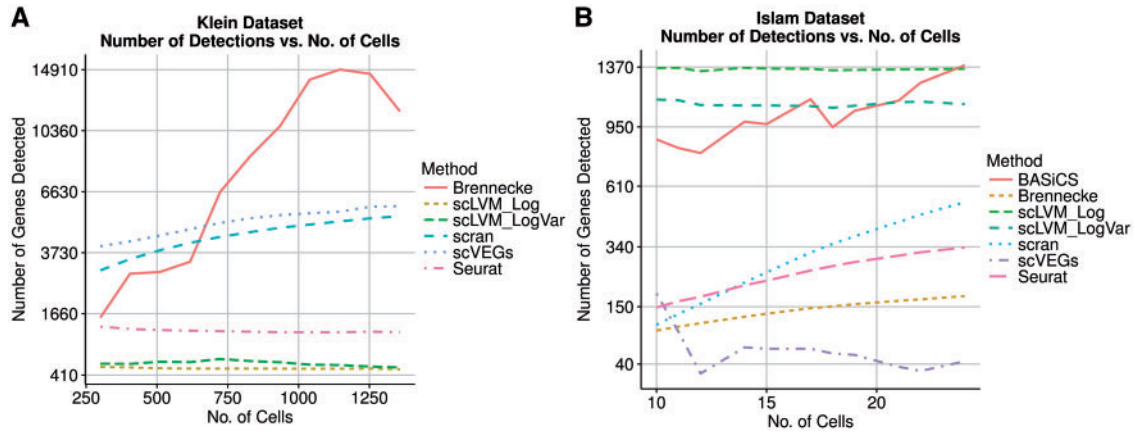
Figure 4. Number of HVGs versus sample size. (**A**) Klein data set. (**B**) Islam data set.
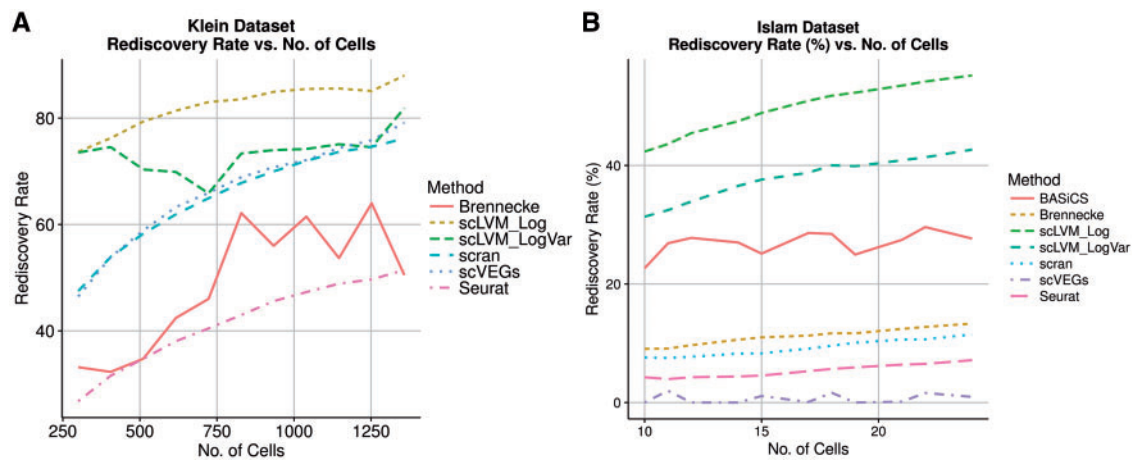

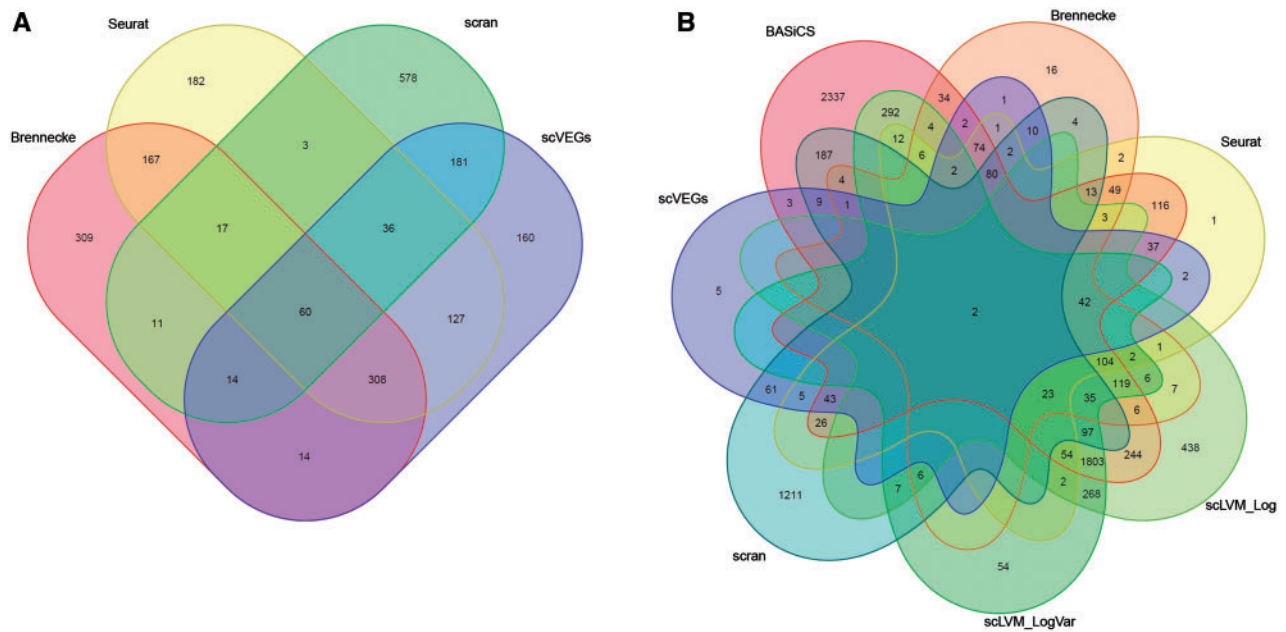**Figure 5.** Rediscovery rate versus sample size. (**A**) Klein data set. (**B**) Islam data set.

Figure 6. Venn diagrams of each tool's result with the Islam data set. (**A**) The overlap between the most significant 900 genes from Brennecke, scran, scVEGs and Seurat. BASiCS and scLVM do not output ranks, and they are skipped. (**B**) The overlap between all tools, using the P-value threshold of 0.05.

Table 2. Percent overlap between a method's detected genes with the same number of highest expressing genes

Method	Dataset			Average (%)
	Deng (%)	Islam (%)	Yan (%)	
Brennecke	23.22	25.49	15.78	21.50
scLVM_Log	82.84	63.68	87.75	78.09
scLVM_LogVar	68.28	50.09	63.44	60.60
scran	22.65	29.29	26.04	25.99
scVEGs	NA	1.10	NA	1.10
Seurat	10.03	22.17	5.27	12.49

Note: Results from the Deng, Islam and Yan data sets are reported. BASiCS is not tested because of its low stability.

highest expressing genes is 12.49%. The two scLVM's results have higher dependencies on the mean than the other methods; consequently, they have percentage overlaps that range from 50.09 to 87.75%. Overall, Brennecke, scran and Seurat show consistent levels of low to medium amounts of overlap with the highest expressing genes. This suggests that their dependencies with the mean are well eliminated.

Runtime

Figure 7 illustrates the differing running time of each tool. In the Deng data set, from shortest to longest time, scLVM_Log, Brennecke, scran, Seurat, scVEGs, scLVM_LogVar and BASiCS took 2.9, 6.7, 11, 16, 82, 285 and 1680 s, respectively. scLVM_Log, Brennecke, scran and Seurat only needed seconds, but scLVM_LogVar and BASiCS needed minutes to half an hour. This analysis is performed on a computing workstation with 32 Intel(R) Xeon(R) CPU E5-2650 v2 processors and 128 GB of RAM.

Conclusion

This study reports a detailed comparison of seven HVG methods for scRNA-seq data. Five real scRNA-seq data sets, Gierahn, Klein, Islam, Deng and Yan, were used. Overall, large differences were observed among the methods, and each tool performed optimally in different situations.

Clustering analysis was used to access each tool's accuracy with real scRNA-seq data sets. Brennecke was shown to have stable good performances in a wide range of data sets. scran and Seurat were shown to perform optimally with some of the data sets. Simulated scRNA-seq data sets were used to further access the performance of each method. Interestingly, BASiCS had the best performance with the simulated data set despite its lower performance with the real data sets. This suggests that the utilization of spike-ins has an advantage when the spike-ins are true stable genes, as shown by the simulated data set. However, noises in real scRNA-seq can cause lower performances in methods that use spike-ins [23].

Type I/II error rates can be inferred from the number of detected genes. By definition, because P-value is equivalent to the proportion of genes when no HVG exists, scVEGs and Seurat's low number of detections signal potential inflated type II error rates. In contrast, BASiCS and Brennecke would call much more HVGs than the other methods in some data sets. This signals potential inflated type I error rates if the other methods are assumed to have good type I/II error rates.

Heteroscedasticity is an important issue in HVG detection algorithms. We showed that both scLVMs' performances are
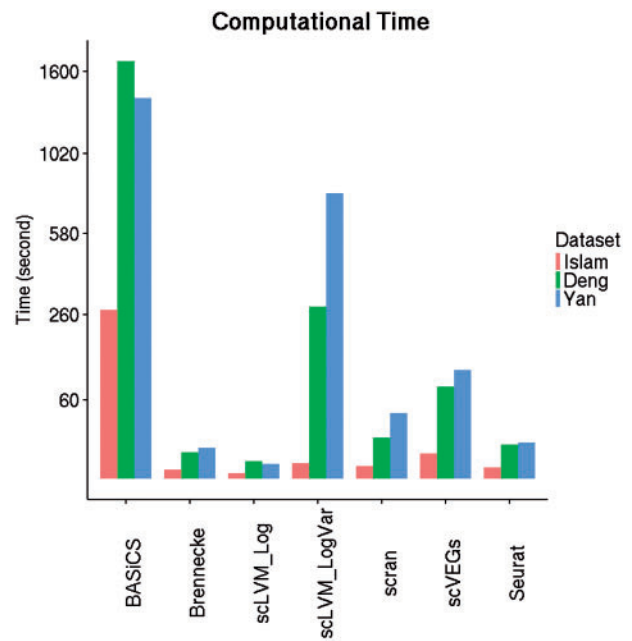
Figure 7. Runtime of each method with the Islam, Deng and Yan data sets. This analysis is performed on a computing workstation with 32 Intel(R) Xeon(R) CPU E5-2650 v2 processors and 128 GB of RAM.

similar to selecting the highest expressing genes as 'highly variable', which raises a concern. On a side note, this test showed that expression mean is more reproducible than expression variance (Figure 5 and Table 2). This suggests that statistical analyses that depend on the expression mean, such as differentially expressed gene analysis [11, 24–29], would show higher reproducibility than HVG analysis.

In conclusion, high discrepancies between methods were discovered. Overall, all methods, except scran, raised issues of concern in the results. In comparison with the other methods, scran has a good performance in clustering, stable number of detected HVGs, good independency from the mean and good running time. Nevertheless, some considerations should be taken to ensure reproducibility of independent experiments. First, downstream data analysis packages, such as SC3 [30], can be used to improve clustering analysis results. Second, HVG analysis requires large sample sizes to be reproducible. Many methods show increasing rediscovery rates as the sample size increases. Third, the rediscovery rates results suggest that higher sequencing depth can also improve reproducibility (Figure 5 and Supplementary Figure S1). Finally, utilization of overlaps between more tools is suggested. The stringency of assigning P-value between each tool is different, which can result in different numbers of detected genes when different tools are used.

Supplementary Data

Supplementary data are available online at https://academic.oup.com/bib.

Funding

Key Points

- BASiCS' results ranked top with simulated data sets but ranked last in real data set. This shows that noises in real scRNA-seq can cause lower performances in methods that use spike-ins.
- Reproducibilities are low, among different tools and among different samples analyzed by the same tool. A higher number of cells can improve rediscovery rates.
- BASiCS, Brennecke, scVEGs and Seurat are shown to have high type I/II error rates.
- The two scLVM algorithms can have high amount of overlaps with the highest expressing genes, casting concerns in their results.
- BASiCS and scLVM_LogVar are much slower than the other methods.

References

1. Islam S, Kjallquist U, Moliner A, *et al*. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* 2011;**21**(7):1160–7.
2. Tang F, Barbacioru C, Wang Y, *et al*. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009;**6**(5): 377–82.
3. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;**10**(1):57–63.
4. Vallejos CA, Marioni JC, Richardson S. BASiCS: Bayesian analysis of single-cell sequencing data. *PLoS Comput Biol* 2015; **11**(6):e1004333.
5. Brennecke P, Anders S, Kim JK, *et al*. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* 2013; **10**(11):1093–5.
6. Buettner F, Natarajan KN, Casale FP, *et al*. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* 2015;**33**(2):155–60.
7. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* 2016;**5**:2122.
8. Chen H-IH, Jin Y, Huang Y, *et al*. Detection of high variability in gene expression from single-cell RNA-seq profiling. *BMC Genomics* 2016;**17**(S7):508.
9. Satija R, Farrell JA, Gennert D, *et al*. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;**33**(5):495–502.
10. Finak G, McDavid A, Yajima M, *et al*. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 2015;**16**(1):278.
11. Yip SH, Wang P, Kocher J-PA, *et al*. Linnorm: improved statistical analysis for single cell RNA-seq expression data. *Nucleic Acids Res* 2017;**45**(22):e179.
12. Law CW, Chen Y, Shi W, *et al*. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;**15**:R29.
13. Korthauer KD, Chu LF, Newton MA, *et al*. A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol* 2016;**17**(1):222.
14. Lun AT, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* 2016;**17**:75.
15. Gierahn TM, Wadsworth MH, II, Hughes TK, *et al*. Seq-well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* 2017;**14**(4):395–8.
16. Klein AM, Mazutis L, Akartuna I, *et al*. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 2015;**161**(5):1187–201.
17. Deng Q, Ramskold D, Reinius B, *et al*. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 2014;**343**(6167):193–6.
18. Yan L, Yang M, Guo H, *et al*. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* 2013;**20**(9):1131–9.
19. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**(15): 2114–20.
20. Bray NL, Pimentel H, Melsted P, *et al*. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016;**34**(8): 888–527.
21. Yates A, Akanni W, Amode MR, *et al*. Ensembl 2016. *Nucleic Acids Res* 2016;**44**(D1):D710–6.
22. Ganna A, Lee D, Ingelsson E, *et al*. Rediscovery rate estimation for assessing the validation of significant findings in high-throughput studies. *Brief Bioinform* 2015;**16**(4):563–75.
23. McDavid A, Finak G, Gottardo R. The contribution of cell cycle to heterogeneity in single-cell RNA-seq data. *Nat Biotechnol* 2016;**34**(6):591–3.
24. Leng N, Li Y, McIntosh BE, *et al*. EBSeq-HMM: a Bayesian approach for identifying gene-expression changes in ordered RNA-seq experiments. *Bioinformatics* 2015;**31**(16): 2614–22.
25. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Res* 2014;**42**(11):e91.
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**(12):550.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**(1): 139–40.
28. Hardcastle TJ, Kelly KA. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 2010;**11**(1):422.
29. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;**11**(10):R106.
30. Kiselev VY, Kirschner K, Schaub MT, *et al*. SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* 2017;**14**(5): 483–6.