Evaluating Model-free Directional Dependency Methods on Single-cell RNA Sequencing Data with Severe Dropout

Eliška Dvořáková*
Department of Computer Science
Czech Technical University
Prague, Czech Republic
dvorael1@fel.cvut.cz

Sajal Kumar*
Department of Computer Science
New Mexico State University
Las Cruces, USA
sajal49@nmsu.edu

Jiří Kléma
Department of Computer Science
Czech Technical University
Prague, Czech Republic
klema@fel.cvut.cz

Filip Železný
Department of Computer Science
Czech Technical University
Prague, Czech Republic
zelezny@fel.cvut.cz

Karel Drbal
Department of Cell Biology
Charles University
Prague, Czech Republic
karel.drbal@natur.cuni.cz

Mingzhou Song†
Department of Computer Science
New Mexico State University
Las Cruces, USA
joemsong@cs.nmsu.edu

ABSTRACT

As severe dropout in single-cell RNA sequencing (scRNA-seq) degrades data quality, current methods for network inference face increased uncertainty from such data. To examine how dropout influences directional dependency inference from scRNA-seq data, we thus studied four methods based on discrete data that are model-free without parametric model assumptions. They include two established methods: conditional entropy and Kruskal-Wallis test, and two recent methods: causal inference by stochastic complexity and function index. We also included three non-directional methods for a contrast. On simulated data, function index performed most favorably at varying dropout rates, sample sizes, and discrete levels. On an scRNA-seq dataset from developing mouse cerebella, function index and Kruskal-Wallis test performed favorably over other methods in detecting expression of developmental genes as a function of time. Overall among the four methods, function index is most resistant to dropout for both directional and dependency inference. The next best choice, Kruskal-Wallis test, carries a directional bias towards a uniformly distributed variable. We conclude that a method robust to marginal distributions with a sufficiently large sample size can reap benefits of single-cell over bulk RNA sequencing in understanding molecular mechanisms at the cellular resolution.

CCS CONCEPTS

• **Mathematics of computing** → *Nonparametric statistics*; • **Applied computing** → **Biological networks**.

*Both authors contributed equally to the work.
†Corresponding author.

ICBRA '19, December 19–21, 2019, Seoul, Republic of Korea

ACM ISBN 978-1-4503-7218-3/19/12…$15.00
https://doi.org/10.1145/3383783.3383793

KEYWORDS

model-free, directional dependency, single-cell sequencing

1 INTRODUCTION

Measuring RNA abundance in individual cells, single-cell RNA sequencing (scRNA-seq) provides an unprecedented resolution to study molecular mechanisms underlying different cell types within a tissue. This enables biology inquires impossible with bulk RNA sequencing, where one measurement of a transcript represents the total of that transcript over all cells in a sample. One such inquiry is cell-type-specific gene network inference. This opportunity is however hampered by a great deal of uncertainty in the readout of transcript abundance in scRNA-seq. Dropout, also known as observed zeros, is a limitation of current scRNA library preparation techniques [10]. It refers to RNA molecules in a cell not being captured for reverse transcription—a necessary step during library preparation. It leads to a multitude of zero transcript levels for a gene in 30–90% cells [2]. Several popular network inference software packages developed without considerations of dropout amounted to nearly random guessing on scRNA-seq data [7]. Network inference after data imputation remains unexplored for a risk of false positives due to a circular argument [1].

Necessary to many network inference applications are both directional and dependency inference between two random variables X and Y, which can represent levels of gene expression or covariates in scRNA-seq experiments. In directional inference, we test Y being a function of X versus X being a function of Y; in dependency inference, we test Y being a function of X versus X and Y being statistically independent. Existing methods that simultaneously address both types of inference are mostly parametric regression analysis [9]. With the throughput of one scRNA-seq experiment approaching millions of cells, it is feasible to study complex nonlinear patterns model-free without predefined mathematical models. Thus, we focus on model-free statistics based on discrete data here. The established conditional entropy and the recent causal inference by stochastic complexity (CISC) [4] are popular choices for directional inference applicable to discrete data. Kruskal-Wallis test [11], based on the discrepancy between group rank means, can do both directional and dependency inference. However, it is not optimized

to capture functional dependency ($f: X \rightarrow Y$), favoring monotonic over non-monotonic functions.

A recently introduced measure called function index claims functional optimality—it is maximized if and only if Y is a non-constant function of X [12, 18, 35]. Measuring the effect size of the functional chi-squared statistic (FunChisq) [18, 33, 36], function index ranges from 0 to 1 and has been used for analyzing bulk transcriptome sequencing data [12, 35]. Similar to other dependency measures, function index is minimized if and only if X and Y are empirically independent.

We evaluated the performance of these methods on both simulated data and real scRNA-seq data in contrast to Pearson's chi-squared test, mutual information, and Pearson's correlation. The performance is measured by percentage of incorrect directions and areas under the receiver-operating characteristic (ROC) and precision-recall (PR) curves. New from previous scRNA-seq dropout studies, we delved into how dropout rates, sample sizes, and discrete levels can influence the performance. We revealed that function index performed robustly relatively to other methods in both directional and dependency inference. We also observed that increasing the sample size can remediate dropout to a demonstrated extent. Our findings suggest that with a sufficient sample size one can make an informed choice of method for directional dependency inference to overcome dropout in scRNA sequencing to a definite advantage.

2 METHODS

2.1 Function index

As function index was just recently introduced [12, 18, 35], we provide its formulation and a justification on its application to scRNA-seq data. Let X and Y be two discrete random variables of r and s levels, respectively. Given n pairs of (X, Y) observations, the FunChisq test is defined on the $r \times s$ contingency table formed by X being the row variable and Y being the column variable. The value at row i and column j of the table, n_{ij}, denotes the number of pairs with $X = i$ and $Y = j$. We use $n_{i\cdot}$ for the sum of row i and $n_{\cdot j}$ for the sum of column j. The FunChisq test statistic is defined by

$$\chi^2_f(X \rightarrow Y) = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{[n_{ij} - (n_{i\cdot}/s)]^2}{n_{i\cdot}/s} - \sum_{j=1}^{s} \frac{[n_{\cdot j} - (n/s)]^2}{n/s} \quad (1)$$

Mathematical and statistical properties of FunChisq are derived in [18, 33, 36]. Function index is designed to measure the effect size of functional dependency independent of sample size. It was first introduced and applied to bulk transcriptome data analysis in [12, 18, 35]. The function index ξ_f is defined by the square root of the ratio of FunChisq test statistic to its maximum attainable value:

$$\xi_f = \sqrt{\frac{\chi^2_f(X \rightarrow Y)}{n(s-1) - \sum_{j=1}^{s}[n_{\cdot j} - (n/s)]^2/(n/s)}} \quad (2)$$

The numerator and denominator within the square root scale with n. Thus function index is independent of sample size, suitable for measuring the effect size. It is proven [18] that $\xi_f \in [0, 1]$ is minimized if and only if X and Y are empirically independent and maximized if and only if Y is a non-constant function of X. We thus call function index asymmetrically functionally optimal. A larger ξ_f indicates a stronger functional dependency of Y on X.

On data with a large sample size, P-value is not necessarily effective: a test statistic tends to have a small P-value but with a weak effect. Thus a quantitative measure of effect size is important [27]. As single-cell data are considered "big data", where effect size statistics are more relevant [3], exploring function index usage on scRNA-seq data is justified.

2.2 Implementations of all seven methods

The R package `FunChisq` [34] implements function index within the function `fun.chisq.test()`. Function index is returned in the function output. Details including examples on the function usage are provided in package documentation. Conditional entropy and mutual information [8] are implemented as functions `condentropy()` and `mutinformation()` in the R package `infotheo` [16]. Pearson's chi-squared test [20], Kruskal-Wallis test [11] and Pearson correlation test [26] are available via R functions `chisq.test()`, `kruskal.test()` and `cor.test()` in the package `stats` [21]. CISC was ported to R from the provided Python implementation [4]. Pearson's correlation is included as a baseline for comparison.

3 RESULTS

To understand how the dropout rate, the sample size, and the discrete level can influence directional dependency inference, we present the performance of the seven methods on simulated and real scRNA-seq datasets.

3.1 Performance on simulation studies

Generating ground truth and noise-free tables. To evaluate the performance of both directional and dependency inference, we simulated 50,000 contingency tables at five dropout rates, five table sizes, and five sample sizes. For directional inference evaluation, we generated (1) 100 many-to-one functional patterns from X to Y to represent *true* directional patterns and (2) the transpose of the 100 many-to-one patterns, which we also call one-to-many non-functional patterns, to represent *false* directional patterns. For dependency inference evaluation, we generated (3) 100 functional patterns as *true* dependent patterns and (4) another 100 patterns where X and Y were independent as *false* dependent patterns. The four table types as illustrated in Figure 1 are generated as contingency tables by the R function `simulate_tables()` [24] at varying sample sizes and table sizes. In each contingency table, X and Y are always the row and column variables, respectively. Functional tables have a uniform marginal for X; one-to-many tables have a uniform marginal for Y; independent tables have uniform marginals for both X and Y.

Mimicking noise and simulating dropout. To simulate biological variations or technical noise not due to dropout, we applied house noise at level 0.2 to all tables. This noise level estimated from bulk transcriptome data [32] represents a possible scenario for real data. The first row of Figure 1 shows four table types with noise but no dropout.

We randomly chose samples at a given percentage, or dropout rate, in each variable to reassign them to the first level that is analogous to zero expression level. Figure 1 illustrates how an

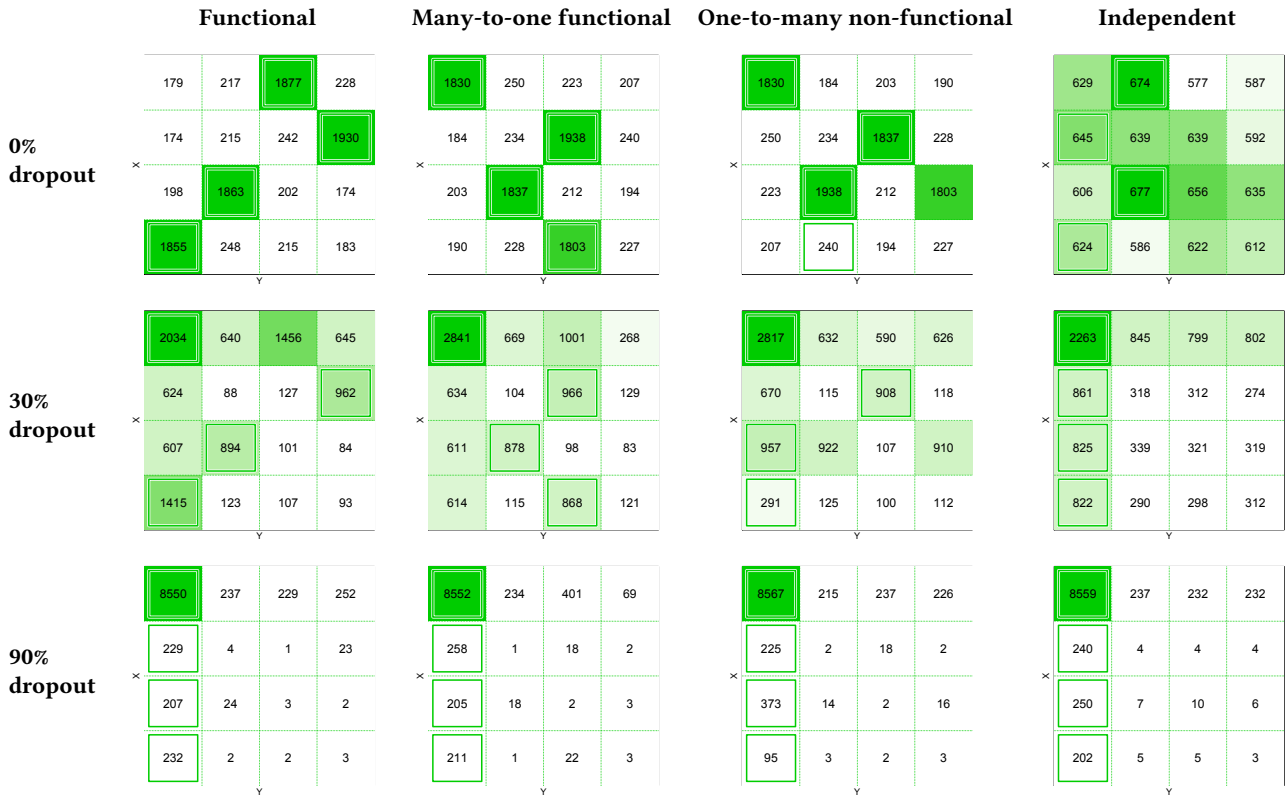

Figure 1: Four table types with noise and dropout. Each table has 10,000 samples at a house noise level of 0.2. Each column illustrates a table type. Dropout rates are 0, 0.3 and 0.9 for each row. The maximum value in each table row is marked by an extra box to indicate the best functional patterns from X (row) to Y (column). A cell in a darker shade has more samples in it. Many-to-one versus one-to-many tables are used in directional inference. Functional versus independent tables are used in dependence inference.

increasing dropout rate gradually erases the signal in the tables within each column.

Performance of directional inference. We first compared the four asymmetric measures on inferring direction. Figure 2 shows percentages of incorrect directions for all four methods over 100 pairs of a many-to-one functional table and its transpose. Even though dropout rates are highly variable, a typical scRNA-seq experiment can have about 50% dropout on average [2]. At a fixed dropout rate of 50% and a fixed table size of 5×5, more importantly,function index has a significant drop in error ratefrom 36% to 15% when the sample size increased from 100 to 100K, while other methods have insignificant or no drop in the error rate, not being able to benefit from larger sample sizes (Figure 2a). All methods, except CISC, reduce error rates over decreasing dropout rates at a fixed sample size of 10,000 and a fixed table size of 5×5 (Figure 2b) or increasing discrete levels at a fixed dropout rate of 50% and a fixed sample size of 10,000 (Figure 2c). Surprisingly, CISC makes no correct prediction at 0% dropout and its accuracy is always close to a random guess. Overall, function index makes the fewest number of mistakes.

Next, we evaluated how well a method ranks directions among multiple tables including their transpose at varying sample sizes, dropout rates and discrete levels. The metrics are area under the ROC curve (AUROC) and area under the PR curve (AUPR). As expected, the three symmetric methods are ineffective (Figure 3). As seen in Figure 2, CISC is unable to determine directionality. Kruskal-Wallis test is always third in terms of both AUROC and AUPR.

Function index benefited the most from increased sample sizes in AUROC at the dropout rate of 50% and the table size of 5×5 (Figure 3a,b). Although second to conditional entropy in AUPR at small sample sizes, it catches up at large sample sizes.

All performance worsens over increasing dropout rates at sample size 10,000 and table size 5×5 (Figure 3c,d). Function index is notably better than other methods in AUROC; however second to conditional entropy in AUPR.

At increasing table sizes with the sample size at 10,000 and the dropout rate at 50% (Figure 3e,f), both conditional entropy and function index show an overall increasing trend. Function index is notably better than all other methods in AUROC, while both function index and conditional entropy perform similarly in AUPR.

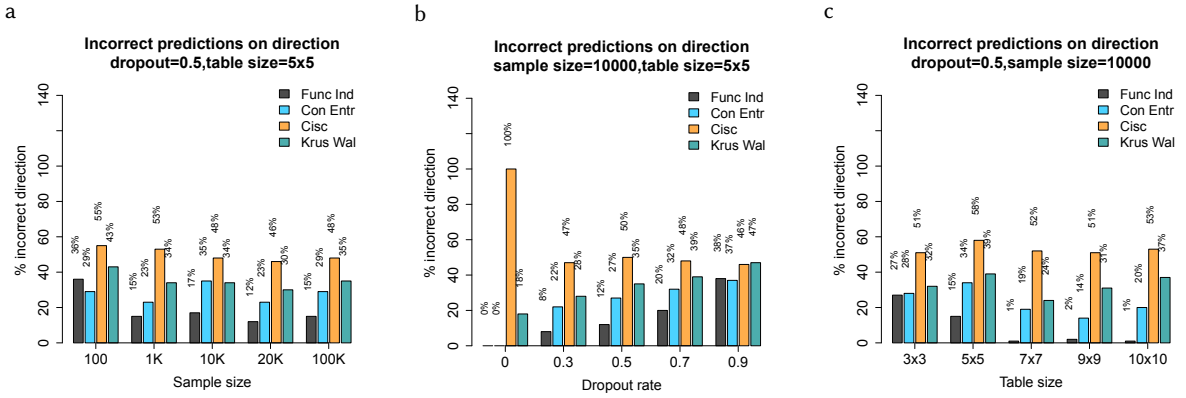

Figure 2: Percentage of incorrectly inferred directions on simulated data at varying dropout rates, sample sizes and table sizes. Function index exhibits a definite advantage over other directional measures with increasing (a) sample sizes, (b) dropout rates, and (c) table sizes, except at the smallest sample size or the highest dropout rate.

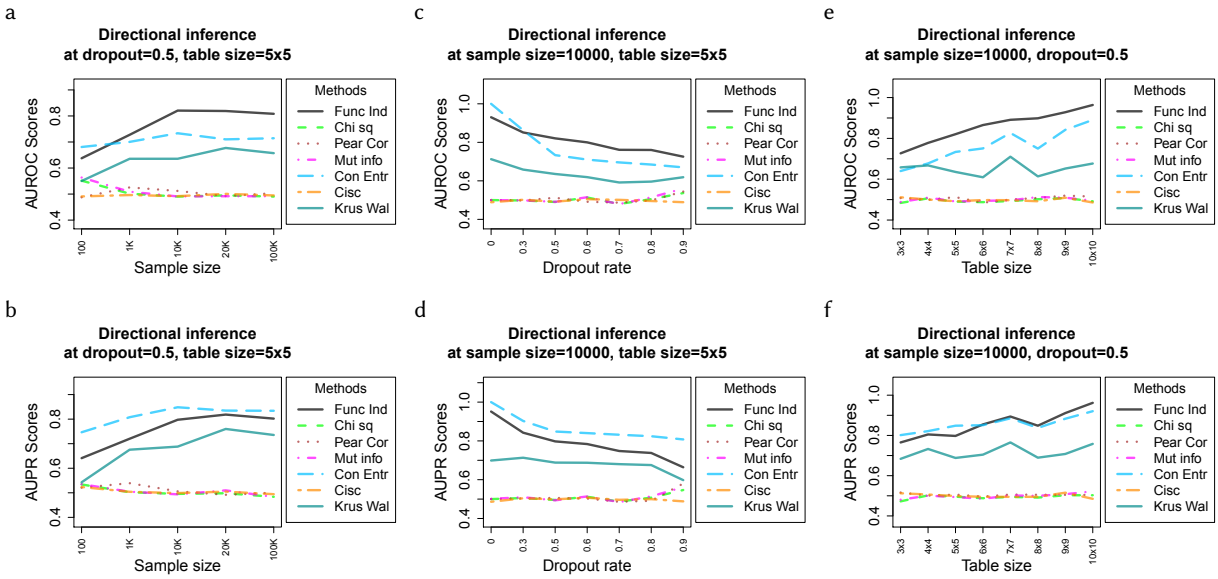

Figure 3: Directional inference performance on simulated data at varying dropout rates, sample sizes and table sizes. The effect of sample size: (a) AUROCs and (b) AUPRs. The effect of dropout rate: (c) AUROCs and (d) AUPRs. The effect of discrete level: (e) AUROCs and (f) AUPRs.

Performance of dependency inference. Figure 4 presents the performance of the five methods on dependency inference. Sample sizes, dropout rates, and table sizes are setup identically with directional inference (Figure 3). In Figure 4a,b, Pearson's correlation noticeably underperforms at small sample sizes, followed by Kruskal Wallis test and CISC. In Figure 4c,d, function index, Pearson's chi-squared test and mutual information demonstrate robustness to dropout with perfect performance at a large sample size. Conditional entropy, CISC and Kruskal-Wallis test experience a major degradation in performance at higher dropout. Pearson's correlation consistently underperformsat all dropout rates. In Figure 4e,f, function index, Pearson's chi-squared test and mutual information still maintain perfect performance. Conditional entropy and CISC did poorly at smaller table sizes. Pearson's correlation consistently underperforms at all table sizes. In summary, function index is also effective for dependency inference. If directionality is not a concern, Pearson's chi-squared test and mutual information can also be used at large sample sizes, while conditional entropy and CISC should be avoided.

3.2 Genes for mouse cerebellar development

We also evaluated the performance of the seven methods for detecting directional dependency of gene on time on a developing

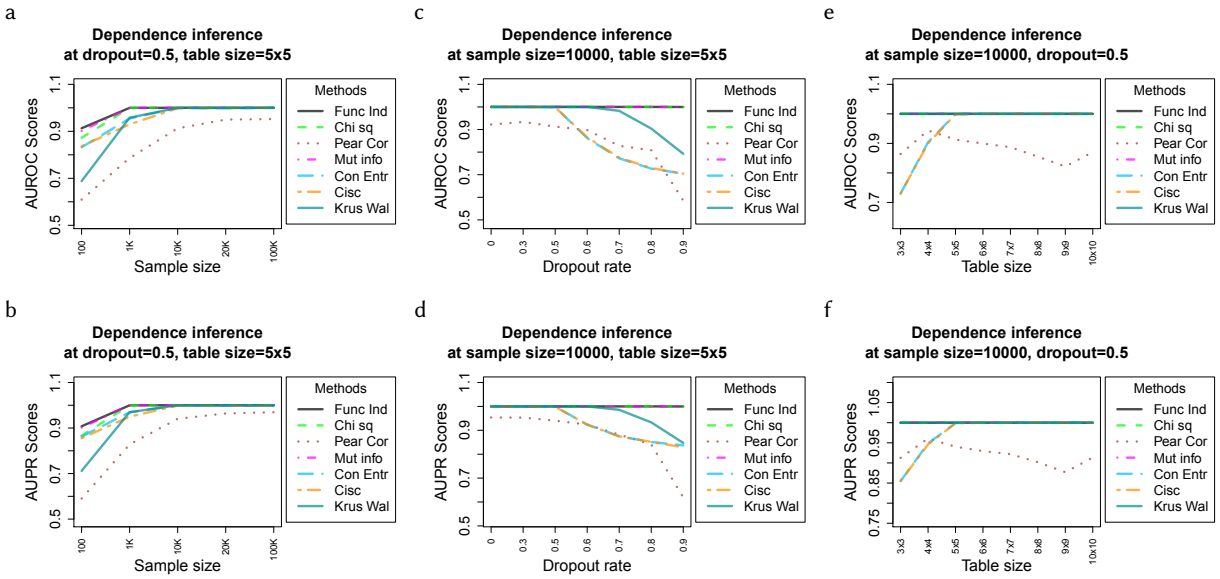

Figure 4: Dependency inference performance on simulated data at varying dropout rates, sample sizes and table sizes. The effect of sample size: (a) AUROCs and (b) AUPRs. The effect of dropout rate: (c) AUROCs and (d) AUPRs. The effect of discrete level: (e) AUROCs and (f) AUPRs.

murine cerebellum (Cb) scRNA-seq dataset, profiled at 12 development time points [6], using an independent ground truth from the Mouse Genome Database (MGD) [5].

The ground truth and data preprocessing. From the MGD, we selected 271 genes important to murine Cb development as the ground truth. These genes are associated with GO term *cerebellum development* (GO:0021549) or the term *abnormal Cb development*.

As in [6], cells with fewer than 3,500 unique molecular indices (UMIs) or greater than 15,000 UMIs were removed, accounting for dead or multiple cells in one droplet. Cells containing over 10% mitochondrial UMIs were removed. Zero-expression, mitochondrial, and ribosomal protein genes were removed. Raw counts were normalized by $U_{c,g} = \log_2(U_{c,g} \cdot U_m/U_c + 1)$, where $U_{c,g}$ is the UMI count for gene g in cell c, U_c is the UMI count for cell c and U_m is the median UMI across all cells.

To identify cell types, 1000 most overdispersed genes were selected by a Z-score for dispersion [15]. Ward hierarchical clustering [30] and dynamic tree cut [13, 14] were used to group cells by expression of the 1000 genes. Expression profiles of known marker genes were used to assign clusters to cell types [6], including progenitor cells and granule neuron precursors (GNPs).

We only used cells identified as progenitors or GNPs, as most progenitors would eventually differentiate into a granule neuron [17], making these two cell types relatively consistent and homogenous over development stages. Also, progenitors were abundant at earlier stages and GNPs were so at later stages [6], spanning the full developmental range. Together progenitors and GNPs comprised of over half of the total ~40,000 captured cells.

The normalized expression values for each of 271 selected gene in progenitor cells and GNPs were optimally discretized using `Ckmeans.1d.dp` [25, 29]. To avoid bias between genes due to table size, we clustered each gene into 12 levels, equal to the number of time points. Genes with fewer than 12 unique values were used as discrete variable directly.

Three evaluation configurations. We designed three test configurations: 1) Genes known to affect Cb development should be a function of time rather than the opposite: the direction from time to gene is *true* and from gene to time is *false*; 2) Similar to Configuration 1, but with dropout injected into the time variable, as if time is another RNA so as to mimic RNA-RNA interactions. We did not use two RNAs due to the lack of RNA→RNA directional ground truth specific to cell types; and 3) We randomly shuffled the 12 time points destroying functional dependency from time to gene. The *true* dependency is from original time to gene and the *false* dependency is from shuffled time to gene.

For Configuration 1, we built contingency tables with time as row and discrete gene level as column, and transposed it for the test of directionality. For Configuration 2, the dropout rate for the time variable of each time-gene contingency table was estimated from the percentage of zeros in the gene ranging 20–90%. For Configuration 3, false time-to-gene contingency tables were created by randomly shuffling time points but preserving gene levels, so that time and gene are independent.

Performance. Dropout imposed a strong impact on the performance of directional inference, as shown by percentages of incorrect directions inferred by the four asymmetric methods (Figure 5). Each method was evaluated for choosing the correct direction from time to gene from pairs of a time-gene contingency table versus its transpose *with* (Configuration 1) and *without* (Configuration 2) dropout added to the time variable. The four methods differ sharply

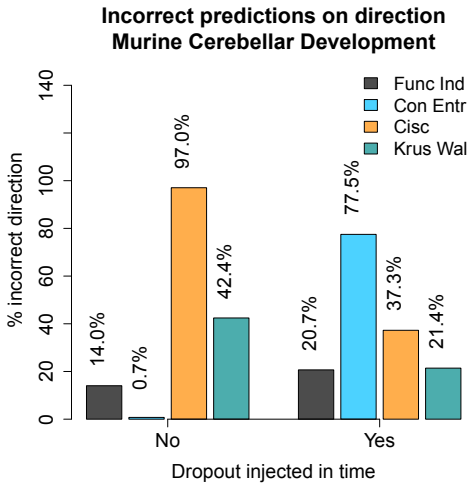

Figure 5: The impact of dropout on percentages of incorrect directions detected on cerebellar development genes by four methods. The correct direction is from time to gene.

in response to dropout: Function index is the most stable with an increase of 6.7% in error; conditional entropy extraordinarily increased its error rate by 76.8%; Surprisingly, CISC and Kruskal-Wallis test reduced in error rates by 59.7% and 21%, respectively. Dropout turned the distribution of time variable from perfectly uniform to highly non-uniform. This change of marginal distribution caused distinct responses by the methods. Function index is the most robust to this change; conditional entropy is biased in favoring a direction from uniform to non-uniform; CISC and Kruskal Wallis are biased towards a direction from non-uniform to uniform.

Figure 6 depicts ROC and PR curves under the three configurations. Configuration 1 (Figure 6a,b) tests directionality from time to gene, without dropout on the time variable. Conditional entropy has an advantage over all methods, same as explained for Figure 5: the uniform marginal distribution of time gave an unfair favor to conditional entropy. The negative effect of this bias is demonstrated in Figure 7, where top genes inferred by conditional entropy are highly uniform in time but non-uniform in gene, with nearly no dynamics and thus uninteresting for this study. Both CISC and Kruskal-Wallis performed worse than random guessing, suggesting the adverse effect of marginal distributions as opposed to that favoring conditional entropy. Function index is the second best in terms of both AUROC and AUPR.

Figure 6c,d show ROC and PR curves of Configuration 2 to test the directionality from time with dropout to gene. Conditional entropy loses the advantage and degrades to random guessing, as the non-uniformity of time and gene is similar. Kruskal-Wallis becomes the best performer. Function index demonstrates relative immunity to marginal interventions, experiencing a small drop in performance. CISC does not improve in either AUROC or AUPR.

Configuration 3 (Figure 6e,f) reveals a major flaw with conditional entropy and CISC wherein they cannot distinguish true from randomly shuffled patterns. Meanwhile, function index performed best, slightly outperforming well-established methods including Pearson's chi-squared test, Kruskal-Wallis test, mutual information, and Pearson's correlation.

Ranking time-to-gene and gene-to-time patterns. On the original data, we applied each method to rank time-to-gene and gene-to-time patterns for all 271 developmental genes. The gene-to-time patterns are included to evaluate if a method can be confused by the wrong direction. Function index and Kruskal-Wallis obtained patterns reflective of development biology literature. Figure 7a–d illustrate the top two time-gene interactions inferred by both function index and Kruskal-Wallis, long established for roles in Cb development: *Id3* was present at P5 in the internal granular layer of the Cb in the postnatal rat brain [28], consistent with the dynamic pattern in Figure 7a. *Id3* has been identified among top seven genes over-expressed in the clusters of Cb progenitors [6]. *Neurod1*, a bHLH transcription factor, is well known for Cb development via mediating the differentiation of Cb granule cells. *Neurod1* is one of the most highly expressed genes along the GNP trajectory [6]. Its conditional knockout led to widespread granule cell necrosis [19]. The top two genes by conditional entropy are *Pou1f1* and *Hoxb1* (Figure 7e,f). Their ectopic expression or mutation leads to abnormal Cb development [22, 23, 31]. However, in the wild-type developing Cb, these genes are suppressed; selecting them indicates inadequacy of conditional entropy and CISC. Finally, CISC ranked *Pou1f1*→time highly, suggesting a deficiency in capturing directions here.

4 DISCUSSION

Function index behaves most robustly at varying dropout rates. Kruskal-Wallis test is biased to marginal distributions wherein it favored a direction to a uniformly distributed variable. With a similar bias, CISC performed poorly overall, offering no benefit in either directional or dependency inference. Conditional entropy carries an opposite bias to promote a direction from a uniformly distributed variable. It seems not practical for use on real data.

From our simulation studies, it is encouraging to learn that increasing the sample size can improve the quality of directional inference. Additionally, all methods except CISC benefited from greater discrete levels in directional inference. This implies that even without renovating the scRNA-seq library preparation technique, one still has alternatives to deal with severe dropout. These possibilities seem not previously reported in the literature.

We did not evaluate directional RNA-RNA interactions, which is desirable but experimentally infeasible at present for several reasons: Coding RNAs typically do not directly interact with each other; a non-coding RNA can regulate another RNA but the former typically does not have a polyA tail, thus not measured with current scRNA-seq techniques; and the cell-type specific ground truth for directional RNA interactions is vastly lacking. However, we expect this situation to improve as scRNA-seq technology matures.

5 CONCLUSIONS

Although single-cell RNA sequencing moves us closer to cell-type specific understanding of molecular mechanisms, the dropout phenomenon represents a thin glass wall before reaching the goal. Directional and dependency inference are the workhorses for a network inference method. Although our studies suggest that the dropout challenge is greater to the former than the latter, it may

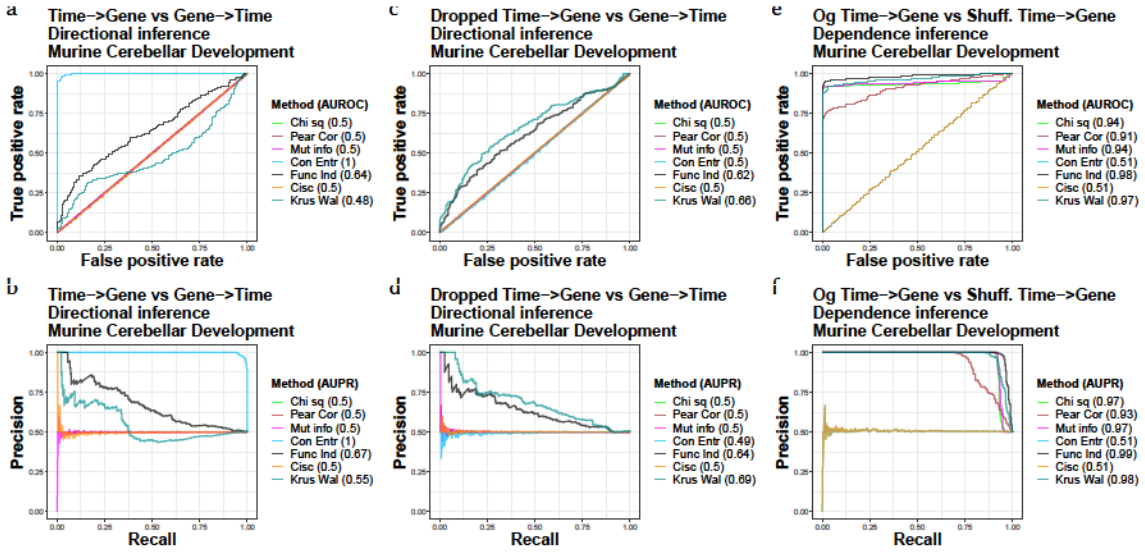

Figure 6: Performance on detecting cerebellar development genes in three configurations. Directional inference from time to gene with no dropout (Configuration 1): (a) ROC and (b) PR. Directional inference from time with dropout to gene (Configuration 2): (c) ROC and (d) PR. Dependency inference from original (Og) time versus shuffled time to gene (Configuration 3): (e) ROC and (f) PR.

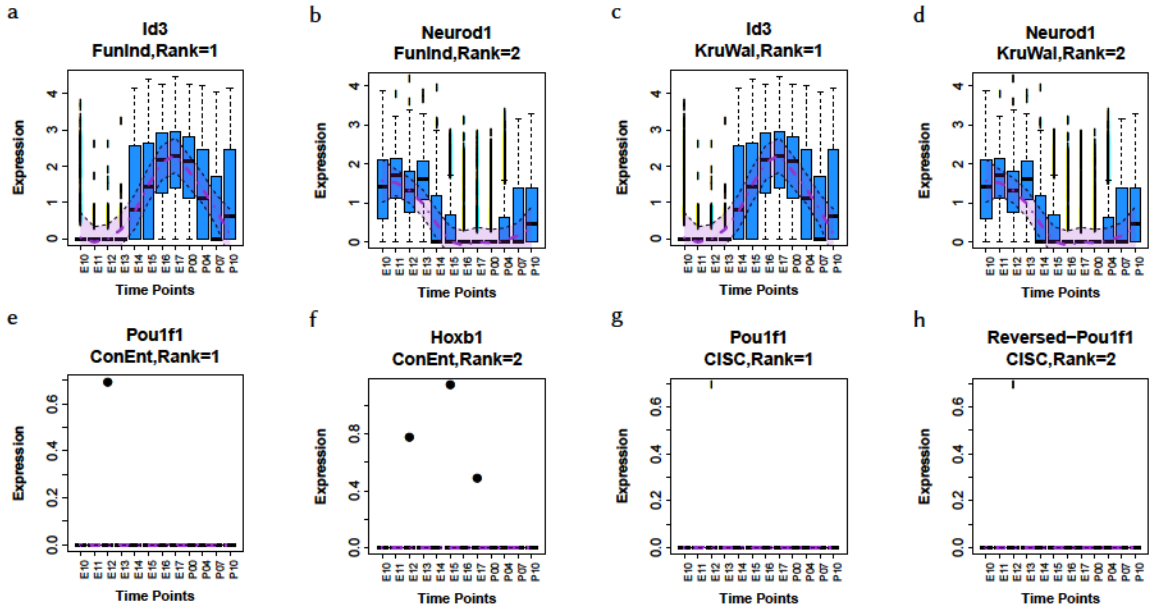

Figure 7: Expression dynamics of top two genes detected by the four asymmetric measures on cerebellar development data. Top two genes by function index: (a) Inhibitor of DNA binding 3 gene (*Id3*), and (b) Neuronal differentiation 1 (*Neurod1*). Top two genes by Kruskal-Wallis test are also (c) *Id3* and (d) *Neurod1*. Top two genes by conditional entropy: (e) POU domain, class 1, transcription factor 1 gene (*Pou1f1*), and (f) Homeobox B1 gene (*Hoxb1*). Top two genes by CISC: (g)*Pou1f1*, and (h) Reversed gene *Pou1f1*→Time.

be overcome with a method robust to marginal distributions at a sufficiently large sample size. Our findings support the use of function index as a basic statistic, which is both robust to marginal distributions and model-free to encourage novel directional pattern discovery. It enables cell-type-specific network characterization of molecular mechanisms.

6 ACKNOWLEDGMENTS

The reported work is supported by the US-Czech Fulbright Scholarship Program, USDA grant 2016-51181-25408, and US National Science Foundation grant 1661331.

REFERENCES

[1] T. S. Andrews and M. Hemberg. False signals induced by single-cell imputation. *F1000Res*, 7:1740, 2018.

[2] T. S. Andrews and M. Hemberg. M3Drop: dropout-based feature selection for scRNASeq. *Bioinformatics*, 35(16):2865–2867, 2019.

[3] P. Angerer, L. Simon, S. Tritschler, F. A. Wolf, D. Fischer, and F. J. Theis. Single cells make big data: New challenges and opportunities in transcriptomics. *Curr Opin Syst Biol*, 4:85–91, 2017.

[4] K. Budhathoki and J. Vreeken. MDL for causal inference on discrete data. In *Proc IEEE Int Conf Data Min*, pages 751–756, 2017.

[5] C. J. Bult, J. A. Blake, C. L. Smith, J. A. Kadin, and J. E. Richardson. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*, 47(D1):D801–D806, Jan 2019.

[6] R. A. Carter, L. Bihannic, C. Rosencrance, J. L. Hadley, Y. Tong, T. N. Phoenix, S. Natarajan, J. Easton, P. A. Northcott, and C. Gawad. A single-cell transcriptional atlas of the developing murine cerebellum. *Curr Biol*, 28(18):2910–2920, 2018.

[7] S. Chen and J. C. Mar. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*, 19(1):232, 2018.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, New Jersey, second edition, 2012.

[9] S. M. Hill, L. M. Heiser, T. Cokelaer, M. Unger, N. K. Nesser, D. E. Carlin, Y. Zhang, A. Sokolov, E. O. Paull, C. K. Wong, K. Graim, A. Bivol, H. Wang, F. Zhu, B. Afsari, L. V. Danilova, A. V. Favorov, W. S. Lee, D. Taylor, C. W. Hu, B. L. Long, D. P. Noren, A. J. Bisberg, The HPN-DREAM Consortium, G. B. Mills, J. W. Gray, M. Kellen, T. Norman, S. Friend, A. A. Qutub, E. J. Fertig, Y. Guan, M. Song, J. M. Stuart, P. T. Spellman, H. Koeppl, G. Stolovitzky, J. Saez-Rodriguez, and S. Mukherjee. Inferring causal molecular networks: empirical assessment through a community-based effort. *Nat Methods*, 13(4):310–318, 2016.

[10] S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lonnerberg, and S. Linnarsson. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods*, 11(2):163–166, 2014.

[11] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc*, 47(260):583–621, 1952.

[12] S. Kumar, H. Zhong, R. Sharma, Y. Li, and M. Song. Scrutinizing functional interaction networks from RNA-binding proteins to their targets in cancer. In *Proceedings (IEEE Int Conf Bioinformatics Biomed)*, pages 185–190, Madrid, Spain, 2018.

[13] P. Langfelder, B. Zhang, and S. Horvath. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, 24(5):719–720, 2007.

[14] A. T. L. Lun, D. J. McCarthy, and J. C. Marioni. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*, 5:2122, 2016.

[15] E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, and S. A. McCarroll. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

[16] P. E. Meyer. *infotheo: Information-Theoretic Measures*, 2014. R package version 1.2.0.

[17] K. Nakashima, H. Umeshima, and M. Kengaku. Cerebellar granule cells are predominantly generated by terminal symmetric divisions of granule cell precursors. *Dev Dyn*, 244(6):748–758, 2015.

[18] H. H. Nguyen. *Inference of Functional Dependency via Asymmetric, Optimal, and Model-free Statistics*. PhD thesis, Department of Computer Science, New Mexico State University, Las Cruces, USA, 2018.

[19] N. Pan, I. Jahan, J. E. Lee, and B. Fritzsch. Defects in the cerebella of conditional Neurod1 null mice correlate with effective Tg (Atoh1-cre) recombination and granule cell requirements for Neurod1 for differentiation. *Cell Tissue Res*, 337(3):407–428, 2009.

[20] K. Pearson. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

[21] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[22] C. Raoul, B. Pettmann, and C. E. Henderson. Active killing of neurons during development and following stress: a role for p75(NTR) and Fas? *Curr Opin Neurobiol*, 10(1):111–117, 2000.

[23] M. Sekiguchi, H. Abe, M. Moriya, O. Tanaka, and R. Nowakowski. Cerebellar microfolia and other abnormalities of neuronal growth, migration, and lamination in the Pit1dw-jhomozygote mutant mouse. *J Comp Neurol*, 400(3):363–374, 1998.

[24] R. Sharma, S. Kumar, H. Zhong, and M. Song. Simulating noisy, non-parametric, and multivariate discrete patterns. *R J*, 9(2):366–377, 2017.

[25] J. Song, H. Zhong, and H. Wang. *Ckmeans.1d.dp: Optimal, Fast, and Reproducible Univariate Clustering*, 2019. R package version 4.3.0.

[26] S. M. Stigler. Francis Galton's account of the invention of correlation. *Statistical Science*, pages 73–79, 1989.

[27] G. M. Sullivan and R. Feinn. Using effect size—or why the P value is not enough. *J Grad Med Educ*, 4(3):279–282, 2012.

[28] S.-F. Tzeng and J. De Vellis. Id1, Id2, and Id3 gene expression in neural cells during development. *Glia*, 24(4):372–381, 1998.

[29] H. Wang and M. Song. Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming. *R J*, 3(2):29–33, 2011.

[30] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*, 58(301):236–244, 1963.

[31] S. Zaffran, G. Odelin, S. Stefanovic, F. Lescroart, and H. C. Etchevers. Ectopic expression of Hoxb1 induces cardiac and craniofacial malformations. *Genesis*, 56(6-7):e23221, 2018.

[32] Y. Zhang, Z. L. Liu, and M. Song. ChiNet uncovers rewired transcription subnetworks in tolerant yeast for advanced biofuels conversion. *Nucleic Acids Res*, 43(9):4393–4407, 2015.

[33] Y. Zhang and M. Song. Deciphering interactions in causal networks without parametric assumptions. *arXiv*, page 1311.2707, 2013.

[34] Y. Zhang, H. Zhong, H. Nguyen, R. Sharma, S. Kumar, and J. Song. *FunChisq: Model-Free Functional Chi-Squared and Exact Tests*, 2019. R package version 2.4.9.1.

[35] H. Zhong and M. Song. Directional association test reveals high-quality putative cancer driver biomarkers including noncoding RNAs. *BMC Med Genomics*, 12(7):129, 2019.

[36] H. Zhong and M. Song. A fast exact functional test for directional association and cancer biology applications. *IEEE/ACM Trans Comput Biol Bioinform*, 16(3):818–826, 2019.