

1 **Dissecting and improving gene regulatory network inference using single-cell
2 transcriptome data**

3 Lingfeng Xue¹, Yan Wu^{1,2}, Yihan Lin^{1,2,3,*}

4 ¹Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking
5 University, Beijing, China

6 ²The MOE Key Laboratory of Cell Proliferation and Differentiation, School of Life Sciences,
7 Peking University, Beijing, China

8 ³Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies,
9 Peking University, Beijing, China

10 *Correspondence: yihan.lin@pku.edu.cn

11 **Running Title:** Pre-mRNA information for GRN inference

16

Abstract

Single-cell transcriptome data has been widely used to reconstruct gene regulatory networks (GRNs) controlling critical biological processes such as development and differentiation. While a growing list of algorithms has been developed to infer GRNs using such data, achieving an inference accuracy consistently higher than random guessing has remained challenging. To address this, it is essential to delineate how the accuracy of regulatory inference is limited. Here, we systematically characterized factors limiting the accuracy of inferred GRNs and demonstrated that using pre-mRNA information can help improve regulatory inference compared to the typically used information (i.e., mature mRNA). Using kinetic modeling and simulated single-cell datasets, we showed that target genes' mature mRNA levels often fail to accurately report upstream regulatory activities due to gene-level and network-level factors, which can be improved by using pre-mRNA levels. We tested this finding on public single-cell RNA-seq datasets using intronic reads as proxies of pre-mRNA levels and can indeed achieve a higher inference accuracy compared to using exonic reads (corresponding to mature mRNAs). Using experimental datasets, we further validated findings from the simulated datasets and identified factors such as transcription factor activity dynamics influencing the accuracy of pre-mRNA-based inference. This work delineates the fundamental limitations of gene regulatory inference and helps improve GRN inference using single-cell RNA-seq data.

34

35

36

37

38

39

40

41

42

Keywords: Gene regulatory network, network inference, single-cell RNA-seq, intronic reads, pre-mRNA

45

46

Introduction

An overarching goal of biology is understanding the regulatory principles of cell states and fates (Waddington 1957; Huang et al. 2009; MacArthur et al. 2009; Ferrell 2012). A critical step towards this goal is to reconstruct gene regulatory networks (GRNs) (Tavazoie et al. 1999; Yeung et al. 2002; Lee et al. 2002; Segal et al. 2003; Friedman 2004; Alon 2006; Levine and Davidson 2005; D'haeseleer et al. 2000; Davidson 2010; Gerstein et al. 2012; Thompson et al. 2015), with which one could delineate the global transcriptional regulatory architecture, and identify key players in the network that control cell states and transitions between them. The reconstruction of GRNs involves inferring edges linking between regulator genes and target genes, and has been traditionally based on population-level microarray or RNA-seq data (Segal et al. 2003; Friedman 2004; D'haeseleer et al. 2000; Hecker et al. 2009; Margolin et al. 2006; Huynh-Thu et al. 2010; Marbach et al. 2012; Sonawane et al. 2017). With the advent of single-cell RNA-seq (scRNA-seq) techniques, it is becoming possible to reconstruct GRNs using scarce and precious tissue samples, yielding unprecedented details into the regulatory underpinnings of key biological processes, including animal development, tumor progression, and immune response (Trapnell et al. 2014; Bendall et al. 2014; Aibar et al. 2017; Chan et al. 2017; Fiers et al. 2018; Iacono et al. 2019; Van de Sande et al. 2020; Matsumoto et al. 2017; Pratapa et al. 2020; Qiu et al. 2020; Deshpande et al. 2022; Qiu et al. 2022). However, despite numerous efforts, reconstructing GRNs from scRNA-seq data has remained challenging (Fiers et al. 2018; Pratapa et al. 2020; Argelaguet et al. 2021).

66

Despite the growing list of methods and algorithms to infer regulatory links from scRNA-seq data (Nguyen et al. 2021), benchmarking results have shown that methods that purely rely on gene expression data often cannot consistently outperform a random predictor (Pratapa et al. 2020). Alternative methods have been developed to overcome this by integrating gene expression data with additional knowledge (Aibar et al. 2017; Van de Sande et al. 2020; Yuan and Bar-Joseph 2019; Qiu et al. 2020). For example, the widely used method, SCENIC (Aibar et al. 2017), leverages regulator binding motif information derived from ChIP-seq datasets to purge links learned from the gene expression data. Nevertheless, despite some success, these

75 alternative methods do not address the fundamental challenge of GRN inference using scRNA-
76 seq data, that is, the challenge to accurately infer gene regulatory relationships from gene
77 expression data only.

78
79 To better understand this challenge, it would be helpful to closely examine gene regulation,
80 and revisit the rationale underlying GRN inference using scRNA-seq data. During gene
81 regulation, the transcription factor (TF) 's activity level influences the target gene's expression
82 level (**Fig. 1A**). The rationale to infer such a relationship using scRNA-seq data is conventionally
83 built upon the assumption that the target's mRNA level can report upstream regulatory activity,
84 and the latter can be approximated by the TF's mRNA level (**Fig. 1B**). However, a series of
85 factors, including the inherent stochasticity of biochemical reactions and expression levels or
86 dynamics of genes in the network, could undermine this assumption, affecting the accuracy of
87 relationship inference. While the effect of these factors in GRN inference has been previously
88 suggested or analyzed (Qiu et al. 2020; Cao et al. 2020; Gupta et al. 2022), we nevertheless
89 lack a systematic understanding of how network inference is limited by various factors, and
90 addressing this issue could help improve GRN inference.

91
92 Here, we quantitatively delineated constraints on GRN inference using chemical kinetic
93 modeling, simulated datasets, and public experimental datasets, and showed that
94 incorporating intronic reads in raw scRNA-seq data can improve GRN inference. Our systematic
95 dissections illustrated fundamental constraints for network inference and helped pave new
96 avenues for improving GRN inference using scRNA-seq data.

97
98 **Results**

99 **A bottom-up dissection of gene regulatory inference using kinetic modeling**

100 To delineate constraints for gene regulatory inference, we first took a bottom-up approach by
101 focusing on the chemical reactions involved in gene regulation, and asked how parameters in
102 these reactions could affect the accuracy of network inference. The rationale for a bottom-up
103 dissection is that network inference involves deducing the relationship between regulator

104 activity and target expression, yet regulator activity level may not be accurately reported by
105 target expression level during gene regulation, placing fundamental limits for network
106 inference. In other words, when the target gene expression level barely reports regulator
107 activity level, it would be very challenging to infer their relationships with any algorithms.

108
109 To quantitatively describe gene regulation, we included the reaction of transcription, which
110 produces pre-mRNAs of the target gene, the reaction for splicing, which produces mature
111 mRNAs from pre-mRNA molecules, and the reaction for mRNA degradation (**Fig. 2A**,
112 **Supplemental Fig. S1A, Methods**). Based on these reactions, we obtained differential
113 equations describing the rates of change for pre-mRNA and mRNA levels (**Supplemental Fig.**
114 **S1A, Methods**). To make the model more amenable to systematic parameter analysis, we used
115 a Boolean description of the gene activity, with its initial state being the OFF state and its
116 subsequent state being controlled by pulse-like regulator activity dynamics (**Methods**). Using
117 this model, we could quantify how accurate target gene expression level can be used to infer
118 upstream regulator activity level, and how this accuracy is affected by a range of factors,
119 including stochasticity in the reactions, regulator dynamics, and rates of reactions for
120 transcription, splicing, and degradation (**Fig. 2A**). These analyses would help elucidate how
121 factors intrinsic to gene regulatory reactions could affect the accuracy of gene regulatory
122 inference.

123
124 We quantified inference accuracy by measuring how accurately the target gene expression
125 level can capture upstream regulator activity level, defined as the fraction of time that gene
126 expression level matches with regulator activity level (**Supplemental Fig. S1B**). We then
127 modeled how a typical gene (e.g., with experimentally determined parameters averaged
128 across genes; see **Methods**) responds to a pulse of regulator activity (**Supplemental Fig. S1C**),
129 and found that using the mRNA level of the target gene to infer the regulator activity level
130 resulted in a much lower accuracy compared to the pre-mRNA level. The difference in
131 inference accuracy was observed under different types of regulator activity dynamics
132 (**Supplemental Fig. S1D**). Mechanistically, such a difference comes from the different half-lives
133 of pre-mRNA and mRNA, as the timescale of splicing is ~10 min and the timescale of mRNA

134 degradation is typically several hours, allowing the pre-mRNA level to reach steady-state faster
135 than the mRNA level. As a result, the degree of improvement in inference accuracy using pre-
136 mRNA depends on the timescales of splicing and mRNA degradation (**Fig. 2B** left). It also
137 depends on the timescale of regulator dynamics (i.e., fast or slow regulation, **Fig. 2B** right,
138 **Supplemental Fig. S1E**). Notably, for most parameters, pre-mRNA can provide a more accurate
139 inference of regulatory activity (**Fig. 2B**).

140
141 While the preceding results indicated that pre-mRNA could better capture upstream regulator
142 activity, we speculated that network inference using pre-mRNA level would be more sensitive
143 to noise than using mRNA level, as the steady-state level of pre-mRNA is much lower than that
144 of mRNA. Indeed, when introducing stochasticity to the model (**Supplemental Fig. S1F**,
145 **Methods**), the advantage of pre-mRNA was reduced for most parameter sets, and could even
146 reverse when the transcription rate is very low and the regulator dynamics are very slow (**Fig.**
147 **2C, Supplemental Fig. S1G**).

148
149 In summary, using a chemical kinetic model, we demonstrated that factors intrinsic to the gene
150 regulatory system can individually or combinatorially affect the accuracy of network inference,
151 resulting in a (theoretical) upper limit of inference accuracy (which depends on gene-specific
152 and regulator-specific kinetic parameters). Importantly, we found that the upper limit of the
153 inference accuracy is generally higher when using the pre-mRNA level of the target gene than
154 using the mRNA level, except for genes whose transcription rates are very low and are under
155 the regulation of very slow regulatory dynamics. It should be noted that while parameters in
156 our model simulations were estimated from experimental data, these simulations are only for
157 deriving general principles of how various factors affect inference accuracy and likely cannot
158 capture many of the biologically relevant scenarios (please refer to **Supplemental Note S1** for
159 further discussions on model simulations).

160
161 **Comparing mRNA-based and pre-mRNA-based inference methods using simulated datasets**
162 Because we only considered the regulation of one gene in the chemical kinetic model, we
163 asked whether the conclusions from this model would apply to complex systems containing

networks of genes. To systematically address this, we resorted to synthetic single-cell datasets generated by a state-of-the-art single-cell simulation engine, dyngen (Cannoodt et al. 2021), which could simulate stochastic pre-mRNA and mRNA dynamics for gene regulatory networks. With this tool, we could assess the effects of network-level and gene-level factors on GRN inference using either pre-mRNA or mRNA levels of target genes (**Fig. 1B**), which may provide a systematic picture of factor dependency of inference accuracy (**Fig. 3A, Supplemental Fig. S2, Methods**).

We first asked whether leveraging pre-mRNA levels of target genes could improve GRN inference compared to using mRNA levels. To address this, we simulated networks containing different backbones (**Supplemental Fig. S3A**), including linear, cycle, bifurcating, and converging backbones. In these networks, we observed four general patterns of regulator activity dynamics, with each network displaying a distinct combination of dynamic patterns (**Supplemental Fig. S3B,C**). We then used a random forest-based algorithm, GENIE3 (Huynh-Thu et al. 2010), to reconstruct GRNs from synthetic single-cell datasets, and quantified the inference accuracy using AUPR (Area Under the Precision-Recall curve) (**Fig. 3A, Methods**). We found that the pre-mRNA-based method confers a significantly higher accuracy of the reconstructed GRN compared to using mRNA levels for all four network backbones (**Fig. 3B**). In addition to GENIE3, we tested five other methods of GRN inference (including Pearson's correlation, propr (Quinn et al. 2017), ARACNE (Margolin et al. 2006), PIDC (Chan et al. 2017), and TIGRESS (Haury et al. 2012)) and found that pre-mRNA-based method consistently outperforms mRNA-based method and that combining pre-mRNA and mRNA counts did not provide an advantage compared to pre-mRNA-based method (**Supplemental Fig. S3D**).

When comparing the relative performance between pre-mRNA-based and mRNA-based methods, we found that the pre-mRNA-based method achieved the highest relative performance for the cycle backbone, while for the bifurcating backbone, the relative performance is the lowest (**Supplemental Fig. S3E**). This result indicated that network topology affects the extent of improvement when using the pre-mRNA-based method, and in networks where regulator activities exhibit more dynamics (e.g., the cycle backbone), the pre-

194 mRNA-based method appears to confer a relatively higher advantage compared to the mRNA-
195 based method. It should be noted that we simulated relatively small networks due to limited
196 computational resources and it would be helpful to further investigate the above findings by
197 simulating larger networks.

198

199 **Analyzing the effect of gene-level factors using simulated datasets**

200 While GRN inference could be improved by using pre-mRNA, the inference accuracy (as
201 quantified by AUPR) is far below 1 (**Fig. 3B**). We speculated that the relatively low accuracy of
202 inferred GRNs could arise from at least two different categories of factors. The first category is
203 gene-level factors, such as kinetic parameters of genes, as suggested by the results from the
204 kinetic model. The other category is related to network topology, such as the abundance of
205 motifs. For example, the two leaf nodes of a bi-fan motif could be falsely connected during
206 inference (Marbach et al. 2010). Both types of factors would lower the theoretical upper
207 bound of the inference accuracy and pose challenges for GRN inference. Quantifying how the
208 accuracy of inferred GRNs depends on gene-level and motif-level factors may provide guidance
209 for improving GRN inference.

210

211 To dissect gene-level factors, we explored how the inference accuracy of all four types of
212 networks is affected by kinetic parameters of genes in the network, such as transcription rate,
213 mRNA degradation rate, and protein degradation rate (which affects regulator activity
214 dynamics). To do so, we measured the parameter dependency of AUPR using simulated
215 datasets generated by scanning individual kinetic parameters across a wide range
216 (**Supplemental Fig. S4**). We found that the pre-mRNA-based method is generally more
217 susceptible to changes in transcription rate compared to the mRNA-based method, and when
218 the transcription rate is very low, the performance of the pre-mRNA-based method drops
219 below that of the mRNA-based method (**Supplemental Fig. S4A**). This result is consistent with
220 the finding from the kinetic model (**Supplemental Fig. S1G**). Notably, the cycle backbone
221 behaved differently from other backbones (**Supplemental Fig. S4A**), which we will dissect later.

222

223 For the second kinetic parameter, mRNA half-life, we found that the mRNA-based method is

much more susceptible to it than the pre-mRNA-based method for all four backbones (**Supplemental Fig. S4B**). More specifically, with a shorter mRNA half-life, the mRNA-based method can achieve higher precision. Such dependency is consistent with the finding from the kinetic model (**Fig. 2B**), that is, when mRNA half-life is shorter, the time delay between mRNA level and upstream regulator activity is shorter, and thus the mRNA level can allow a more accurate inference. For the third kinetic parameter, protein half-life, we found that the pre-mRNA-based method is much more susceptible to it compared to the mRNA-based method for all four backbones (**Supplemental Fig. S4C**). More specifically, with a shorter protein half-life, the pre-mRNA-based method can achieve higher precision, as the regulator dynamics would be more rapid and thus could be better distinguished from other regulators. This dependency was similarly observed in the kinetic model (**Supplemental Fig. S1E**). Notably, these parameters can combinatorially affect the inference accuracy (**Supplemental Fig. S4D,E**). For completeness, we also analyzed the effect of splicing time (**Supplemental Fig. S4F**) and found that it affects pre-mRNA-based method but not mRNA-based method, and that if the splicing time is longer at a given transcription rate, the count number of pre-mRNA will be larger, and thus the inference will be less affected by stochastic noise.

Together, except for the dependency of cycle backbone on transcription rate, other networks display parameter dependencies that are overall consistent with findings from the kinetic model. These results suggest that while the pre-mRNA-based method generally outperforms the mRNA-based method, gene-level factors greatly affect the inference accuracy of either method (**Fig. 3C**).

247 Analyzing network-level factors using simulated datasets

Because genes can assemble into network motifs, gene pairs in the same motif may have correlated expression levels without having regulatory relationships, giving rise to motif-related errors that affect inference accuracy. It was previously shown that four types of motif-related errors are common in network inference, namely fan-out error, fan-in error, cascade error, and feed-forward loop (FFL) error (Marbach et al. 2010). Out of the four motifs, we focused on fan-out and cascade motifs (**Supplemental Fig. S5A,B**), as the other two motifs are

254 not enriched in the four simulated network backbones. By analyzing pre-mRNA-based inferred
255 GRNs from the simulated datasets, we found that fan-out error pervades all four backbones
256 (**Supplemental Fig. S5C, Methods**), while cascade error is only significant in bifurcating and
257 converging backbones (**Supplemental Fig. S5D**). These results demonstrated that motif-
258 related errors (**Supplemental Fig. S5E**) affect the accuracy of inferred GRNs in a network
259 backbone-dependent manner.

260
261 To further dissect network backbone-specific motif errors, we focused on the cycle backbone,
262 as the preceding gene-level analysis showed that the inference accuracy of the cycle network
263 exhibits unexplained dependency on gene transcription rate (**Supplemental Fig. S4A**). We thus
264 sought to address whether such unexplained dependency could be accounted for by motif-
265 related errors.

266
267 As a notable feature in the motif analysis result, the fan-out error in the cycle network has
268 prediction confidence even higher than true links in the network (**Supplemental Fig. S5C**), that
269 is, the two genes at the leaf nodes of the fan-out motif were inferred as having regulatory
270 relationships much more often than gene pairs that have true regulatory relationships. By
271 examining the temporal dynamic traces of all three genes in a fan-out motif inside a cycle
272 network, we found that indeed the two leaf-node genes are temporally correlated
273 (**Supplemental Fig. S5F**). Thus, the relatively high fan-out error appeared to depend on the
274 cycling dynamics of the upstream regulator. We observed that the dependency of network
275 inference accuracy on transcription rate is dependent on protein half-life (e.g., the degree of
276 regulator dynamics) (**Supplemental Fig. S5G**), and thus regulator dynamics-dependent motif
277 errors (e.g., fan-out errors) could potentially account for the abnormal dependency of
278 inference accuracy on gene transcription rate for the cycle backbone (**Supplemental Fig. S4A**).
279 To test this, we calculated fan-out and cascade errors under varying kinetic parameters, and
280 found that the fan-out error becomes more significant as the protein half-life increases to 5
281 hours or above (**Supplemental Fig. S5H**). In contrast, when the protein half-life is short (1 or 2
282 hours), the fan-out error is not as significant (**Supplemental Fig. S5H**), and the abnormal
283 dependency of inference accuracy on gene transcription rate disappears (compare

284 **Supplemental Fig. S4A** cycle backbone with **Supplemental Fig. S5G**).

285
286 Together, these results explained the abnormal dependency of inference accuracy on
287 transcription rate for the cycle network when using the pre-mRNA-based method, and more
288 generally, highlighted that gene-level and network-level factors can combinatorially constrain
289 GRN inference accuracy in a network backbone-dependent manner.

290

291 **Evaluating GRN inference using public scRNA-seq datasets**

292 Thus far, we have systematically dissected factors constraining GRN inference accuracy, and
293 have demonstrated that the pre-mRNA-based method outperforms the mRNA-based method
294 under most scenarios. Yet, the comparison between the two methods was based on *in silico*
295 data, and it is unclear whether the pre-mRNA-based method would still perform as well in real
296 datasets.

297

298 To compare the two methods using real datasets, we needed to address two challenges. First,
299 unlike simulated datasets that contain measurements of mRNA and pre-mRNA levels, typical
300 scRNA-seq data is acquired by isolating and sequencing poly(A)-containing mRNA molecules,
301 and thus in principle does not measure pre-mRNA molecules directly. However, several recent
302 studies have shown that many scRNA-seq data contains reads that can be mapped to intronic
303 regions of genes (Gaidatzis et al. 2015; La Manno et al. 2018; Wu et al. 2022), most likely arising
304 from the binding of Oligo(dT) primers to the internal adenosine-rich regions of pre-mRNA
305 molecules (La Manno et al. 2018), allowing one to estimate the expression levels of pre-mRNAs
306 using intronic reads. Second, compared to simulated datasets that have ground-truth GRNs,
307 real datasets acquired with natural systems do not have ground-truth GRNs, posing a critical
308 challenge to the evaluation of inference accuracy. To alleviate the challenge, we relied on two
309 types of ‘ground-truth’ GRNs for evaluation. The first ground-truth GRN is a network with
310 edges having A or B confidence scores in the DoRothEA database (Garcia-Alonso et al. 2019)
311 (**Methods**), a carefully curated database integrating multiple data sources. DoRothEA network
312 has a relatively low number of nodes and edges, but high confidence levels. The second
313 ground-truth GRN is from transcription factor (TF) binding motifs (curated using ChIP-seq data,

314 designated as Motif GRN) (Aibar et al. 2017), which has much more nodes and edges
315 compared to DoRothEA network, but with relatively low confidence levels (**Methods**).
316

317 Next, by using intronic reads to approximate pre-mRNA levels and using two complementary
318 ground-truth GRNs, we sought to infer GRNs and evaluate inference accuracy using 30 public
319 scRNA-seq human and mouse datasets (**Supplemental Table S1, Methods**). The rationale for
320 choosing these datasets is two-fold: the underlying biological processes are representative of
321 key mammalian processes, and these data were acquired using 10x Genomics platform and
322 are thus comparable in terms of sequencing library construction. To infer GRNs, we created
323 pre-mRNA (i.e., based on intronic reads) and mRNA (i.e. based on exonic reads) expression
324 matrices for each dataset, which were fed into the GEINE3 algorithm (Huynh-Thu et al. 2010)
325 (**Methods**). Due to the large size and the sparse nature of GRNs, we devised a metric, average
326 early precision (AEP), to evaluate the inferred GRNs, which measures the accuracy of the top
327 network (**Methods**).
328

329 Using DoRothEA network as the ground truth, we found that above a threshold level of
330 precision (i.e., ~0.15), the pre-mRNA-based GRNs were more accurate than mRNA-based GRNs
331 (**Fig. 4A,C**). Yet for GRNs with generally low precision (below 0.15), the two methods were
332 comparable (**Fig. 4A,C**), which could be due to the relatively small network size of the
333 DoRothEA GRN. We then asked whether the enhanced inference accuracy of pre-mRNA-based
334 GRNs would be susceptible to dropouts in reads. To address this, we subsampled both intronic
335 and exonic reads to ~30% and performed GRN inference using subsampled reads. We found
336 that while the inference accuracies of both methods were reduced, pre-mRNA-based GRNs
337 were generally still more accurate than mRNA-based GRNs (**Supplemental Fig. S6A**),
338 suggesting that the performance of the pre-mRNA-based method was not affected by
339 dropouts more than that of the mRNA-based method. We next evaluated the results with the
340 Motif GRN, which has a much larger network size. The pre-mRNA-based method
341 outperformed the mRNA-based method for most networks (**Fig. 4B,D**). We further analyzed
342 the performance of our inferred networks using AUPR as a metric, and our evaluation results
343 were found to be similar to those obtained using the AEP (**Fig. 4E,F**). However, we observed

344 that AUPR is generally less sensitive than AEP, particularly when evaluating the results using
345 the Motif network as ground-truth (**Fig. 4F**). For some datasets, the mRNA-based method
346 outperformed the pre-mRNA-based method, suggesting the presence of factors contributing
347 to dataset-specific performance difference (see **Supplemental Note S2** for further discussions).
348 Moreover, because different datasets exhibited varying AUPRs, we asked whether the
349 observed difference in inference accuracy could be attributed to difference in data sparsity in
350 the count matrix. We found that AUPR ratio (i.e., relative to random) from either method is
351 not positively correlated with the numbers of exon or intron UMI per cell (**Supplemental Fig.**
352 **S6B**), implicating that inference accuracy is more affected by factors other than data sparsity.
353

354 Furthermore, we compared inferred GRNs using both methods from datasets where
355 exogenously induced TFs were used to program cell fates, such that we should be able to
356 recover the TFs being induced in the inferred network without needing to rely on GRN
357 databases as the ground truth. By analyzing public datasets taken at two different time points
358 post-induction (Hersbach et al. 2022), we found that while both methods identified
359 exogenously induced TFs as hub nodes in the GRN, only the pre-mRNA-based method
360 appeared to capture the temporal behaviors of the TFs (**Supplemental Fig. S6C-D**,
361 **Supplemental Note S3**). Additionally, we compared the two methods using a single-nucleus
362 RNA-seq dataset of mouse skeletal myofibers (Petrany et al. 2020), and found that even with
363 an increased intronic read fraction (an average of ~74%), the performance of the pre-mRNA-
364 based method was not much enhanced when compared to the mRNA-based method and the
365 evaluation of inferred GRNs is limited by the ground-truth database (i.e., the DoRothEA
366 network may lack regulatory interactions specific to this particular cell type) (**Supplemental**
367 **Fig. S6E**). Therefore, it is necessary to use multiple evaluation methods to accurately assess
368 the performance of different approaches.

369
370 We next focused on the human forebrain dataset (hFB) (La Manno et al. 2018), as the pre-
371 mRNA-based method consistently and greatly outperformed the mRNA-based method when
372 evaluated using two ground-truth GRNs. For the top links in the inferred network, pre-mRNA-
373 based GRN achieved a precision of ~50% when recall is ~4%, while mRNA-based GRN could

374 only achieve a precision of ~20% when the recall is ~2% (**Fig. 5A**).
375

376 To analyze the biological relevance of the inferred GRN from the human forebrain dataset, we
377 extracted the high-precision top network of the pre-mRNA-based GRN, and analyzed the hub
378 TFs in the top network (**Fig. 5B**). We found that the hub TFs are all related to human neural
379 development, consistent with the biological process during which the data was acquired
380 (which contrasts with mRNA-based GRN (**Supplemental Fig. S6F**)). We further evaluated the
381 precision of each hub TF using the Motif GRN as the ground truth, and found that all these hub
382 TFs, except the largest hub TF NEUROD2, have much higher precision than random, with a
383 median precision of ~35% (**Fig. 5C**). To account for the small increase in precision relative to
384 random control for NEUROD2, we speculated that it was due to the lack of reliable motif data
385 for this TF, and thus resorted to cell type-specific ChIP data (Bayam et al. 2015) as the ground
386 truth to further evaluate the inferred targets of NEUROD2. With the cell type-specific ChIP
387 data, NEUROD2 has a precision of ~35% (**Fig. 5C**), illustrating the importance of using relevant
388 ground truth for evaluation. Furthermore, using cross-species target overlap as a ground-
389 truth-free evaluation method, we found that between mouse and human forebrain datasets
390 (mFB and hFB), pre-mRNA-based GRNs contain more overlapping target genes compared to
391 mRNA-based GRNs (**Fig. 5D**).
392

393 Besides the mouse and human forebrain datasets, we also analyzed the biological relevance
394 of top five hub TFs from mRNA-based or pre-mRNA-based GRN for four other datasets, and
395 found that hub TFs inferred by the pre-mRNA-based method were more related to the
396 corresponding biological processes (**Supplemental Table S2**). Additionally, we noted that
397 ribosomal proteins were more frequently inferred as TFs in the mRNA-based method
398 compared to the pre-mRNA-based method (**Supplemental Table S2**).
399

400 Taken together, we demonstrated that the pre-mRNA-based method generally outperforms
401 the mRNA-based method for experimental scRNA-seq datasets, and that the ground-truth-
402 free method of evaluation using forebrain datasets provides consistent support.
403

404 Factor-dependency analysis of inferred GRNs from experimental datasets

405 While we have established the overall advantage of the pre-mRNA-based method over the
406 mRNA-based method with experimental datasets, it remained unclear what affects the
407 relative performance between the two methods and whether such an advantage arises from
408 the same factors as illustrated in the simulated datasets.

409 We focused on the factor-dependency analysis of specific GRNs and explored whether the
410 inference accuracy of links within GRNs depends on factors related to the target gene or the
411 TF (**Fig. 6A**). We chose four experimental datasets, whereby for all datasets, the pre-mRNA-
412 based method outperformed the mRNA-based method using Motif ground-truth GRN, but
413 with varying degrees of improvements (**Fig. 4B**). We first focused on factors related to target
414 genes, including expression level and mRNA half-life, which relate to kinetic parameters of
415 genes and have been shown to affect inference accuracy in simulated datasets. To analyze the
416 effect of the target gene's expression level on inference accuracy, we sorted target genes into
417 two bins by their expression levels (separately for pre-mRNA and mRNA), and averaged the
418 top-10-precision of each target (i.e., mean inference precision of the top 10 inferred TFs of
419 each target evaluated using Motif GRN) within each bin (**Supplemental Fig. S7A**). We found
420 that for all four datasets, the mean top precision of the GRN inferred with the pre-mRNA-based
421 method exhibits a much stronger dependency on the target's pre-mRNA level compared to the
422 dependency of the mRNA-based method on the target's mRNA level (**Fig. 6B**). These results
423 are consistent with the findings from the kinetic model (**Supplemental Fig. S1G**) and simulated
424 datasets (**Fig. 3C**), indicating that target gene expression levels affect inference accuracy in real
425 datasets as well.

427
428 Compared to the analysis of expression level, analyzing the effect of the target gene's mRNA
429 half-life on inference accuracy is more challenging, as we do not have direct measurements of
430 mRNA half-lives for the cell type of interest. We thus adopted mRNA half-life data from
431 TimeLapse-seq measurement in K562 cells (Schofield et al. 2018), sorted target genes into two
432 bins by their mRNA half-lives, and averaged the top-10-precision of each target within each
433 bin (evaluated using Motif GRN). We found that in all four datasets, the mean top precision of

434 the GRN inferred with the mRNA-based method exhibits a stronger dependency on the mRNA
435 half-life compared to that of the pre-mRNA-based method (**Fig. 6C**), consistent with findings
436 from simulated datasets (**Fig. 3C**).

437
438 We next focused on TFs, as the regulation time or dynamics of TFs have been shown to affect
439 inference accuracy in our *in-silico* results. To analyze the potential effect of TF dynamics in real
440 datasets, we resorted to an indirect characterization of TF dynamics measured from single-cell
441 ATAC-seq data (Buenrostro et al. 2015) (**Methods**). Across four different datasets, we found
442 that the advantage of the pre-mRNA-based method over the mRNA-based method becomes
443 more obvious when TF activities are more dynamic (**Fig. 6D, Supplemental Fig. S7B**),
444 consistent with findings from simulated datasets (**Supplemental Fig. S4C**). Moreover, such a
445 comparison would be further facilitated by using cell type-specific ATAC-seq data
446 (**Supplemental Fig. S7C**).

447
448 In summary, we applied factor-dependency analysis of GRN inference using scRNA-seq
449 datasets. Our analyses indicated that the precision of GRN inferred with the pre-mRNA-based
450 or the mRNA-based method could be explained by factors related to the target gene and the
451 TF. This result supported the hypothesis that the advantage of the pre-mRNA-based method
452 over the mRNA-based method in the experimental datasets arises because pre-mRNA levels
453 could better report upstream TF activities in these data. Thus, GRN inferred using the pre-
454 mRNA-based method generally has a higher precision compared to GRN inferred using the
455 mRNA-based method for both simulated and experimental single-cell data.

456
457 **Discussion**

458 Researchers are leveraging the ever-growing single-cell transcriptome data for reconstructing
459 GRNs, and many efforts have been devoted to the improvement of inference methods and
460 algorithms (Matsumoto et al. 2017; Aibar et al. 2017; Chan et al. 2017; Qiu et al. 2020; Nguyen
461 et al. 2021; Pratapa et al. 2020). A parallel line of work has been centered on integrating
462 additional omics data with single-cell transcriptome to improve GRN inference (Argelaguet et

al. 2021; Welch et al. 2017; Stuart et al. 2019; Hao et al. 2021; Efremova and Teichmann 2020). While GRN inference using single-cell transcriptome data can indeed be improved by using new methods and algorithms, or by incorporating additional measurement modalities, it is critical to deciphering the limits of GRN inference using single-cell gene expression data alone, as it would provide key guiding principles for improving GRN inference. In this work, we provided a quantitative framework for delineating the effects of gene-level and network-level factors on the accuracy of GRN inference using single-cell transcriptome data, identified fundamental limitations of the inference accuracy, and described a new method that leverages typically disregarded pre-mRNA information for improving GRN inference using single-cell transcriptome measurement alone.

Both the kinetic model and the simulated single-cell datasets helped depict the fundamental limits of GRN inference using the mRNA-based method, which can be alleviated by leveraging pre-mRNA information. In particular, because mRNA is generally more long-lived compared to pre-mRNA, the level of mRNA exhibits a generally lower accuracy compared to pre-mRNA level for reporting the upstream regulator activity. Yet, the accuracy of the pre-mRNA-based method diminishes when the gene is transcribed at a low level, limiting the overall advantage of the pre-mRNA-based method over the mRNA-based method. More generally, while the pre-mRNA-based method overall outperforms the mRNA-based method for GRN inference, *in silico* results provided a systematic picture of the limitations of GRN inference using either method imposed by network backbones and motifs, gene expression level, and TF activity dynamics, demonstrating that inference accuracy is impacted by characteristics intrinsic to genes and that dynamics in the network can be leveraged by the pre-mRNA-based method for improved network inference compared to the mRNA-based method.

The overall advantage of the pre-mRNA-based method was also evident when inferring networks using experimental single-cell RNA-seq datasets. Such an advantage was demonstrated by utilizing two independently sourced ground-truth GRNs (Aibar et al. 2017; Garcia-Alonso et al. 2019) to evaluate inferred networks, and was further supported by ground-truth-free evaluation of two related datasets. Notably, the factor-dependency analysis

493 provided quantitative insights into the effects of gene-level factors on the accuracy of inferred
494 networks, and the observed effects were consistent with findings from corresponding *in silico*
495 studies. These results provided critical support for the conclusion that the pre-mRNA-based
496 method can offer improved GRN inference with real datasets compared to the mRNA-based
497 method, as we could explain how such improvements are achieved. Yet, several aspects of the
498 proposed GRN inference method could be further improved. For example, to mitigate the
499 generally low intronic read counts, it is possible to aggregate cells with similar transcriptomes;
500 to improve the versatility of the method, it would be intriguing to explore the possibility of a
501 hybrid method whereby both mRNA and pre-mRNA reads are leveraged; to improve the
502 ground-truth networks, it would be helpful to leverage experimental datasets whereby many
503 TFs are directly perturbed to alter cell fates (Hersbach et al. 2022; Joung et al. 2023).

504
505 More generally, this work highlights the importance of characterizing and understanding the
506 temporal dynamics in gene regulatory networks. Fundamentally, GRN inference using single-
507 cell gene expression is constrained by temporal mismatches between the target expression
508 level and upstream TF expression level. Leveraging the pre-mRNA level of the target gene
509 partially addresses such temporal mismatches, and the pervasively dynamic TFs in the GRN
510 (Levine et al. 2013) further help alleviate the challenge. Nevertheless, existing approaches
511 cannot address the challenge that temporal and dynamic TF activity may not be accurately
512 captured by TF mRNA expression level, i.e., TF mRNA level may not a good proxy of its activity.
513 This challenge is reflected in the modest improvements of our new inference method on
514 experimental datasets compared to the mRNA-based method (**Fig. 4A-D**), which also
515 demonstrated the inherent limitations when using mRNA levels of TFs as a proxy of their
516 activities as it ignores other factors affecting TF activities. It is thus conceivable that a better
517 estimation of TF activity could be introduced to alleviate this challenge, such as by summing
518 over pre-mRNAs from genes in the same regulons (Wu et al. 2022). Furthermore, to overcome
519 the sparsity of intronic reads in typical scRNA-seq datasets, single-cell transcriptome method
520 using metabolic labeling can be leveraged to measure newly produced mRNAs, which has been
521 shown to improve gene regulatory inference (Cao et al. 2020).

523

524

Methods

525

Kinetic model of gene expression

526

Details on model and model simulations

527

We used the kinetic model of gene expression in the literature (La Manno et al. 2018), where both unspliced (pre-mRNA) and spliced mRNA species are included in the model (designated as u and s , respectively). In the first equation (Eq. 1), the unspliced mRNA level (u) is increased by transcription at the rate of α and is reduced by splicing at the rate of βu . In the second equation (Eq. 2), spliced mRNA level (s) is increased by splicing at the rate of βu and is reduced by degradation at the rate of γs . More specifically, in these equations, α denotes the transcription rate, β denotes the splicing rate constant of unspliced mRNA, and γ denotes the degradation rate constant of mRNA.

535

$$\frac{du}{dt} = \alpha - \beta u \quad (\text{Eq. 1})$$

536

$$\frac{ds}{dt} = \beta u - \gamma s \quad (\text{Eq. 2})$$

537

Because transcriptional regulation is often dynamic, i.e., the transcription factor is temporally activated and deactivated (Wu et al. 2022), we needed to explore how such dynamics affect gene regulatory inference. To do so, we included TF dynamics in the above equations by modeling a temporally changing transcription rate that is governed by TF activation and deactivation. More specifically, we assumed that the transcription rate follows that:

542

$$\alpha(t) = \begin{cases} \alpha_1, & 0 < t < t1 \\ \alpha_2, & t1 < t < t2 \\ \alpha_1, & t2 < t < t3 \end{cases} \quad (\text{Eq. 3})$$

543

In Eq. 3, α_1 denotes the basal transcription rate without TF activation, α_2 denotes the transcription rate upon TF activation, and $t1, t2$, and $t3$ denote the time points of step-like changes. More generally, we used $g(t)$ in other places of the manuscript (e.g., **Fig. 2A** and **Supplemental Fig. S1**) to denote the temporal pattern of the transcription rate. Typical splicing rate constant and degradation rate constant are 0.1/min and 0.01/min respectively, which were adopted from experimental values (Rabani et al. 2014). To evaluate the accuracy of gene regulatory inference, we solved these equations analytically and simulated the corresponding master equations. For stochastic simulations, we used master equation to model the dynamics,

551 which were solved by the ‘GillespieSSA’ package in R (R Core Team 2022).

552

553 *The calculation of inference accuracy*

554 The inference accuracy is intended to describe the ability of pre-mRNA and mRNA levels to
555 accurately infer upstream regulatory activity. By quantifying inference accuracies under
556 various model parameters, we can study how diverse factors affect gene regulatory inference
557 (as illustrated in **Fig. 2A**). In our model, the regulatory activity has two distinct states: ON and
558 OFF. To calculate accuracy, we first binarize the downstream expression levels (pre-mRNA or
559 mRNA) by assigning the state ON if the expression level is larger than half of the maximal value;
560 otherwise, the state is considered OFF. After binarizing the downstream expression levels, we
561 can compare them with the upstream regulatory activity dynamics (which is a binary trace) to
562 calculate the accuracy. The accuracy is defined as the proportion of time when the
563 downstream expression level (pre-mRNA or mRNA) can correctly match the upstream
564 regulatory activity:

565 Accuracy = (Number of time points with correct predictions) / (Total number of time points).

566 For example, in **Supplemental Fig. S1B**, the mRNA level can correctly reflect the regulatory
567 activity 55% of the time, while the pre-mRNA level can do so 95% of the time. In **Fig. 2**, we
568 used the above metric to determine how well output signals (i.e., pre-mRNA dynamics or
569 mRNA dynamics) track input signal (i.e., upstream regulatory signal which is binary). As we
570 illustrated, we thus needed to binarize target gene expression levels based on half maximal
571 value, which allows quantitative determination of whether input and output are matching in
572 time.

573

574 **Evaluating network inference accuracy using simulated single-cell datasets**

575 *Generation of simulated single-cell datasets*

576 To evaluate and compare mRNA-based and pre-mRNA-based methods in network inference,
577 we generated synthetic single-cell gene expression data using stochastic simulations of pre-
578 defined networks. These simulated datasets were generated by the dyngen package
579 (Cannoodt et al. 2021). More specially, we selected 4 of the default backbones (linear, cycle,
580 bifurcating, and converging), set the number of genes and cells, and generated a series of

581 dynamical trajectories, which were then converted into expression matrices for network
582 inference (note that the outputs of the simulation are mRNA and pre-mRNA counts, instead
583 of intron and exon reads as in experimental datasets). It should be noted that the default
584 splicing rate in the dyngen package (v0.4.0) was inappropriately set for mammalian cells, and
585 we used the splicing rate from the experimentally measured value (Rabani et al. 2014) to
586 replace the default values (splicing_rate = log(2)/splice_time, splice_time = 10 min). For other
587 parameters, we used the default values in the package.

588

589 *Network inference for simulated single-cell data*

590 To generate the single-cell expression matrix from the dynamical trajectories output by dyngen,
591 we sampled the trajectories in equal time intervals (1 h) to obtain the input matrix for the GRN
592 inference algorithm. Because experimental single-cell RNA-seq data is typically log normalized,
593 we thus log-normalized the simulated expression counts (i.e., log(count+1)) in order to better
594 compare with experimental data. For network inference, we used a random forest regression
595 method, GENIE3 (Huynh-Thu et al. 2010), which was developed for use with bulk
596 transcriptomes and has become a standard method for GRN inference with single-cell
597 transcriptomes. Note that in the input gene expression matrix, the information of transcription
598 factor has been utilized and that only regulatory relationships from one TF to another gene
599 (including other TFs) were considered for inference. For the mRNA-based method, because it
600 is a typically used standard method, we followed the guideline provided in the package
601 instruction and provided GENIE3 with mRNA expression matrices for both TFs and target genes.
602 For the pre-mRNA-based method, we provided GENIE3 with the pre-mRNA expression matrix
603 of the target genes and the mRNA expression matrix for the TFs, whereby the mRNA
604 expression levels of TFs were used to regress the pre-mRNA expression levels of target genes
605 by GENIE3.

606

607 *Motif error analysis*

608 For the analysis of motif-related errors, we followed the previous work (Marbach et al. 2010).
609 The prediction confidence of a potential regulatory relationship is the rank of this relationship
610 in the inferred network. The link with the highest weight in the inferred network has a

prediction confidence of 100%, while the link with the lowest weight in the inferred network has a prediction confidence of 0. In **Supplemental Fig. S5**, the links for all motifs were extracted, and the prediction confidence for those links was calculated. The distributions of prediction confidence were shown in the figure. As a comparison, the prediction confidence for background links is also calculated (for all links in the synthetic network, i.e., TRUE, and for links that are not in the synthetic network, i.e., FALSE). Fan-out error and cascade error were analyzed and the prediction confidence was calculated for motif errors and background (all true links and all false links).

619 620 **Evaluating network inference accuracy using experimental single-cell datasets**

621 *Brief description of experimental datasets*

622 We investigated a total of 30 scRNA-seq datasets, all acquired using 10x Genomics platform
623 (Zheng et al. 2017). These data include single cells of diverse types and states, ranging from
624 human to mouse cells, from cells extracted from tissues to *in vitro* cultured cells, and from
625 cells in differentiated states to cells with pluripotency. Detailed information regarding datasets
626 can be found in the supplementary table (**Supplemental Table S1**).

627 628 *Preprocessing of experimental datasets*

629 Following the established convention (La Manno et al. 2018), we used exon read counts to
630 quantify mRNA levels, and intron read counts to quantify pre-mRNA levels. For a few datasets,
631 the expression matrices (including both intron and exon counts) have been provided by
632 indicated publications, and we directly incorporated these matrices for downstream analysis.
633 However, for most datasets, we needed to download the raw FASTQ files and calculated the
634 expression matrices ourselves. Specifically, we used cellranger (Zheng et al. 2017) (v3.02) to
635 map the reads, and then used velocyto (La Manno et al. 2018) (v0.17.17) or UMI-tools (Smith
636 et al. 2017) (v1.01) to quantify the unspliced mRNA and spliced mRNA levels. After obtaining
637 the expression matrices, Seurat (Stuart et al. 2019) was used for performing quality control (to
638 remove low-quality cells) and for normalizing expression matrices (LogNormalize).

639 640 *Network inference procedure*

641 Expression matrices were fed into GENIE3 for GRN inference. To enhance accuracy, we utilized
642 the TF information (from RcisTarget (Aibar et al. 2017)) to label regulators and only regulatory
643 relationships from one TF to another gene (including other TFs) were considered for inference.
644 Because of the sparsity of scRNA-seq data, only genes expressed in more than 10% of cells
645 were included for analysis. For the mRNA-based method, only spliced mRNA matrices were
646 used to infer the network, and for the pre-mRNA-based method, both unspliced pre-mRNA
647 and spliced mRNA matrices were used.

648

649 *Ground-truth for network evaluation*

650 We mainly used two databases as ground-truth GRNs for evaluating GRNs inferred from
651 experimental datasets: 1) the GRN from DoRothEA (Garcia-Alonso et al. 2019), with only edges
652 with confidence A and B included; 2) the GRN from RcisTarget (Aibar et al. 2017) (i.e.,
653 RcisTarget was used to identify enriched TF-binding motifs for all genes and to obtain
654 candidate TFs). The rationale for choosing these two ground-truth GRNs was that they are
655 complementary, i.e., the regulatory links in DoRothEA have high confidence but low coverage,
656 while the regulatory links in RcisTarget have high coverage but low confidence. Besides these
657 two ground-truth GRNs, we also used cell type-specific ChIP-seq for Neurod2 (Bayam et al.
658 2015) to evaluate the network inferred from the human forebrain dataset (**Fig. 5C**).
659

660 *Metrics for evaluating network inference results*

661 We used AUPR (area under the precision-recall curve) and AEP (average early precision) to
662 evaluate the quality of the inferred GRNs, whereby AUPR has been widely used. We proposed
663 to use AEP for networks inferred from typical single-cell RNA data, as explained below. More
664 specifically, when comparing inferred GRN with ground-truth GRN, each link can be assigned
665 into one of the four categories: TP (true positive), FP (false positive), TN (true negative), and
666 FN (false negative). The precision and recall were defined as following: precision = TP/(TP+FP),
667 recall = TP/(TP+FN). Because the output of the inferred GRN is a ranked list of potential
668 regulatory links, a threshold should be chosen to determine the inference results. For each
669 threshold, there is a pair of precision and recall values, and thus a curve can be plotted in the
670 precision-recall plane. The area under the precision-recall curve is defined as AUPR. In this

work, AUPR was adopted to evaluate simulated single-cell datasets. As for networks inferred from experimental single-cell datasets, although we also used AUPR, we believed that AEP is a more appropriate and informative metric, which focuses on the most confident links in the inferred network, analogous to the Early Precision Ratio (EPR) used in a recent network inference evaluation study (Pratapa et al. 2020). Compared to EPR, which computes the ratio of the fraction of true positives in the top-k edges over the early precision for a random predictor, AEP measures the average precision in the top network. The top network includes links that have confidence levels ranked at a top threshold percentage, where the threshold is set to be 10% or 1% when evaluated by DoRothEA or Motif GRN respectively. The rationale for using AEP is that, by using the average, instead of a single data point as in EPR, AEP could, in principle, provide a more robust measure of early precision.

Factor-dependency analysis

In the networks inferred from four representative experimental datasets, we analyzed the effect of mRNA half-life, the expression level of mRNA or pre-mRNA, and TF dynamics on network inference accuracy by using Motif GRN as the ground-truth (**Fig. 6**). We calculated the top precision for each target gene, i.e., mean inference precision of the top 10 inferred TFs of each target, and divided all target genes into two groups according to mRNA half-life or expression level. We then compared the top precision of the two groups and calculated the p-values using Wilcoxon rank sum tests. The mRNA half-life data was from TimeLapse-seq in K562 cells (Schofield et al. 2018). For the effect of TF dynamics, we used the accessibility variability data (Buenrostro et al. 2015) to approximate the degree of TF dynamics, and the average variabilities of 8 wild-type cells were used. More specifically, regarding the estimation of TF dynamics, Buenrostro et al (Buenrostro et al. 2015) first estimated the activity of TFs in individual cells based on the accessibility of the motifs bound by these TFs (derived from single-cell ATAC-seq data), and then estimated the dynamics of TFs through evaluating the cell-to-cell variability in TF activity. We sorted all TFs by variability score and calculated the fraction of TFs (that are above each variability score) whose performance is better with the pre-mRNA-based method than with the mRNA-based method. The rationale is that by focusing on TFs that are more dynamic, we could ask whether the pre-mRNA-based method would be more

701 advantageous in the presence of dynamic regulatory activities, as demonstrated in the kinetic
702 model and simulated datasets (**Supplemental Figs. S1E, S4C**). We also further investigated the
703 human forebrain dataset using cell type-specific TF variability data (Trevino et al. 2020).

704

705

706 **Software availability**

707 The source codes of this study are freely available through GitHub
708 (<https://github.com/LingfengXue1999/NetworkInference>) and as Supplemental Code. The
709 associated datasets used for GRN inference are available through Figshare
710 (<https://figshare.com/articles/dataset/Datasets/21342117> or
711 <https://doi.org/10.6084/m9.figshare.21342117.v1>).

712

713 **Acknowledgments**

714 The authors would like to thank the High-Performance Computing Platform of the Peking-
715 Tsinghua Center for Life Sciences for computational support. Y.L. acknowledges the supports
716 from National Key R&D Program of China (Grant # 2020YFA0906900, and 2018YFA0900703)
717 and National Natural Science Foundation of China (Grant # 32088101, and 31771425).

718

719 **Author contributions**

720 L.X.: Conceptualization, Investigation, Methodology, Software, Writing – Original Draft
721 Preparation, Writing – Review & Editing; W.Y.: Methodology, Writing – Review & Editing; Y.L.:
722 Conceptualization, Funding Acquisition, Supervision, Writing – Original Draft Preparation,
723 Writing – Review & Editing

724

725 **Competing interests**

726 None of the authors have any competing interests.

727

728 **References**

- 729 Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F,
 730 Marine J-C, Geurts P, Aerts J, et al. 2017. SCENIC: single-cell regulatory network inference
 731 and clustering. *Nature Methods* **14**: 1083–1086.
- 732 Alon U. 2006. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman
 733 and Hall/CRC, New York.
- 734 Argelaguet R, Cuomo ASE, Stegle O, Marioni JC. 2021. Computational principles and challenges in
 735 single-cell data integration. *Nat Biotechnol* **39**: 1202–1215.
- 736 Bayam E, Sahin GS, Guzelsoy G, Guner G, Kabakcioglu A, Ince-Dunn G. 2015. Genome-wide target
 737 analysis of NEUROD2 provides new insights into regulation of cortical projection neuron
 738 migration and differentiation. *BMC Genomics* **16**: 681.
- 739 Bendall SC, Davis KL, Amir ED, Tadmor MD, Simonds EF, Chen TJ, Shenfeld DK, Nolan GP, Pe'er D.
 740 2014. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination
 741 in Human B Cell Development. *Cell* **157**: 714–725.
- 742 Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ.
 743 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*
 744 **523**: 486–490.
- 745 Cannoodt R, Saelens W, Deconinck L, Saeys Y. 2021. Spearheading future omics analyses using
 746 dyngen, a multi-modal simulator of single cells. *Nat Commun* **12**: 3942.
- 747 Cao J, Zhou W, Steemers F, Trapnell C, Shendure J. 2020. Sci-fate characterizes the dynamics of gene
 748 expression in single cells. *Nature Biotechnology* **1**–9.
- 749 Chan TE, Stumpf MPH, Babtie AC. 2017. Gene Regulatory Network Inference from Single-Cell Data
 750 Using Multivariate Information Measures. *Cell Systems* **5**: 251–267.e3.
- 751 Davidson EH. 2010. Emerging properties of animal gene regulatory networks. *Nature* **468**: 911–920.
- 752 Deshpande A, Chu L-F, Stewart R, Gitter A. 2022. Network inference with Granger causality
 753 ensembles on single-cell transcriptomics. *Cell Reports* **38**: 110333.
- 754 D'haeseleer P, Liang S, Somogyi R. 2000. Genetic network inference: from co-expression clustering
 755 to reverse engineering. *Bioinformatics* **16**: 707–726.
- 756 Efremova M, Teichmann SA. 2020. Computational methods for single-cell omics across modalities.
 757 *Nat Methods* **17**: 14–17.
- 758 Ferrell JE. 2012. Bistability, Bifurcations, and Waddington's Epigenetic Landscape. *Current Biology*
 759 **22**: R458–R466.
- 760 Fiers MWEJ, Minnoye L, Aibar S, Bravo González-Blas C, Kalender Atak Z, Aerts S. 2018. Mapping
 761 gene regulatory networks from single-cell omics data. *Briefings in Functional Genomics* **17**:
 762 246–254.
- 763 Friedman N. 2004. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science* **303**:
 764 799–805.
- 765 Gaidatzis D, Burger L, Florescu M, Stadler MB. 2015. Analysis of intronic and exonic reads in RNA-
 766 seq data characterizes transcriptional and post-transcriptional regulation. *Nature*
 767 *Biotechnology* **33**: 722–729.
- 768 Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. 2019. Benchmark and
 769 integration of resources for the estimation of human transcription factor activities.
 770 *Genome Res* **29**: 1363–1375.
- 771 Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J,
 772 Alexander R, et al. 2012. Architecture of the human regulatory network derived from
 773 ENCODE data. *Nature* **489**: 91–100.
- 774 Gupta A, Martin-Rufino JD, Jones TR, Subramanian V, Qiu X, Grody EE, Bloemendaal A, Weng C, Niu
 775 S-Y, Min KH, et al. 2022. Inferring gene regulation from stochastic transcriptional variation
 776 across single cells at steady state. *Proceedings of the National Academy of Sciences* **119**:
 777 e2207392119.
- 778 Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M,
 779 et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* **184**: 3573–3587.e29.
- 780 Haury A-C, Mordelet F, Vera-Licona P, Vert J-P. 2012. TIGRESS: Trustful Inference of Gene REgulation
 781 using Stability Selection. *BMC Systems Biology* **6**: 145.
- 782 Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R. 2009. Gene regulatory network

- 783 inference: Data integration in dynamic models—A review. *Biosystems* **96**: 86–103.
- 784 Hersbach BA, Fischer DS, Masserdotti G, Deeksha null, Mojžišová K, Waltzhöni T, Rodriguez-
- 785 Terrones D, Heinig M, Theis FJ, Götz M, et al. 2022. Probing cell identity hierarchies by fate
- 786 titration and collision during direct reprogramming. *Mol Syst Biol* **18**: e11129.
- 787 Huang S, Ernberg I, Kauffman S. 2009. Cancer attractors: A systems view of tumors from a gene
- 788 network dynamics and developmental perspective. *Seminars in Cell & Developmental*
- 789 *Biology* **20**: 869–876.
- 790 Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. 2010. Inferring Regulatory Networks from
- 791 Expression Data Using Tree-Based Methods. *PLOS ONE* **5**: e12776.
- 792 Iacono G, Massoni-Badosa R, Heyn H. 2019. Single-cell transcriptomics unveils gene regulatory
- 793 network plasticity. *Genome Biology* **20**: 110.
- 794 Joung J, Ma S, Tay T, Geiger-Schuller KR, Kirchgatterer PC, Verdine VK, Guo B, Arias-Garcia MA, Allen
- 795 WE, Singh A, et al. 2023. A transcription factor atlas of directed differentiation. *Cell* **186**:
- 796 209–229.e26.
- 797 La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME,
- 798 Lönnberg P, Furlan A, et al. 2018. RNA velocity of single cells. *Nature* **560**: 494–498.
- 799 Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT,
- 800 Thompson CM, Simon I, et al. 2002. Transcriptional Regulatory Networks in *Saccharomyces*
- 801 *cerevisiae*. *Science* **298**: 799–804.
- 802 Levine JH, Lin Y, Elowitz MB. 2013. Functional Roles of Pulsing in Genetic Circuits. *Science* **342**:
- 803 1193–1200.
- 804 Levine M, Davidson EH. 2005. Gene regulatory networks for development. *Proceedings of the*
- 805 *National Academy of Sciences* **102**: 4936–4942.
- 806 MacArthur BD, Ma'ayan A, Lemischka IR. 2009. Systems biology of stem cell fate and cellular
- 807 reprogramming. *Nature Reviews Molecular Cell Biology* **10**: 672–681.
- 808 Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ,
- 809 Stolovitzky G. 2012. Wisdom of crowds for robust gene network inference. *Nature Methods* **9**: 796–804.
- 810 Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G. 2010. Revealing strengths
- 811 and weaknesses of methods for gene network inference. *Proceedings of the National*
- 812 *Academy of Sciences* **107**: 6286–6291.
- 813 Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, Califano A. 2006. ARACNE:
- 814 An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian
- 815 Cellular Context. *BMC Bioinformatics* **7**: S7.
- 816 Matsumoto H, Kiryu H, Furusawa C, Ko MSH, Ko SBH, Gouda N, Hayashi T, Nikaido I. 2017. SCODE:
- 817 an efficient regulatory network inference algorithm from single-cell RNA-Seq during
- 818 differentiation. *Bioinformatics* **33**: 2314–2321.
- 819 Nguyen H, Tran D, Tran B, Pehlivan B, Nguyen T. 2021. A comprehensive survey of regulatory
- 820 network inference methods using single cell RNA sequencing data. *Briefings in*
- 821 *Bioinformatics* **22**. <https://doi.org/10.1093/bib/bbaa190> (Accessed August 22, 2021).
- 822 Petrany MJ, Swoboda CO, Sun C, Chetal K, Chen X, Weirauch MT, Salomonis N, Millay DP. 2020.
- 823 Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal
- 824 myofibers. *Nat Commun* **11**: 6374.
- 825 Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. 2020. Benchmarking algorithms for gene
- 826 regulatory network inference from single-cell transcriptomic data. *Nat Methods* **17**: 147–
- 827 154.
- 828 Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell
- 829 C, Kannan S. 2020. Inferring Causal Gene Regulatory Networks from Coupled Single-Cell
- 830 Expression Dynamics Using Scribe. *Cell Systems* **10**: 265–274.e11.
- 831 Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, Pogson AN, Hein MY, Hoi (Joseph)
- 832 Min K, Wang L, et al. 2022. Mapping transcriptomic vector fields of single cells. *Cell* **185**:
- 833 690–711.e45.
- 834 Quinn TP, Richardson MF, Lovell D, Crowley TM. 2017. propr: An R-package for Identifying
- 835 Proportionally Abundant Features Using Compositional Data Analysis. *Sci Rep* **7**: 16252.

- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org>.
- Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, Hacohen N, Schier AF, Blackshear PJ, Friedman N, et al. 2014. High-Resolution Sequencing and Modeling Identifies Distinct Dynamic RNA Regulatory Strategies. *Cell* **159**: 1698–1710.
- Schofield JA, Duffy EE, Kiefer L, Sullivan MC, Simon MD. 2018. TimeLapse-seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* **15**: 221–225.
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34**: 166–176.
- Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499.
- Sonawane AR, Platig J, Fagny M, Chen C-Y, Paulson JN, Lopes-Ramos CM, DeMeo DL, Quackenbush J, Glass K, Kuijjer ML. 2017. Understanding Tissue-Specific Gene Regulation. *Cell Reports* **21**: 1077–1088.
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**: 1888-1902.e21.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. 1999. Systematic determination of genetic network architecture. *Nat Genet* **22**: 281–285.
- Thompson D, Regev A, Roy S. 2015. Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. *Annual Review of Cell and Developmental Biology* **31**: 399–428.
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32**: 381–386.
- Trevino AE, Sinnott-Armstrong N, Andersen J, Yoon S-J, Huber N, Pritchard JK, Chang HY, Greenleaf WJ, Paşa SP. 2020. Chromatin accessibility dynamics in a model of human forebrain development. *Science* **367**: eaay1645.
- Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W, Cannoodt R, Rouchon Q, et al. 2020. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**: 2247–2276.
- Waddington CH. 1957. *The Strategy of the Genes*. Routledge, London.
- Welch JD, Hartemink AJ, Prins JF. 2017. MATCHER: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics. *Genome Biology* **18**: 138.
- Wu Y, Xue L, Huang W, Deng M, Lin Y. 2022. Profiling transcription factor activity dynamics using intronic reads in time-series transcriptome data. *PLOS Computational Biology* **18**: e1009762.
- Yeung MKS, Tegnér J, Collins JJ. 2002. Reverse engineering gene networks using singular value decomposition and robust regression. *PNAS* **99**: 6163–6168.
- Yuan Y, Bar-Joseph Z. 2019. Deep learning for inferring gene relationships from single-cell expression data. *Proceedings of the National Academy of Sciences* **116**: 27151–27158.
- Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**: 1–12.

Figure Legends**Figure 1. Schematics of gene regulation and regulatory inference methods.**

(A) Schematic diagram of molecular species involved in transcriptional regulation.

(B) Schematics comparing conventional and proposed methods for gene regulatory network (GRN) inference. Conventionally, mRNA levels of transcription factors and target genes from scRNA-seq data are used to infer GRN (i.e., the mRNA-based method). In contrast, we propose to use mRNA levels of transcription factors and pre-mRNA levels of target genes for inferring GRN (i.e., the pre-mRNA-based method).

Figure 2. Using kinetic modeling to compare pre-mRNA-based and mRNA-based GRN inference methods.

(A) The inference of regulatory activity, $g(t)$, using target gene expression level (either pre-mRNA level or mRNA level) can be affected by various factors. These factors include the dynamics of the regulatory activity (i.e., T_{on}), the transcription rate of the target gene (α), the splicing rate of the pre-mRNA (β), and the degradation rate of the mRNA (γ).

(B) Heatmaps showing the relative accuracy between pre-mRNA-based and mRNA-based inference of the regulatory activity in the absence of stochasticity. The model (**Supplemental Figs. S1A**) was analytically solved using indicated parameter combinations and the relative inference accuracies between pre-mRNA-based and mRNA-based methods were calculated (Methods). The redder the color, the more accurate the pre-mRNA-based method is.

(C) Heatmap showing the relative accuracy between pre-mRNA-based and mRNA-based inference of the regulatory activity in the presence of stochasticity. The model (**Supplemental Figs. S1A**) was stochastically solved using indicated parameter combinations and the relative inference accuracies between pre-mRNA-based and mRNA-based methods were calculated. Note that under low gene expression levels and slow regulatory dynamics, the pre-mRNA-based method underperforms compared to the mRNA-based method.

Figure 3. Using simulated single-cell datasets to compare pre-mRNA-based and mRNA-based methods.

917 **(A)** Schematic illustration of pipelines used for the generation of simulated datasets, the
918 inference of the GRN, the evaluation of inferred GRN, and factor-dependency analysis.
919 Simulated datasets were generated using the dyngen package, whereby different network
920 backbones and kinetic parameters were used for performing dynamic simulations. The output
921 count matrices were then used for network inference using GENIE3. The accuracies of the
922 inferred networks were then calculated and used for factor-dependency analysis. See
923 **Methods** for details.

924 **(B)** Boxplots comparing the performance of pre-mRNA-based and mRNA-based methods in
925 four different network backbones. The performance was determined by the accuracy of the
926 inferred network, measured by AUPR (area under precision-recall curve). Meanwhile, AUPR
927 for a random predictor is included for comparison. N = 20.

928 **(C)** Factor-dependency analysis for pre-mRNA-based and mRNA-based methods. Simulations
929 were performed under parameter ranges of three gene-level factors (transcription rate, mRNA
930 half-life, and protein half-life). The effect of each factor on the inference accuracies of different
931 network backbones was evaluated using AUPR ratio, calculated as the ratio of the AUPR under
932 the largest parameter value to the AUPR under the lowest parameter value (see also
933 **Supplemental Fig. S4A-C**). Black dashed lines indicate AUPR ratio of 1 (i.e., no effect of the
934 parameter choice on inference accuracy).

935

936 **Figure 4. Using experimental single-cell datasets to compare pre-mRNA-based and mRNA-**
937 **based methods.**

938 **(A)** Comparison between pre-mRNA-based and mRNA-based methods for 30 scRNA-seq
939 datasets of human and mouse cells using DoRothEA GRN as the ground truth. The
940 performance was measured by average early precision (**Methods**). Precision for a random
941 predictor is shown for comparison.

942 **(B)** Analogus to **(A)**, except that Motif GRN was used as the ground truth.

943 **(C)** Scatter plots showing the comparison between pre-mRNA-based and mRNA-based
944 methods for 30 scRNA-seq datasets of human and mouse cells using DoRothEA GRN as the
945 ground truth. Diagonal line denotes equal performance. Dashed lines denote precision value
946 at 0.15.

947 **(D)** Scatter plots analogous to **(A)**, except that Motif GRN was used as the ground truth.

948 **(E-F)** Analogous to **(A-B)** except that AUPR was used as the metric.

949
950 **Figure 5. Using the forebrain datasets to demonstrate the improved performance of the pre-**
951 **mRNA-based method.**

952 **(A)** Precision-recall curves for GRNs inferred by pre-mRNA-based and mRNA-based methods
953 for the human forebrain dataset. The ground-truth network is DoRothEA GRN.

954 **(B)** Inferred GRN for the human forebrain dataset using the pre-mRNA-based method. In this
955 network, the edges represent the inferred transcriptional regulation from one TF (transcription
956 factor) to one target gene. The size of the node represents the number of inferred target genes
957 for the TF. 300 interactions (edges) of the highest confidence were shown (i.e., Top300
958 network).

959 **(C)** Evaluation of the hub TFs in the pre-mRNA-based Top300 network using different types of
960 ground-truth network. Note that because the DoRothEA ground-truth GRN contains much
961 fewer nodes and edges compared to the Motif ground-truth GRN, only one of the hub TFs
962 could be evaluated using DoRothEA GRN while all of them could be evaluated with the Motif
963 GRN.

964 **(D)** Ground-truth-free comparison between GRNs inferred by pre-mRNA-based and mRNA-
965 based methods using cross-species target overlap. In particular, GRNs from the human
966 forebrain dataset (hFB) and the mouse forebrain (mFB) were evaluated, whereby the overlap
967 ratio of top 500 targets in two networks for each TF was calculated. Such overlap ratio
968 represents the cross-species similarity between the two networks, and was compared
969 between pre-mRNA-based and mRNA-based GRNs. N = 372 TFs and p-value = 5×10^{-9} from
970 Wilcoxon test. Grey horizontal line indicates random overlap ratio.

971
972 **Figure 6. Factor-dependency analysis for GRNs inferred from experimental datasets.**

973 **(A)** Schematic diagram of the factor-dependency analysis. Three factors that affect the
974 accuracy of the inferred network were analyzed, i.e., the dynamics of the TF, the expression
975 level of the target gene (pre-mRNA or mRNA), and the mRNA half-life of the target gene.

976 **(B)** The dependency of the network inference accuracy on mRNA (or pre-mRNA) expression

977 levels of target genes in four datasets. The dependency was quantified by using the p-value as
978 in **Supplemental Fig. S7A**, calculated by comparing the inference accuracies between targets
979 of low or high expression levels for each network. Horizontal dashed line indicates $p = 0.05$.

980 **(C)** Analogous to **(B)** for the mRNA half-life of the target gene.

981 **(D)** The dependency of the network inference accuracy on the TF dynamics. The dynamics of
982 TFs were approximated by using cell-to-cell variabilities of TF activities measured by public
983 single-cell ATAC-seq data (**Methods**). TFs from all four datasets were sorted according to TF
984 dynamics, and the fraction of TFs more accurately inferred by the pre-mRNA-based method
985 than by the mRNA-based method was calculated for TFs above the indicated value of TF
986 dynamics on x-axis (see also **Supplemental Fig. S7B-C**).

Figure 1

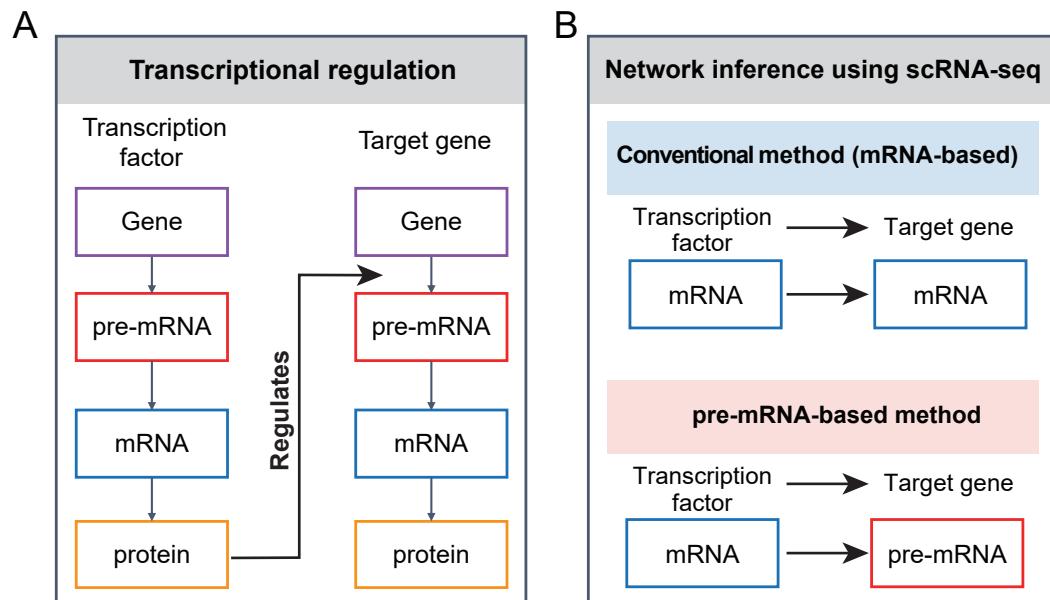
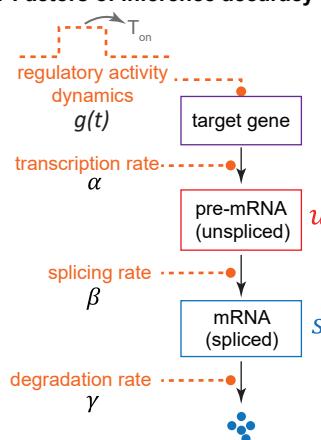
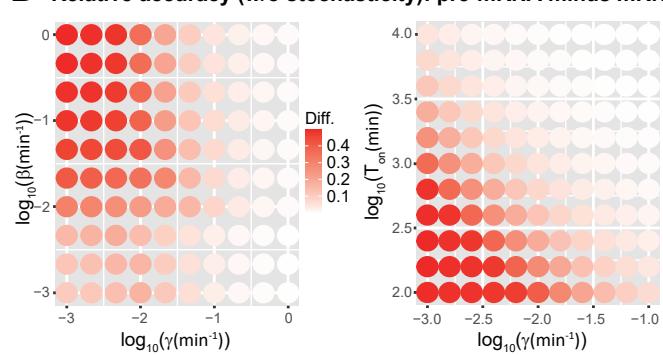


Figure 2

A Factors of inference accuracy



B Relative accuracy (w/o stochasticity): pre-mRNA minus mRNA



C Relative accuracy (with stochasticity)

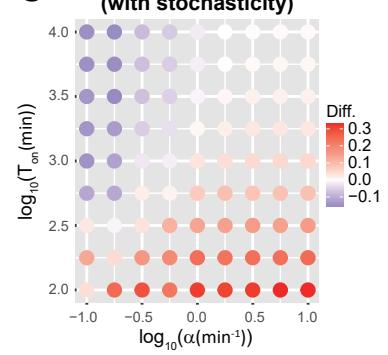


Figure 3

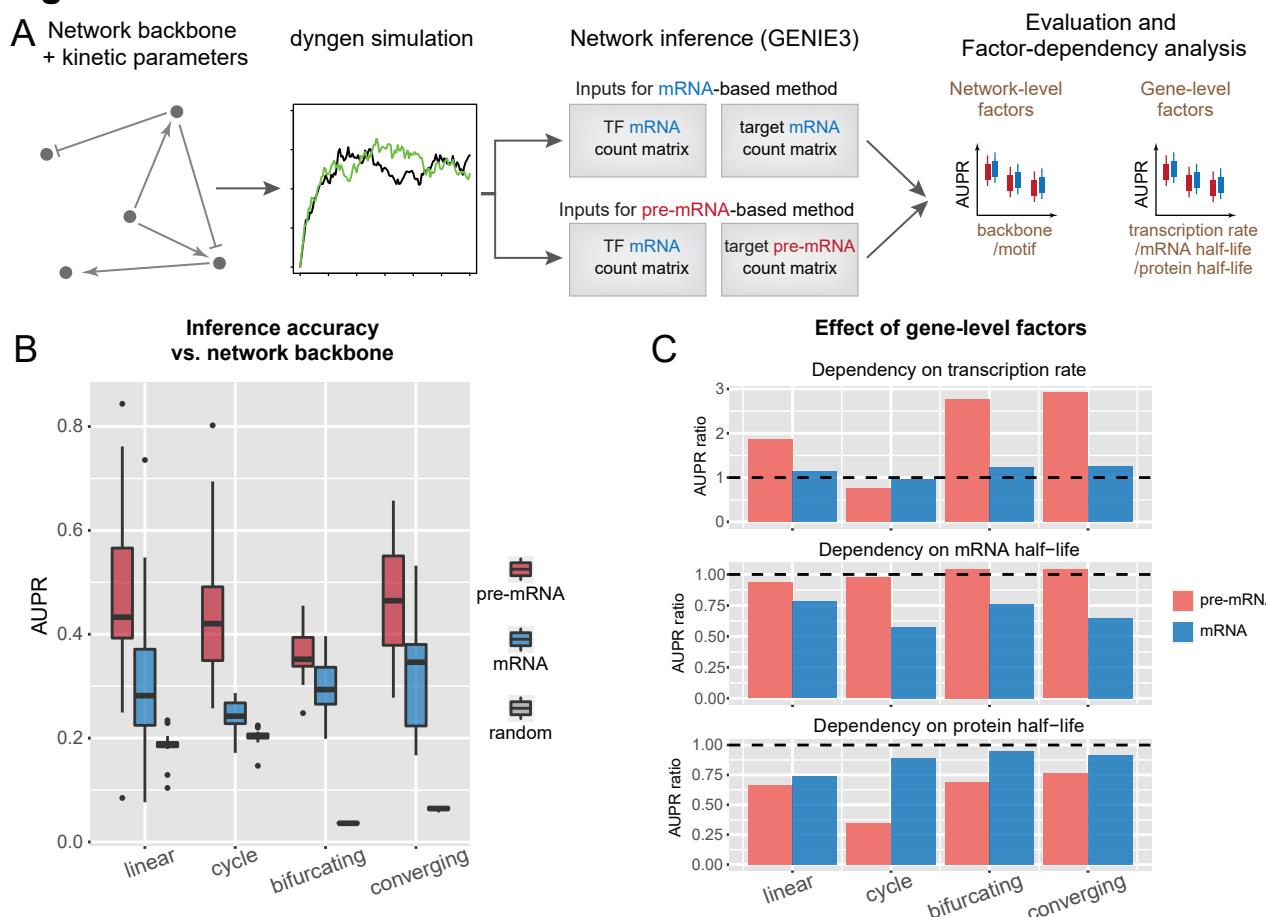


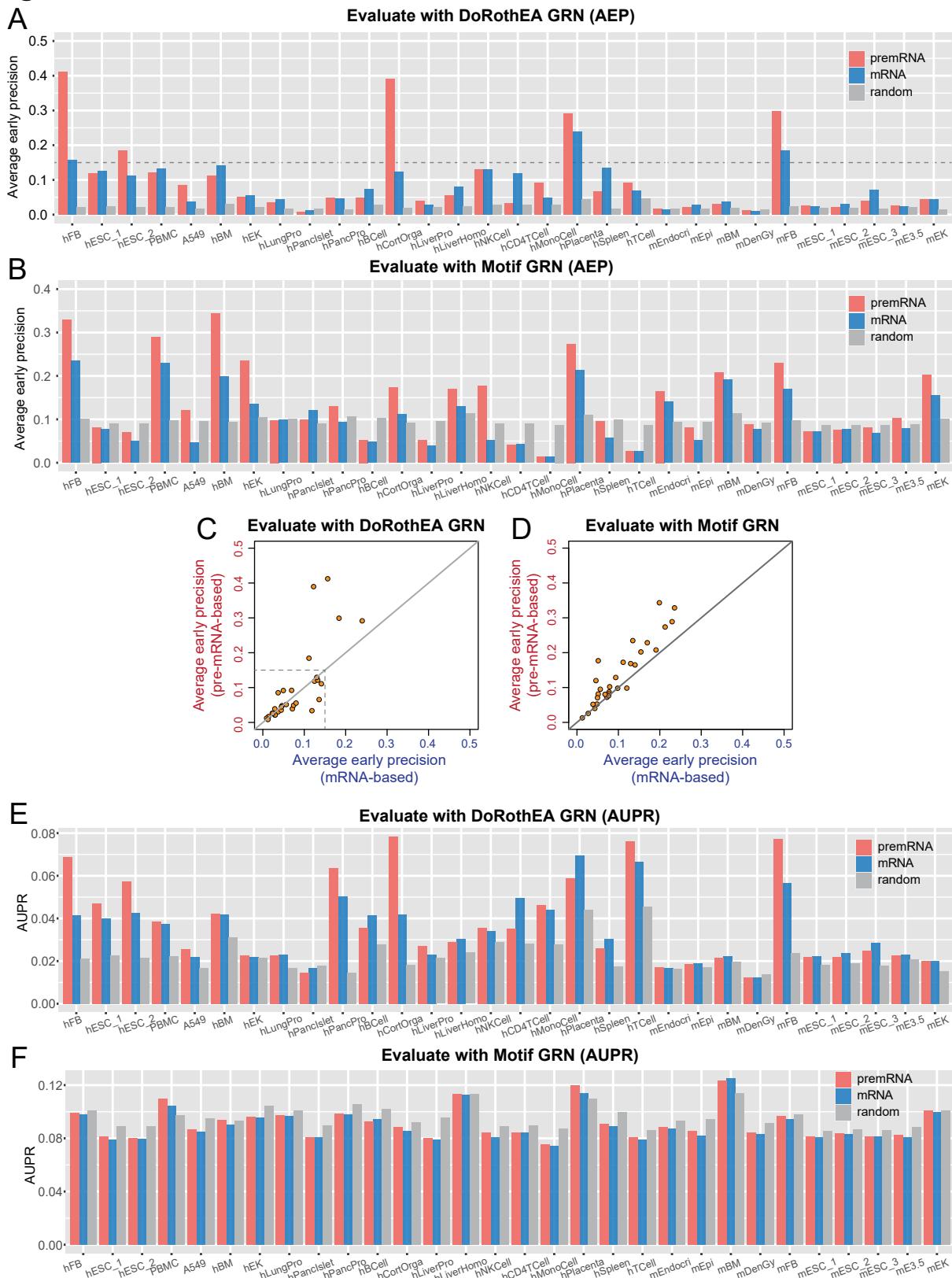
Figure 4

Figure 5

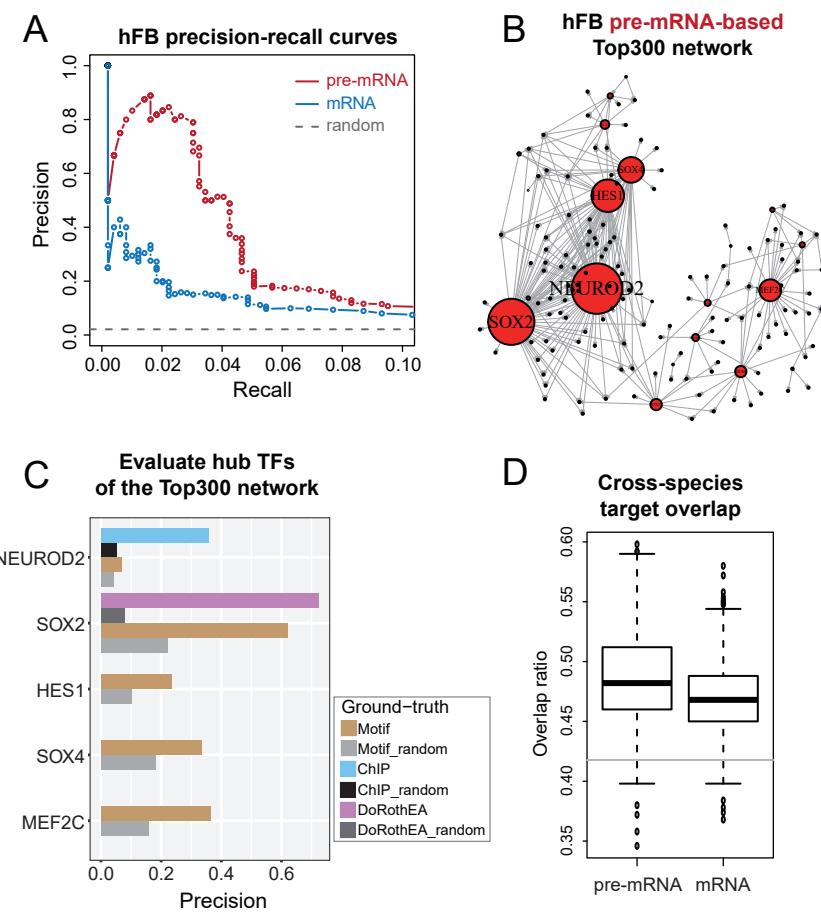
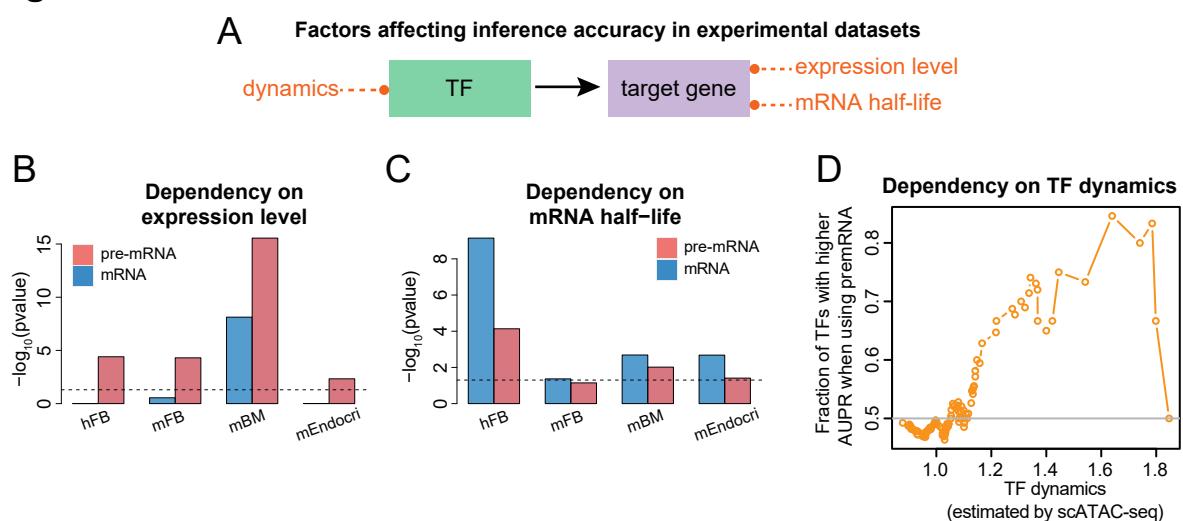


Figure 6





Dissecting and improving gene regulatory network inference using single-cell transcriptome data

Lingfeng Xue, Yan Wu and Yihan Lin

Genome Res. published online August 14, 2023

Access the most recent version at doi:[10.1101/gr.277488.122](https://doi.org/10.1101/gr.277488.122)

P<P Published online August 14, 2023 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Doing science doesn't
have to be wasteful.

USA
SCIENTIFIC

[LEARN MORE](#)

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
