Check for updates

METHOD ARTICLE

Improving prediction of core transcription factors for cell reprogramming and transdifferentiation [version 1; peer review: 2 not approved]

Mikhail Raevskiy ^{1,2}, Anna Kondrashina¹, Yulia Medvedeva^{1,3,4}

¹Department of Biological and Medical Physics, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, 141701, Russian Federation

²Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

³Institute of Bioengineering, Research Center of Biotechnology, Moscow, 119071, Russian Federation

⁴National Medical Research Center for Endocrinology, Moscow, 115478, Russian Federation

v1 First published: 13 Jan 2022, 11:38
<https://doi.org/10.12688/f1000research.75321.1>

Latest published: 13 Jan 2022, 11:38
<https://doi.org/10.12688/f1000research.75321.1>

Abstract

Identification of transcription factors (TFs) that could induce and direct cell conversion remains a challenge. Though several hundreds of TFs are usually transcribed in each cell type, the identity of a cell is controlled and can be achieved through the ectopic overexpression of only a small subset of so-called core TFs. Currently, the experimental identification of the core TFs for a broad spectrum of cell types remains challenging. Computational solutions to this problem would provide a better understanding of the mechanisms controlling cell identity during natural embryonic or malignant development, as well as give a foundation for cell-based therapy. Herein, we propose a computational approach based on over-enrichment of transcription factors binding sites (TFBS) in differentially accessible chromatin regions that could identify the potential core TFs for a variety of primary human cells involved in hematopoiesis. Our approach enables the integration of both transcriptomic (single-cell RNA sequencing, scRNA-seq) and epigenomic (single-cell assay for transposable-accessible chromatin, scATAC-seq) data at the single-cell resolution to search for core TFs, and can be scalable to predict subsets of core TFs and their role in a given conversion between cells.

Keywords

cell conversion, scRNA-seq, scATAC-seq, transcription factors, epigenetics

Open Peer Review

Approval Status

	1	2
version 1
13 Jan 2022	view	view
<p>1. Valentiva Boeva, ETH Zurich, Zurich, Switzerland</p> <p>Samuel Gunz, ETH Zurich, Zurich, Switzerland</p> <p>2. Erdem B. Dashinimaev , Russian Academy of Sciences, Moscow, Russian Federation</p> <p>Pirogov Russian National Research Medical University, Moscow, Russian Federation</p> <p>Any reports and responses or comments on the article can be found at the end of the article.</p>

This article is included in the **Bioinformatics** gateway.

This article is included in the **Cell & Molecular Biology** gateway.

Corresponding authors: Mikhail Raevskiy (raevskymichail@gmail.com), Yulia Medvedeva (ju.medvedeva@gmail.com)

Author roles: Raevskiy M: Formal Analysis, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; Kondrashina A: Software, Validation, Visualization; Medvedeva Y: Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was supported by Ministry of Science and Higher Education of the Russian Federation (agreement no.075- 15-2020-899).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Raevskiy M *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Raevskiy M, Kondrashina A and Medvedeva Y. **Improving prediction of core transcription factors for cell reprogramming and transdifferentiation [version 1; peer review: 2 not approved]** F1000Research 2022, 11:38 <https://doi.org/10.12688/f1000research.75321.1>

First published: 13 Jan 2022, 11:38 <https://doi.org/10.12688/f1000research.75321.1>

Introduction

The cell identity is largely controlled by transcription factors (TFs). TFs regulate gene expression by binding DNA in a sequence-specific manner, targeting short sequences called transcription factor binding sites (TFBS). Although almost half of all TFs are expressed in a particular cell type,²¹ only a minor share of these TFs — so-called core TFs — are sufficient to maintain cell identity by defining the corresponding gene expression programs.^{11,22,23} The identification of core TFs for a large number of cell types would be a valuable addition for an atlas of transcription regulators supplementing the Encyclopedia of Regulatory DNA Elements (ENCODE, Ref. 16). Such an atlas, in turn, would facilitate systematic investigation of regulatory networks and contribute to establishing and refining direct cell conversion protocols for clinically relevant cell types.^{6,7}

Systematic determination of core TFs controlling individual cell type identity has previously been attempted. Initial efforts were mainly focused on the experimental screening of the TFs, presumably regulating the differentially expressed genes (DEGs) in the comparison between query cell type, and a small number of alternative cell types that could potentially serve as an initial stage for conversion. Some of these TFs could play a role as regulators controlling cellular identities. For example, studies showed that over-expression of *MyoD1* in fibroblasts leads to its conversion into the muscle cells,¹⁹ while inhibition of *Oct4* resulted in the suppression of the pluripotent stem cell population during mammalian embryo development.¹² Recent experiments with TF over-expression leading to conversion of cells to another cell type appeared to be used as a stringent test of the potential of specific TFs to establish and maintain cell identity.^{11,22,23} Nonetheless, while being illustrative validation for each TF, such experiments are still time- and labor-consuming, and resulting observations are limited to specific cell types.

The growth of genome-wide sequencing technologies allowed to develop computational systems capable of predicting candidate core TFs.^{2,9,14,17} However, being broad in scope and easily scalable, these methods infer predictions using preferably only bulk RNA sequencing (RNA-seq) data, which estimates the average gene expression level across a hundred thousands to millions of cells. As a result, they are insufficient for analysis of heterogeneous systems, such as early embryonic populations or complex tissues, including brain or bone marrow.

Here we propose an approach that uses single-cell expression and DNA accessibility data to select core TFs for cell differentiation or directed conversion. A distinct feature of the approach is incorporating not only TFs expression levels in the original and target cell types, but also (1) the chromatin conditions in gene regulatory elements, as well as (2) TF putative binding sites. Thus, this method simultaneously takes into account the accessibility and expression profile of the initial and terminal cell types involved in the conversion. Additionally, our method uses modified gene set enrichment analysis (GSEA)¹⁸ for the selection of core TFs, thus reducing the number of arbitrary thresholds in the pipeline.

Results

To validate our method, we applied it to hematopoietic differentiation datasets,^{5,1} since this process has been extensively studied. We provided TFs for the hematopoietic stem cells (HSC) differentiation into CD4(+) cells as an example (Table 1). The detected TFs are critical for the HSC-to-CD4(+) cells differentiation. The top-ranked TF, TCF7, is a transcription activator recruited in T-cell lymphocyte differentiation and is necessary for the survival of immature CD4(+) and CD8(+) thymocytes.^{13,10} RORA gene plays a crucial role in the regulation of embryonic development, differentiation and immunity.⁴ TBX21 is a lineage-defining TF, which initiates Th1(CD4(+)) lineage development from naive T helper (CD4(+)) precursor cells.^{24,10} The LEF1 TF has a higher affinity to a functionally important site in the T-cell receptor-alpha enhancer, and thereby its presence in these regions increases the activity of the enhancer.³

Methods

The proposed approach (Figure 1) consists of the following steps. First, for two given cell types involved in cell differentiation or conversion pathways, the minimal spanning tree (MST) is reconstructed based on the open chromatin in

Table 1. A predicted list of transcription factors for HSC to CD4(+) lymphocytes differentiation.

HGNC gene	GSEA <i>p-val</i>	GSEA <i>q-val</i>
TCF7	4.36×10^{-33}	1.50×10^{-31}
RORA	6.98×10^{-31}	2.17×10^{-29}
NR1D1	1.83×10^{-29}	5.29×10^{-28}
TBX21	1.57×10^{-8}	6.20×10^{-8}
LEF1	2.53×10^{-7}	8.24×10^{-7}

Figure 1. Schematic overview of the proposed approach within the typical pipeline of TFs selection.

regulatory regions (Figure 2, Figure 3). Then, a differential accessibility analysis (DAA) between initial and final cell types is performed to retrieve a list of genomic regions (ATAC-seq peaks) ranked by the statistical significance of a change in chromatin accessibility for a given cell conversion (Figures 4, 5). Next, the sequences corresponding to each of the ranked regions undergo the functional annotation with TFBS. Finally, TFs ranking is inferred by GSEA,¹⁸ which was adjusted to estimate the tendency of TFBS for given TF under investigation to be over-represented at the most statistically significant genomic regions for a given cell differentiation or conversion.

Reconstruction of cell trajectories with scATAC-seq data

scATAC-seq data (GEO: (GSE96769, GSE111586)) were used to reconstruct the minimal spanning tree (MST) of hematopoietic cell types, the hierarchy of which was aligned along pseudo-time, reflecting a degree of pluripotency of the cells observed in the single-cell assay for transposable-accessible chromatin (scATAC-seq) dataset.¹⁵ Thus, the obtained MST presents a collection of possible cell trajectories among the analyzed cell types.

Differential accessibility analysis

Similarly to DEG analysis,²⁰ a differential accessibility analysis (DAA) of genomic regions was performed between two given cell types on the cell trajectory by <https://www.bioconductor.org/packages/devel/bioc/manuals/slingshot/man/slingshot.pdf> Slingshot v2.3. Accordingly, for each cell population on the MST, such a subset of regions ranked by p-value can be obtained, discriminating given cell population from others.

Figure 2. The Minimal Spanning Tree (MST), reconstructed on scATAC-seq data (GSE74912) for the system of the 8 hematopoietic cell types. HSC, hematopoietic stem cells; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; NK, natural killer cells.

TFs filtration and TFBS annotation

We excluded from the downstream analysis TFs that had either a near-zero median expression (below 5% percentile) in the final cell type or had a higher expression in the original cell types based on scRNA-seq data (GEO: GSE74912). Thus, only TFs uniquely expressed in a final cell population were considered.

Genomic regions (scATAC-seq peaks from GSE74912) were listed and ranked based on the significance of DAA (p-value < 0.01) performed by Monocle2, and used for functional annotation by TFBS using position weight matrices (PWM, p-value < 0.0001) from the HOCOMOCO database.⁸

TFs ranking via GSEA-like enrichment analysis

GSEA¹⁸ was modified to perform the TF ranking according to their significance for a given cell conversion.

Since TF sequence preferences and, therefore, the quantity of TFBS for each TF is different, TFs annotations are presented highly unequally in the regions ranking. Thereby, GSEA here was utilized to infer the degree of TFBS abundance at the top of the regions ranking for a given conversion.

Consequently, for GSEA, the genomic regions ranking annotated with TFBS was taken as a pre-ranked list of TFs and each separate factor as a signature gene set. The final TFs ranking obtained from GSEA, thus, represents the significance of distinct TFs for cell differentiation or conversion.

Figure 3. UMAP clustering of (A) scATAC-seq and (B) scRNA-seq of the 13 primary hematopoietic cell types (GSE74912).

Figure 4. Heatmap of the most differentially accessible scATAC-seq regions between hematopoietic stem cell (HSC) and CD4(+) T helper cells (CD4Tcell) cells (GSE74912).

Discussion

The proposed pipeline utilizes both transcriptomic and epigenomic data at the single-cell resolution to search for core TFs that enable cell differentiation and conversion within the human hematopoietic system. The transcription factors rankings obtained (Table 1) suggest that the current approach is capable of predicting subsets of core TFs as well as reflecting their importance for cell differentiation and conversion between cells.

Figure 5. Heatmap of the most differentially expressed genes from scRNA-seq data between hematopoietic stem cell (HSC) and CD4(+) T helper cells (CD4Tcell) cells (GSE74912).

Conclusions

Herein, we described a method for integrating single-cell chromatin accessibility and gene expression data that can successfully select core TFs for cell differentiation and conversion *in silico*.

Data availability

Underlying data

Gene Expression Omnibus: A Single-Cell Atlas of in vivo Mammalian Chromatin Accessibility, [https://identifiers.org/](https://identifiers.org/geo:GSE111586)

geo: GSE111586

Gene Expression Omnibus: Single-cell epigenomics maps the continuous regulatory landscape of human hematopoietic differentiation [scATAC-Seq], <https://identifiers.org/geo:GSE96769>

Gene Expression Omnibus: ATAC-seq data, <https://identifiers.org/geo:GSE74912>

Extended data

Analysis code available from: <https://github.com/annykay/transFactorsPrediction>

Archived analysis code as at time of publication: <https://doi.org/10.5281/zenodo.5799254>

License: MIT

Competing interests

No competing interests were disclosed.

Grant information

The study was supported by Ministry of Science and Higher Education of the Russian Federation (agreement no. 075-15-2020-899).

References

- Buenrostro JD, Corces MR, Lareau CA, *et al.*: **Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation.** *Cell.* 2018; **173**(6): 1535–1548.e16. [PubMed Abstract](#) | [Publisher Full Text](#)
- Cahan P, Li H, Morris SA, *et al.*: **CellNet: Network biology applied to stem cell engineering.** *Cell.* 2014; **158**: 903–915. [PubMed Abstract](#) | [Publisher Full Text](#)
- Choi YS, Gullicksrud JA, Xing S, *et al.*: **LEF-1 and TCF-1 orchestrate TFH differentiation by regulating differentiation circuits upstream of the transcriptional repressor Bcl6.** *Nat. Immunol.* 2015; **16**: 980–990. [PubMed Abstract](#) | [Publisher Full Text](#)
- Cook DN, Kang HS, Jetten AM: **Retinoic Acid-Related Orphan Receptors (RORs): Regulatory Functions in Immunity, Development, Circadian Rhythm, and Metabolism.** *Nuclear Receptor Research.* 2015; **2**. [PubMed Abstract](#) | [Publisher Full Text](#)
- Corces MR, Buenrostro JD, Wu B, *et al.*: **Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution.** *Nat. Genet.* 2016; **48**: 1193–1203. [PubMed Abstract](#) | [Publisher Full Text](#)
- Henriques T, Gilchrist DA, Nechaev S, *et al.*: **Stable pausing by rna polymerase II provides an opportunity to target and integrate regulatory signals.** *Mol. Cell.* 2013; **52**: 517–528. [PubMed Abstract](#) | [Publisher Full Text](#)
- Iwafuchi-Doi M, Zaret KS: **Pioneer transcription factors in cell reprogramming.** 2014.
- Kulakovskiy IV, Vorontsov IE, Yevshin IS, *et al.*: **HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis.** *Nucleic Acids Res.* 2018; **46**: D252–D259. [PubMed Abstract](#) | [Publisher Full Text](#)
- Lang AH, Li H, Collins JJ, *et al.*: **Epigenetic Landscapes Explain Partially Reprogrammed Cells and Identify Key Reprogramming Genes.** *PLoS Comput. Biol.* 2014; **10**: e1003734. [PubMed Abstract](#) | [Publisher Full Text](#)
- Luckheeram RV, Zhou R, Verma AD, *et al.*: **CD4 +T cells: Differentiation and functions.** 2012.
- Morris SA, Daley GQ: **A blueprint for engineering cell fate: Current technologies to reprogram cell identity.** 2013.
- Nichols J, Zevnik B, Anastassiadis K, *et al.*: **Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.** *Cell.* 1998; **95**: 379–391. [PubMed Abstract](#) | [Publisher Full Text](#)
- Nish SA, Zens KD, Kratchmarov R, *et al.*: **CD4+ T cell effector commitment coupled to self-renewal by asymmetric cell divisions.** *J. Exp. Med.* 2017; **214**: 39–47. [PubMed Abstract](#) | [Publisher Full Text](#)
- Rackham OJ, Firas J, Fang H, *et al.*: **A predictive computational framework for direct reprogramming between human cell types.** *Nat. Genet.* 2016; **48**: 331–335. [PubMed Abstract](#) | [Publisher Full Text](#)
- Reid JE, Wernisch L: **Pseudotime estimation: Deconfounding single cell time series.** *Bioinformatics.* 2016; **32**: 2973–2980. [PubMed Abstract](#) | [Publisher Full Text](#)
- Rivera CM, Ren B: **Mapping human epigenomes.** 2013.
- Roost MS, Van Iperen L, Ariyurek Y, *et al.*: **KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas.** *Stem Cell Reports.* 2015; **4**: 1112–1124. [PubMed Abstract](#) | [Publisher Full Text](#)
- Subramanian A, Tamayo P, Mootha VK, *et al.*: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.** *Proc. Natl. Acad. Sci. U. S. A.* 2005; **102**: 15545–15550. [PubMed Abstract](#) | [Publisher Full Text](#)
- Tapscott SJ, Davis RL, Thayer MJ, *et al.*: **MyoDL: a Myc Requiring Nuclear Phosphoprotein to Convert Region Homology Myoblasts Fibroblasts to.** *Adv. Sci.* 2010.
- Tarazona S: **Differential Expression in RNA-Seq.** *Gene Expr.* 2011.
- Vaquerez JM, Kummerfeld SK, Teichmann SA, *et al.*: **A census of human transcription factors: Function, expression and evolution.** 2009.
- Vierbuchen T, Wernig M: **Molecular Roadblocks for Cellular Reprogramming.** 2012.
- Yamanaka S: **Induced pluripotent stem cells: Past, present, and future.** 2012.
- Zhu J, Yamane H, Paul WE: **Differentiation of Effector CD4 T Cell Populations.** *Annu. Rev. Immunol.* 2010; **28**: 445–489. [PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 31 January 2022

<https://doi.org/10.5256/f1000research.79178.r119675>

© 2022 Dashinimaev E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Erdem B. Dashinimaev

¹ Koltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russian Federation

² Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian National Research Medical University, Moscow, Russian Federation

Major comments:

- On the whole, the article is unclear. I have no doubt that some thoughtful and valuable work has been done in which interesting data have been obtained, but in this manner, the entire work is presented incomprehensibly.
- The abstract and introduction chapters are well written.
- The Results chapter is poorly written and unclear, especially since the Methods chapter, which describes the proposed Pipeline, has moved behind the Results chapter. Considering that the main point of the article is the publication of the pipeline under development, perhaps it makes sense to eliminate the Methods chapter and move it to the Results chapter, taking into account the logic of the narrative. For example, the fact that the main result of the applied method was Table 1 becomes clear only at the end of the article from the Discussion chapter
- The Discussion chapter was written extremely poorly. I suggest that this chapter should include references to similar published works (pipelines) from other teams in the field, and discuss their similarities and differences from your work (pipelines). It is also necessary to discuss the results in terms of the value of the data obtained, their potential applications in different fields of science and biotechnology. It is also necessary to highlight the disadvantages and limitations of your method and possible ways to solve them.

Minor comments:

- Fig.2. - missing the transcription of the abbreviations of cell type names.

- Perhaps, based on the logic of the narrative, swap Fig. 2 and Fig. 3.
- The text contains a lot of abbreviations that are not explained in any way - for example TFBS, ATAC-seq, scATAC-seq, GSEA, DEG, Monocle2, PWM, HOCOMOCO, etc.

Of course, for narrow specialists these acronyms make sense, but one of the tasks of scientific publications is to convey information to a wider audience in the most accessible way possible. All the more so given the multidisciplinary nature of F1000Research.

- "Pluripotency" is better replaced with "stemness". ¹

Conclusion:

- I believe that the article requires significant revisions and rewrites.

References

1. Melton D: 'Stemness'. 2014. 7-17 [Publisher Full Text](#)

Is the rationale for developing the new method (or application) clearly explained?

Partly

Is the description of the method technically sound?

Partly

Are sufficient details provided to allow replication of the method development and its use by others?

Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cell biology, human cell reprogramming, regenerative biomedicine

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 27 January 2022

<https://doi.org/10.5256/f1000research.79178.r119678>

© 2022 Boeva V et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Valentiva Boeva

Department of Computer Science, Institute for Machine Learning, ETH Zurich, Zurich, Switzerland

Samuel Gunz

Department of Computer Science, Institute for Machine Learning, ETH Zurich, Zurich, Switzerland

Rationale:

The rationale of developing this method is clearly stated. The authors identify experimental work to be the bottleneck in identifying core transcription factors (TFs) and therefore suggest a computational approach to solve this challenge. However, the authors do not state if there are existing methods in the field that address this problem. This information should be included, also in case there are similar methods that are related to the problem of identifying core TFs.

Methods:

The described methodology is overall understandable and seems technically sound. Nevertheless, parts of the described methods are too short and should be further specified. It includes

1. Details on the construction of the MST (what were exactly the input data? Some details on the algorithm or a reference would be useful too).
2. More details on how GSEA was adjusted in this method. What is the input into GSEA? Why is it applied?
3. Details on the parsing of the obtained TF rankings (e.g. if a threshold was used).

No information is given about data pre-processing requirements (scRNA-seq, scATAC-seq) to use the presented method. If the authors want their method to be used on other datasets this information should be included.

Data used in this study includes 'A Single-Cell Atlas of in vivo Mammalian Chromatin Accessibility' (GSE111586) which was obtained from mice. An explanation why this data was used should be included.

Reproducibility:

The analysis of this manuscript cannot be reproduced for a number of reasons. First, the code cannot be found from the link of the GitHub repository. (I assume that the correct link should be <https://github.com/annykay/transFactorsPrediction->) Second, major parts of the analysis are missing in the GitHub repository, as well as in the [archived code repository](#). The missing parts include the construction of the MST, TF filtration and TFBS annotation and TF ranking via GSEA-like enrichment analysis and parsing of the obtained TF rankings.

In addition, there were a number of issues to reproduce the figures using the scripts that were available in the [archived code repository](#).

1. Data referenced for Figure 3 (A) (GSE74912) do not correspond to scATAC-seq but to bulk ATAC-seq data.
2. The data referenced for Figure 3 (B) scRNA-seq are not retrievable from the accession number the authors specify. Another accession number was listed in the code (GSE74246) which seems to correspond to bulk RNA-seq (not scRNA-seq) data.
3. A random seed for the UMAP algorithm and other parameters should be specified in the scripts such that the results are reproducible.

Results and discussion:

The main result of this study is a table of important TFs in hematopoietic cell differentiation. It is not clear why only these 5 TFs are reported (see comment on parsing of the obtained TF rankings). Moreover, the result is not clearly put into context in the discussion. Questions that have to be discussed include:

1. Are the reported TFs all TFs involved in hematopoietic cell differentiation known from the literature?
2. Are TFs involved in hematopoietic differentiation absent in the final enrichment list, if yes how many and what could be the reason that the method did not identify them?
3. Were some core TFs newly discovered using this method?
4. Can alternative methods, if they exist, retrieve these TFs?

Given the title, the authors claim that they can identify core TFs for cell reprogramming. It is clear why the same idea could in principle be used in both cell reprogramming and differentiation. However, the method is not applied to identify core TFs in cell reprogramming in the manuscript. While the authors claim that their method can determine core TFs for cell differentiation this does not directly imply that the method can also be used to determine core TFs for cell reprogramming. If this result should be included in the title and the conclusion, results supporting this claim must be shown in the manuscript.

Other comments:

- There is a typo in referencing slingshot in paragraph Differential accessibility analysis.
- The links to the Gene Expression Omnibus do not always work.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epigenetics, transcriptional control, bioinformatics

We confirm that we have read this submission and believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research