A Manuscript Microscope 'Sentence Audit' of:

Network diffusion for scalable embedding of massive single-cell ATAC-seq data

What is the Manuscript Microscope Sentence Audit?

The Manuscript Microscope Sentence Audit is a research paper introspection system that parses the text of your manuscript into minimal sentence components for faster, more accurate, enhanced proofreading.

Why use a Sentence Audit to proofread your manuscript?

- Accelerated Proofreading: Examine long technical texts in a fraction of the usual time.
- Superior Proofreading: Detect subtle errors that are invisible to traditional methods.
- Focused Proofreading: Inspect each individual sentence component in isolation.
- Reliable Proofreading: Ensure every single word of your manuscript is correct.
- Easier Proofreading: Take the hardship out of crafting academic papers.

Bonus 1: Improved Productivity: Rapidly refine rough drafts to polished papers.
Bonus 2: Improved Authorship: Cultivate a clear, concise, consistent, writing style.
Bonus 3: Improved Reputation: Become known for rigorously precise publications.

Manuscript Source: https://www.biorxiv.org/content/10.1101/2021.03.05.434093v1

Manuscript Authors: Kangning Dong & Shihua Zhang

Audit Date: 29/03/21 **Audit Identifier:** 6Y3C4EB48429YJ6 **Code Version:** 3.6

Features of the Sentence Audit:

The Sentence Audit combines two complementary proofreading approaches:

1. Each sentence of your text is parsed and displayed in isolation for focused inspection.
2. Each individual sentence is further parsed into Minimal Sentence Components for a deeper review of the clarity, composition and consistency of the language you used.

The Minimal Sentence Components shown are the smallest coherent elements of each sentence of your text as derived from it's conjunctions, prepositions and selected punctuation symbols (i.e. commas, semicolons, round and square brackets).

The combined approaches ensure easier, faster, more effective proofreading.

Comments and Caveats:

• The sentence parsing is achieved using a prototype natural language processing pipeline written in Python and may include occasional errors in sentence segmentation.

• Depending on the source of the input text, the Sentence Audit may contain occasional html artefacts that are parsed as sentences (E.g. "Download figure. Open in new tab").

• Always consult the original research paper as the true reference source for the text.

Contact Information:

To get a Manuscript Microscope Sentence Audit of any other research paper, simply forward any copy of the text to John.James@OxfordResearchServices.com.

All queries, feedback or suggestions are also very welcome.

Research Paper Sections:

The sections of the research paper input text parsed in this audit.

Section No.	Headings	Sentences
Section: 1	ABSTRACT	11
Section: 2	Introduction	16
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0

The Sentence Audit Of The Research Paper

Title **Network diffusion for scalable embedding of massive single-cell ATAC-seq data**

S1 [001] ABSTRACT

S1 [002] With the rapid development of single-cell ATAC-seq technology, it has become possible to profile the chromatin accessibility of massive individual cells.

With the rapid development ...
... of single-cell ATAC-seq technology, ...
... it has become possible ...
... to profile the chromatin accessibility ...
... of massive individual cells.

S1 [003] However, it remains challenging to characterize their regulatory heterogeneity due to the high-dimensional, sparse and near-binary nature of data.

However, ...
... it remains challenging ...
... to characterize their regulatory heterogeneity ...
... due to the high-dimensional, ...
... sparse ...
... and near-binary nature ...
... of data.

S1 [004] Most existing data representation methods were designed based on correlation, which may be ill-defined for sparse data.

Most existing data representation methods were designed based ...
... on correlation, ...
... which may be ill-defined ...
... for sparse data.

S1 [005] Moreover, these methods do not well address the issue of excessive zeros.

Moreover, ...
... these methods do not well address the issue ...
... of excessive zeros.

S1 [006] Thus, a simple, fast and scalable approach is needed to analyze single-cell ATAC-seq data with massive cells, address the “missingness” and accurately categorize cell types.

Thus, ...
... a simple, ...
... fast ...
... and scalable approach is needed ...
... to analyze single-cell ATAC-seq data ...
... with massive cells, ...
... address the “missingness” ...

... and accurately categorize cell types.

S1 [007] To this end, we developed a network diffusion method for scalable embedding of massive single-cell ATAC-seq data (named as scAND).

To this end, ...
... we developed a network diffusion method ...
... for scalable embedding ...
... of massive single-cell ATAC-seq data ...
... (named ...
... as scAND).

S1 [008] Specifically, we considered the near-binary single-cell ATAC-seq data as a bipartite network that reflects the accessible relationship between cells and accessible regions, and further adopted a simple and scalable network diffusion method to embed it.

Specifically, ...
... we considered the near-binary single-cell ATAC-seq data ...
... as a bipartite network ...
... that reflects the accessible relationship ...
... between cells ...
... and accessible regions, ...
... and further adopted a simple ...
... and scalable network diffusion method ...
... to embed it.

S1 [009] scAND can take information from similar cells to alleviate the sparsity and improve cell type identification.

scAND can take information ...
... from similar cells ...
... to alleviate the sparsity ...
... and improve cell type identification.

S1 [010] Extensive tests and comparison with existing methods using synthetic and real data as benchmarks demonstrated its distinct superiorities in terms of clustering accuracy, robustness, scalability and data integration.

Extensive tests ...
... and comparison ...
... with existing methods ...
... using synthetic ...
... and real data ...
... as benchmarks demonstrated its distinct superiorities ...
... in terms ...
... of clustering accuracy, ...
... robustness, ...
... scalability ...
... and data integration.

S1 [011] Availability The Python-based scAND tool is freely available at http://page.amss.ac.cn/shihua.zhang/software.html.

Availability The Python-based scAND tool is freely available ...

... at ...

... http://page.amss.ac.cn/shihua.zhang/software.html.

S2 [012] Introduction

S2 [013] Cell type-specific genomic regulation is driven by the binding of transcription factors (TFs) in accessible genomic regions.

Cell type-specific genomic regulation is driven ...

... by the binding ...

... of transcription factors ...

... (TFs) ...

... in accessible genomic regions.

S2 [014] Thus, chromatin accessibility can be used to identify cis-regulatory elements and directly depict cellular identity [1].

Thus, ...

... chromatin accessibility can be used ...

... to identify cis-regulatory elements ...

... and directly depict cellular identity ...

... [1].

S2 [015] Single-cell Assay for Transposase-Accessible Chromatin using sequencing (Single-cell ATAC-seq or scATAC-seq) has enabled genome-wide profiling of chromatin accessibility at single-cell resolution and can thus reveal epigenetic heterogeneity at cellular level [2, 3].

Single-cell Assay ...

... for Transposase-Accessible Chromatin ...

... using sequencing ...

... (Single-cell ATAC-seq ...

... or scATAC-seq) ...

... has enabled genome-wide profiling ...

... of chromatin accessibility ...

... at single-cell resolution ...

... and can thus reveal epigenetic heterogeneity ...

... at cellular level ...

... [2, 3]...

... .

S2 [016] However, clustering of single cells based on scATAC-seq data is a challenging task due to their massive dimensionality and extremely sparse nature (Supplementary Figure S1).

However, ...

... clustering ...

... of single cells based ...

... on scATAC-seq data is a challenging task ...

... due to their massive dimensionality ...

... and extremely sparse nature ...

... (Supplementary Figure S1).

S2 [017] For example, a recent study produced an unprecedented large-scale scATAC-seq dataset containing ~140k cells in a single study [4].

For example, ...
... a recent study produced an unprecedented large-scale scATAC-seq dataset containing ~140k cells ...
... in a single study ...
... [4].

S2 [018] One can expect that massive scATAC-seq data will be available in the near future.

One can expect ...
... that massive scATAC-seq data will be available ...
... in the near future.

S2 [019] On the other hand, the feature dimension of scATAC-seq data is about 10 times larger than that of single-cell RNA-seq data.

On the other hand, ...
... the feature dimension ...
... of scATAC-seq data is ...
... about 10 times larger ...
... than that ...
... of single-cell RNA-seq data.

S2 [020] Thus, clustering methods for scATAC-seq data must be computationally efficient to satisfy the basic requirement of scalability and high-dimensionality.

Thus, ...
... clustering methods ...
... for scATAC-seq data must be computationally efficient ...
... to satisfy the basic requirement ...
... of scalability ...
... and high-dimensionality.

S2 [021] Moreover, sparsity is intrinsic to single-cell epigenomic data due to the low DNA copy number; that is, only 0, 1 or 2 reads can be captured at any genomic locus within a diploid genome.

Moreover, ...
... sparsity is intrinsic ...
... to single-cell epigenomic data ...
... due to the low DNA copy number; ...
... that is, ...
... only 0, ...
... 1 ...
... or 2 reads can be captured ...
... at any genomic locus ...
... within a diploid genome.

S2 [022] Normally, because of the limited number of captured reads, only 1–10% of expected accessible regions (i.e. peaks) could be detected per cell [5].

Normally, ...

End of Sample Audit

This is a truncated Manuscript Microscope Sample Audit.

To get the full audit of this text (or any other research paper),
forward a copy of the research paper to John James at
John.James@OxfordResearchServices.com
