Choice of alignment pipeline strongly influences clustering quality of scRNA-seq datasets

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: Review long technical papers 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, one at a time.
• Easier Proofreading: Take the hardship out of examining complex academic text.
• Safer Proofreading: Ensure every single word of your manuscript is correct.

+ 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/10.1101/2021.03.05.434032v1
Manuscript Authors: Inbal Shainer & Manuel Stemmer

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 Python natural language processing pipeline and may result in occasional sentence segmentation or parsing errors.

• 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 of 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	15
Section: 2	Background	28
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

Title	**Choice of alignment pipeline strongly influences clustering quality of scRNA-seq datasets**

S1 [001]   Abstract

S1 [002]   Background

> Background

S1 [003]   Single-cell RNA sequencing (scRNA-seq) has quickly become one of the most dominant techniques in modern transcriptome assessment.

> Single-cell RNA sequencing ...
>
> ... (scRNA-seq) ...
>
> ... has quickly become one ...
>
> ... of the most dominant techniques ...
>
> ... in modern transcriptome assessment.

S1 [004]   In particular, 10X Genomics' Chromium system, with its high throughput approach, turn key and thorough user guide made this cutting-edge technique accessible to many laboratories using diverse animal models.

> In particular, ...
>
> ... 10X Genomics' Chromium system, ...
>
> ... with its high throughput approach, ...
>
> ... turn key ...
>
> ... and thorough user guide made this cutting-edge technique accessible ...
>
> ... to many laboratories ...
>
> ... using diverse animal models.

S1 [005]   However, standard downstream processing, including the alignment and cell filtering pipelines might not be ideal for every organism or tissue.

> However, ...
>
> ... standard downstream processing, ...
>
> ... including the alignment ...
>
> ... and cell filtering pipelines ...
>
> ... might not be ideal ...
>
> ... for every organism ...
>
> ... or tissue.

S1 [006]   Here we applied an alternative strategy, based on the pseudoaligner kallisto, on twenty-two publicly available single cell sequencing datasets from a wide range of tissues of eight organisms and compared the results with the standard 10X Genomics' Cell Ranger pipeline.

> Here we applied an alternative strategy, ...
>
> ... based ...
>
> ... on the pseudoaligner kallisto, ...
>
> ... on twenty-two publicly available single cell sequencing datasets ...

... from a wide range ...

... of tissues ...

... of eight organisms ...

... and compared the results ...

... with the standard 10X Genomics' Cell Ranger pipeline.

S1 [007]   Results

Results

S1 [008]   In most of the tested samples, kallisto outperformed Cell Ranger in sequencing read alignment rates and total gene detection rates.

In most of the tested samples, ...

... kallisto outperformed Cell Ranger ...

... in sequencing read alignment rates ...

... and total gene detection rates.

S1 [009]   Although datasets processed with Cell Ranger had higher cell counts, outside of human and mouse datasets, these additional cells were routinely of low quality, containing low gene detection rates.

Although datasets processed ...

... with Cell Ranger had higher cell counts, ...

... outside ...

... of human ...

... and mouse datasets, ...

... these additional cells were routinely ...

... of low quality, ...

... containing low gene detection rates.

S1 [010]   Thorough downstream analysis of one kallisto processed dataset, obtained from the zebrafish pineal gland, revealed clearer clustering, allowing the identification of an additional photoreceptor cell type that previously went undetected.

Thorough downstream analysis ...

... of one kallisto processed dataset, ...

... obtained ...

... from the zebrafish pineal gland, ...

... revealed clearer clustering, ...

... allowing the identification ...

... of an additional photoreceptor cell type ...

... that previously went undetected.

S1 [011]   The finding of the new cluster suggests that the photoreceptive pineal gland is essentially a bi-chromatic tissue containing both green and red cone-like photoreceptors and implies that the alignment and processing pipeline can affect the discovery of biologically-relevant cell types.

The finding ...

... of the new cluster suggests ...

... that the photoreceptive pineal gland is essentially a bi-chromatic tissue containing both green ...

... and red cone-like photoreceptors ...

... and implies ...

... that the alignment ...

... and processing pipeline can affect the discovery ...

... of biologically-relevant cell types.

S1 [012]   Conclusion

Conclusion

S1 [013]   While Cell Ranger favors higher cell numbers, using kallisto results in datasets with higher median gene detection per cell.

While Cell Ranger favors higher cell numbers, ...

... using kallisto results ...

... in datasets ...

... with higher median gene detection ...

... per cell.

S1 [014]   We could demonstrate that cell type identification was not hampered by the lower cell count, but in fact improved as a result of the high gene detection rate and the more stringent filtering.

We could demonstrate ...

... that cell type identification was not hampered ...

... by the lower cell count, ...

... but in fact improved ...

... as a result ...

... of the high gene detection rate ...

... and the more stringent filtering.

S1 [015]   It is thus beneficial to favor high quality cells and accept a lower cell count, leading to an improved classification of cell types.

It is thus beneficial ...

... to favor high quality cells ...

... and accept a lower cell count, ...

... leading ...

... to an improved classification ...

... of cell types.

S2 [016]   Background

S2 [017]   Single-cell transcriptome sequencing (scRNA-seq) has rapidly become one of the most popular tools for dissecting the transcriptomic states of individual cells in a tissue of interest.

Single-cell transcriptome sequencing ...

... (scRNA-seq) ...

... has rapidly become one ...

... of the most popular tools ...

... for dissecting the transcriptomic states ...

... of individual cells ...

... in a tissue ...

... of interest.

S2 [018]   It can be applied to virtually any biological sample as long as a reference genome is
available.

> It can be applied ...
>
> ... to virtually any biological sample ...
>
> ... as long ...
>
> ... as a reference genome is available.

S2 [019]   Among the available scRNA-seq techniques, the Chromium (10X Genomics) platform is
probably the most widely used at this point.

> Among the available scRNA-seq techniques, ...
>
> ... the Chromium ...
>
> ... (10X Genomics) ...
>
> ... platform is probably the most widely used ...
>
> ... at this point.

S2 [020]   Thanks to its user-friendly design and very well-documented workflow, it has quickly
emerged as the top choice for many researchers and clinicians (Cao et al., 2019; Davie et
al., 2018; Kölsch et al., 2020; Packer et al., 2019; Pandey et al., 2018; Peuß et al., 2020;
Shainer et al., 2019; Wang et al., 2020).

> Thanks ...
>
> ... to its user-friendly design ...
>
> ... and very well-documented workflow, ...
>
> ... it has quickly emerged ...
>
> ... as the top choice ...
>
> ... for many researchers ...
>
> ... and clinicians ...
>
> ... (Cao et al., 2019; ...
>
> ... Davie et al., 2018; ...
>
> ... Kölsch et al., 2020; ...
>
> ... Packer et al., 2019; ...
>
> ... Pandey et al., 2018; ...
>
> ... Peuß et al., 2020; ...
>
> ... Shainer et al., 2019; ...
>
> ... Wang et al., 2020).

S2 [021]   Its droplet-based design and simple workflow make it the ideal technique for surveying
hundreds to thousands of cells in a single experiment, without needing prior knowledge of
the system (Svensson et al., 2018; Zheng et al., 2017).

> Its droplet-based design ...
>
> ... and simple workflow make it the ideal technique ...
>
> ... for surveying hundreds ...
>
> ... to thousands ...
>
> ... of cells ...
>
> ... in a single experiment, ...
>
> ... without needing ...
>
> ... prior knowledge ...
>
> ... of the system ...

End of Sample Audit