

CopyMix: Mixture Model Based Single-Cell Clustering and Copy Number Profiling using Variational Inference

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/10.1101/2020.01.29.926022v2>

Manuscript Authors: Negar Safinianaini, Camila P. E. de Souza & Jens Lagergren

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.

[illegible]

Title **CopyMix: Mixture Model Based Single-Cell Clustering and Copy Number Profiling using Variational Inference**

S1 [001] Abstract

S1 [002] Single-cell sequencing technologies are becoming increasingly more established, in particular, in the study of tumor heterogeneity.

Single-cell sequencing technologies are becoming increasingly more established, ...
... in particular, ...
... in the study ...
... of tumor heterogeneity.

S1 [003] Investigating tumor heterogeneity is imperative to better understand how tumors evolve since each cell subpopulation harbors a unique set of genomic features that yields a unique phenotype, an issue that is bound to have clinical relevance.

Investigating tumor heterogeneity is imperative ...
... to better understand how tumors evolve ...
... since each cell subpopulation harbors a unique set ...
... of genomic features ...
... that yields a unique phenotype, ...
... an issue ...
... that is bound ...
... to have clinical relevance.

S1 [004] Clustering of cells based on copy number data, obtained from single-cell DNA sequencing, provides an opportunity to assess different tumor cell subpopulations.

Clustering ...
... of cells based ...
... on copy number data, ...
... obtained ...
... from single-cell DNA sequencing, ...
... provides an opportunity ...
... to assess different tumor cell subpopulations.

S1 [005] Accordingly, computational methods have emerged for single-cell copy number profiling as well as clustering; however, these two tasks have up to now been handled sequentially, and, included various ad-hoc preprocessing steps; hence, prone to introduce clustering artifacts.

Accordingly, ...
... computational methods have emerged ...
... for single-cell copy number profiling ...
... as well ...
... as clustering; ...
... however, ...
... these two tasks have up to now been handled sequentially, ...

... and, ...
... included various ad-hoc preprocessing steps; ...
... hence, ...
... prone ...
... to introduce clustering artifacts.

S1 [006] We propose CopyMix, a mixture model-based method using Variational Inference, to simultaneously cluster cells and infer a single copy number profile for each cluster, revealing the copy number variation pattern behind each cluster.

We propose CopyMix, ...
... a mixture model-based method ...
... using Variational Inference, ...
... to simultaneously cluster cells ...
... and infer a single copy number profile ...
... for each cluster, ...
... revealing the copy number variation pattern behind each cluster.

S1 [007] CopyMix is evaluated using simulated data as well as published biological data from metastatic colorectal cancer.

CopyMix is evaluated ...
... using simulated data ...
... as well ...
... as published biological data ...
... from metastatic colorectal cancer.

S1 [008] The results reveal high clustering performance and low errors in copy number profiling.

The results reveal high clustering performance ...
... and low errors ...
... in copy number profiling.

S1 [009] These favorable results indicate a considerable potential to obtain clinical impact by using CopyMix in studies of cancer tumor heterogeneity.

These favorable results indicate a considerable potential ...
... to obtain clinical impact ...
... by using CopyMix ...
... in studies ...
... of cancer tumor heterogeneity.

S2 [010] 1 Introduction

S2 [011] A tumor typically consists of a collection of heterogeneous cell populations, each having distinct genetic and phenotypic properties, in particular, concerning the capacity to promote cancer progression, metastasis, and therapy resistance Eirew et al. (2015); Nowell (1976).

A tumor typically consists ...
... of a collection ...

... of heterogeneous cell populations, ...
... each having distinct genetic ...
... and phenotypic properties, ...
... in particular, ...
... concerning the capacity ...
... to promote cancer progression, ...
... metastasis, ...
... and therapy resistance Eirew et al. ...
... (2015); ...
... Nowell ...
... (1976).

S2 [012] Single-cell sequencing technologies Gawad et al. (2016); Navin et al. (2011); Shapiro et al. (2013); Zahn et al. (2017) provide a new opportunity to investigate the genomic profile of individual cells both regarding single nucleotide variation (SNV) and copy number variation (CNV).

Single-cell sequencing technologies Gawad et al. ...
... (2016); ...
... Navin et al. ...
... (2011); ...
... Shapiro et al. ...
... (2013); ...
... Zahn et al. ...
... (2017) ...
... provide a new opportunity ...
... to investigate the genomic profile ...
... of individual cells both regarding single nucleotide variation ...
... (SNV) ...
... and copy number variation ...
... (CNV).

S2 [013] CNVs and SNVs have been shown to be important contributors to phenotypic variation relating to health and disease Baslan et al. (2012); Lawson et al. (2018).

CNVs ...
... and SNVs have been shown ...
... to be important contributors ...
... to phenotypic variation relating ...
... to health ...
... and disease Baslan et al. ...
... (2012); ...
... Lawson et al. ...
... (2018).

S2 [014] Moreover, although single-cell SNV profiling is hampered by experimental imperfections such as drop-outs, copy number profiling (detecting single-cell copy number variations) is feasible at least at coarser resolutions.

Moreover, ...
... although single-cell SNV profiling is hampered ...
... by experimental imperfections ...
... such as drop-outs, ...
... copy number profiling ...

... (detecting single-cell copy number variations) ...
... is feasible ...
... at least ...
... at coarser resolutions.

S2 [015] Clustering of cells based on their individual copy number profiles provides the opportunity for SNV profiling of the clusters and, in general, opportunities to better understand tumor subpopulations and tumor heterogeneity, issues that are bound to have clinical relevance.

Clustering ...
... of cells based ...
... on their individual copy number profiles provides the opportunity ...
... for SNV profiling ...
... of the clusters and, ...
... in general, ...
... opportunities ...
... to better understand tumor subpopulations ...
... and tumor heterogeneity, ...
... issues ...
... that are bound ...
... to have clinical relevance.

S2 [016] Current single-cell datasets pose a wealth of computational challenges.

Current single-cell datasets pose a wealth ...
... of computational challenges.

S2 [017] As answers to some of those, methods have emerged for single-cell copy number profiling as well as clustering of cells based on their already inferred copy number profiles, e.g., Garvin et al. (2015); Zahn et al. (2017); Leung et al. (2017); Vitak et al. (2017); however, these two tasks have up to now been handled sequentially with various adhoc preprocessing steps.

As answers ...
... to some of those, ...
... methods have emerged ...
... for single-cell copy number profiling ...
... as well ...
... as clustering ...
... of cells based ...
... on their already inferred copy number profiles, ...
... e.g., ...
... Garvin et al. ...
... (2015); ...
... Zahn et al. ...
... (2017); ...
... Leung et al. ...
... (2017); ...
... Vitak et al. ...
... (2017); ...
... however, ...
... these two tasks have up to now been handled sequentially ...
... with various adhoc preprocessing steps.

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
