

Cross-tissue eQTL mapping in the presence of missing data via surrogate outcome analysis

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/10.1101/2020.11.29.403063v3>

Manuscript Authors: Zachary R. McCaw, Sheila M. Gaynor, Ryan Sun & Xihong Lin

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 **Cross-tissue eQTL mapping in the presence of missing data via surrogate outcome analysis**

S1 [001] Abstract

S1 [002] Missing data are prevalent in the Genotype-Tissue Expression (GTEx) project, where measurements from certain inaccessible tissues, such as the substantia nigra (SSN), are available at much smaller sample sizes than those from accessible tissues, such as blood.

Missing data are prevalent ...
... in the Genotype-Tissue Expression ...
... (GTEx) ...
... project, ...
... where measurements ...
... from certain inaccessible tissues, ...
... such as the substantia nigra ...
... (SSN), ...
... are available ...
... at much smaller sample sizes ...
... than those ...
... from accessible tissues, ...
... such as blood.

S1 [003] This severely limits power for identifying tissue-specific expression quantitative trait loci (eQTL).

This severely limits power ...
... for identifying tissue-specific expression quantitative trait loci ...
... (eQTL).

S1 [004] Here we propose Surrogate Phenotype Regression Analysis (Spray) for leveraging information from a correlated surrogate outcome (e.g. expression in blood) to improve inference on a partially missing target outcome (e.g. expression in SSN).

Here we propose Surrogate Phenotype Regression Analysis ...
... (Spray) ...
... for leveraging information ...
... from a correlated surrogate outcome ...
... (e.g. expression ...
... in blood) ...
... to improve inference ...
... on a partially missing target outcome ...
... (e.g. expression ...
... in SSN).

S1 [005] Rather than regarding the surrogate outcome as a proxy for the target outcome, Spray jointly models the target and surrogate outcomes within a bivariate regression framework.

Rather ...

... than regarding the surrogate outcome ...
 ... as a proxy ...
 ... for the target outcome, ...
 ... Spray jointly models the target ...
 ... and surrogate outcomes ...
 ... within a bivariate regression framework.

S1 [006] Unobserved values of either outcome are regarded as missing data.

Unobserved values ...
 ... of either outcome are regarded ...
 ... as missing data.

S1 [007] We describe and implement an expectation conditional maximization algorithm for performing estimation in the presence of bilateral outcome missingness.

We describe ...
 ... and implement an expectation conditional maximization algorithm ...
 ... for performing estimation ...
 ... in the presence ...
 ... of bilateral outcome missingness.

S1 [008] Spray estimates the same association parameter estimated by standard eQTL mapping, and controls the type I error even when the target and surrogate outcomes are truly uncorrelated.

Spray estimates the same association parameter estimated ...
 ... by standard eQTL mapping, ...
 ... and controls the type I error even ...
 ... when the target ...
 ... and surrogate outcomes are truly uncorrelated.

S1 [009] We demonstrate analytically and empirically, using simulations and GTEx data, that in comparison with marginally modeling the target outcome, jointly modeling the target and surrogate outcomes increases estimation precision and improves power.

We demonstrate analytically ...
 ... and empirically, ...
 ... using simulations ...
 ... and GTEx data, ...
 ... that in comparison ...
 ... with marginally modeling the target outcome, ...
 ... jointly modeling the target ...
 ... and surrogate outcomes increases estimation precision ...
 ... and improves power.

S2 [011] Tissue-specific expression quantitative trait loci (eQTL) are of substantial biological interest as mechanisms for explaining how the genetic variants identified in genome-wide association studies (GWAS) influence complex traits and diseases (Gamazon et al., 2015; Gusev et al., 2016; Hormozdiari et al., 2016; Zhu et al., 2016; Visscher et al., 2017).

Tissue-specific expression quantitative trait loci ...
... (eQTL) ...
... are ...
... of substantial biological interest ...
... as mechanisms ...
... for explaining how the genetic variants identified ...
... in genome-wide association studies ...
... (GWAS) ...
... influence complex traits ...
... and diseases ...
... (Gamazon et al., 2015; ...
... Gusev et al., 2016; ...
... Hormozdiari et al., 2016; ...
... Zhu et al., 2016; ...
... Visscher et al., 2017).

S2 [012] Traditional eQTL studies have focused on accessible tissues such as blood (Nica et al., 2011; Westra et al., 2013), while eQTL analyses in inaccessible tissues, such as brain, have been hindered by limited sample sizes.

Traditional eQTL studies have focused ...
... on accessible tissues ...
... such as blood ...
... (Nica et al., 2011; ...
... Westra et al., 2013), ...
... while eQTL analyses ...
... in inaccessible tissues, ...
... such as brain, ...
... have been hindered ...
... by limited sample sizes.

S2 [013] Cross-tissue studies, including the Genotype-Tissue Expression Project (GTEx), have demonstrated that the effect sizes of eQTL are heterogeneous across tissues (Consortium, 2017).

Cross-tissue studies, ...
... including the Genotype-Tissue Expression Project ...
... (GTEx), ...
... have demonstrated ...
... that the effect sizes ...
... of eQTL are heterogeneous ...
... across tissues ...
... (Consortium, 2017).

S2 [014] Consequently, studying only accessible tissues is insufficient to understand the genetic basis of gene regulation.

Consequently, ...
... studying ...

... only accessible tissues is insufficient ...
... to understand the genetic basis ...
... of gene regulation.

S2 [015] Larger sample sizes are needed to provide sufficient power for reliable eQTL detection in inaccessible tissues, and there is great interest in borrowing information from accessible tissues to increase the effective sample sizes of inaccessible tissues.

Larger sample sizes are needed ...
... to provide sufficient power ...
... for reliable eQTL detection ...
... in inaccessible tissues, ...
... and there is great interest ...
... in borrowing information ...
... from accessible tissues ...
... to increase the effective sample sizes ...
... of inaccessible tissues.

S2 [016] The goal of improving power for eQTL mapping in the substantia nigra (SSN) motivated our work on cross-tissue eQTL mapping.

The goal ...
... of improving power ...
... for eQTL mapping ...
... in the substantia nigra ...
... (SSN) ...
... motivated our work ...
... on cross-tissue eQTL mapping.

S2 [017] SSN is a region of the midbrain implicated in the development of Parkinson's disease (Poewe et al., 2017).

SSN is a region ...
... of the midbrain implicated ...
... in the development ...
... of Parkinson's disease ...
... (Poewe et al., 2017).

S2 [018] Due to the scarcity of gene expression data, no previous studies have focused on eQTL mapping in this region.

Due ...
... to the scarcity ...
... of gene expression data, ...
... no previous studies have focused ...
... on eQTL mapping ...
... in this region.

S2 [019] At the time of our analysis, only 73 genotyped subjects with expression data in SSN were available from GTEx, in contrast to 369 subjects with expression in whole blood.

At the time ...
... of our analysis, ...
... only 73 genotyped subjects ...

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
