

SOFTWARE

Open Access



Enhanced JBrowse plugins for epigenomics data visualization

Brigitte T. Hofmeister^{1*} and Robert J. Schmitz^{2*}

Abstract

Background: New sequencing techniques require new visualization strategies, as is the case for epigenomics data such as DNA base modifications, small non-coding RNAs, and histone modifications.

Results: We present a set of plugins for the genome browser JBrowse that are targeted for epigenomics visualizations. Specifically, we have focused on visualizing DNA base modifications, small non-coding RNAs, stranded read coverage, and sequence motif density. Additionally, we present several plugins for improved user experience such as configurable, high-quality screenshots.

Conclusions: In visualizing epigenomics with traditional genomics data, we see these plugins improving scientific communication and leading to discoveries within the field of epigenomics.

Keywords: Epigenomics, Genomics, Genome browser, Visualization

Background

As next-generation sequencing techniques for detecting and quantifying DNA nucleotide variants, histone modifications and RNA transcripts become widely implemented, it is imperative that graphical tools such as genome browsers are able to properly visualize these specialized data sets. Current genome browsers such as UCSC genome browser [1], AnnoJ [2], IGV [3], WashU EpiGenome Browser [4], Epiviz [5], IGB [6], and JBrowse [7], have limited capability to visualize these data sets effectively, hindering the visualization and potential discoveries with new sequencing technologies. JBrowse is used by numerous scientific resources, such as Phytozome [8], CoGe [9], WormBase [10], and Araport [11] because it is highly customizable and adaptable with modular plugins [7].

Epigenomics is an emerging area of research that generates a significant amount of specialized sequencing data which cannot be efficiently visualized using standard genome browsers. New sequencing technologies such as whole-genome bisulfite sequencing (WGBS) [2, 12], Tet-assisted bisulfite sequencing (TAB-seq) [13], single-molecule real-time sequencing (SMRT) [14], chromatin immunoprecipitation sequencing (ChIP-

seq) [15], assay for transposase-accessible chromatin sequencing (ATAC-seq) [16], RNA-seq [17–19], and small RNA-seq [20] have been instrumental in advancing the field of epigenomics. Epigenomic data sets generated from these techniques typically include: DNA base modifications, mRNAs, small RNAs, histone modifications and variants, chromatin accessibility, and DNA sequence motifs. These techniques have allowed researchers to map the epigenomic landscape at high resolution, greatly advancing our understanding of gene regulation. DNA methylation (4-methylcytosine, 4mC; 5-methylcytosine, 5mC; 5-hydroxymethylcytosine, 5hmC; and 6-methyladenine, 6 mA) and small non-coding RNAs (smRNAs) are modifications often found in epigenomic data sets, and function to regulate DNA repair and transcription by localizing additional chromatin marks or inducing post-transcriptional gene regulation [21–23].

We have developed several JBrowse plugins to address the current limitations of visualizing epigenomics data, which include visualizing base modifications and small RNAs as well as stranded-coverage tracks and sequence motif density. Additionally, we have developed several plugins that add features for improved user experience with JBrowse, including high-resolution browser screenshots. These plugins are freely available and can be used together or independently as needed. In visualizing epigenomics with traditional genomics data, we see these

* Correspondence: bth29393@uga.edu; schmitz@uga.edu

¹Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA

²Department of Genetics, University of Georgia, Athens, GA 30602, USA



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

plugins improving scientific communication and leading to discoveries within the field of epigenomics.

Implementation

Plugins are implemented to work with JBrowse's modular plugin system. Client-side logic, such as visualization, fetching data, and interaction, are written in JavaScript relying on the Dojo library [24]. This includes JavaScript classes for viewing data and storing data. Raw data files are standard in genomics, including BAM files for next-generation sequencing reads [25] and BigWig files for quantitative coverage tracks [26]. Python scripts are included to convert output from analysis pipelines to BigWig files needed by JBrowse. Additional styling for each plugin is provided using CSS. Wherever possible, colorblind safe colors were used to improve accessibility.

Results

Base modifications

We have developed a plugin to visualize the quantity of 4mC, 5mC, 5hmC, and 6 mA at single base-pair resolution. When studying 5mC, the modification is split into two (CG and CH; where H is any nucleotide except G) sequence contexts for animals or three (CG, CHG, and CHH) sequence contexts for plants, as each context is established and/or maintained by different pathways with different functional roles [22]. Our plugin visualizes the quantity of methylation at each cytosine or adenine using a bar plot (Fig. 1), where values are positive or

negative to signify the DNA strand. In most genome browsers, each sequence context must be shown as a different track (Fig. 1a). This is cumbersome when viewing multiple samples and makes it more difficult to determine overlap between context or samples. Our plugin is advantageous because, we color-code 4mC, 5mC, 5hmC, and 6 mA sequence contexts and display them on a single track (Fig. 1b, Additional file 1: Figure S1). However, focusing on a single context or modification can be important, thus our plugin offers several filtering options including by sequence context and base modification.

Small RNAs

Currently, JBrowse represents each sequenced RNA as a single read and is colored by sequenced strand (Fig. 2a). When analyzing smRNAs, strand alone does not always provide sufficient information; the size (nucleotides [nt]) of smRNA and strandedness indicate potential function [21]. For example, in plants, 21 nt microRNAs can be aligned to single strand and 24 nt small interfering RNAs can be aligned to both strands [27]. Products of RNA degradation, however, have varying sizes and align to one strand. To improve smRNA visualization, we color-code reads by smRNA size and retain strand information by placement of smRNAs within the track relative to the y-axis (Fig. 2b). This plugin also includes the ability to filter the reads in a track or multiple tracks by size, strand, and read quality.

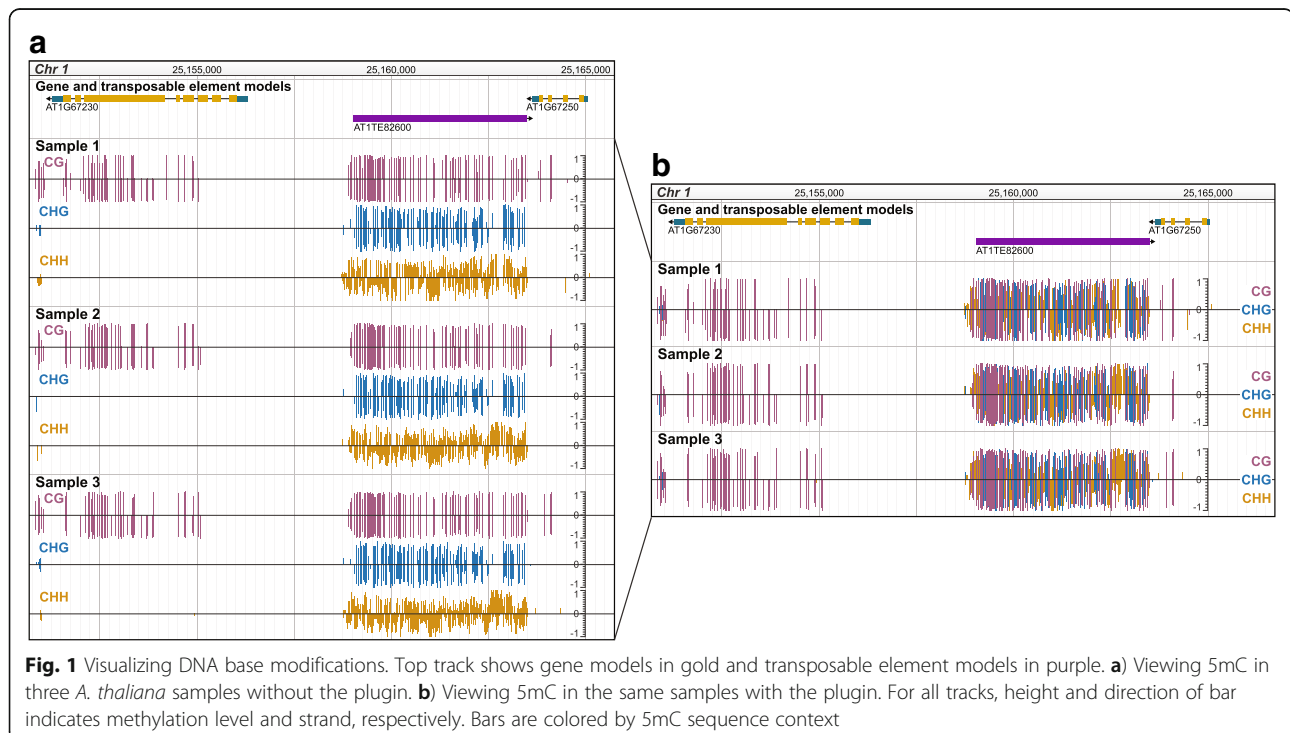




Fig. 2 Visualizing small RNAs. Top track shows gene models in gold and transposable element models in purple. **a)** Viewing smRNA reads, 18 nt - 30 nt, in an *A. thaliana* sample using the general JBrowse alignments track. Color indicates strand; red, forward; blue, reverse. **b)** Viewing the same smRNA reads using the smRNA alignments track provided by the plugin. Color indicates read length. Position above and below the y-axis origin indicates forward and reverse strand, respectively. Unfilled reads map to multiple genomic locations and filled reads map uniquely

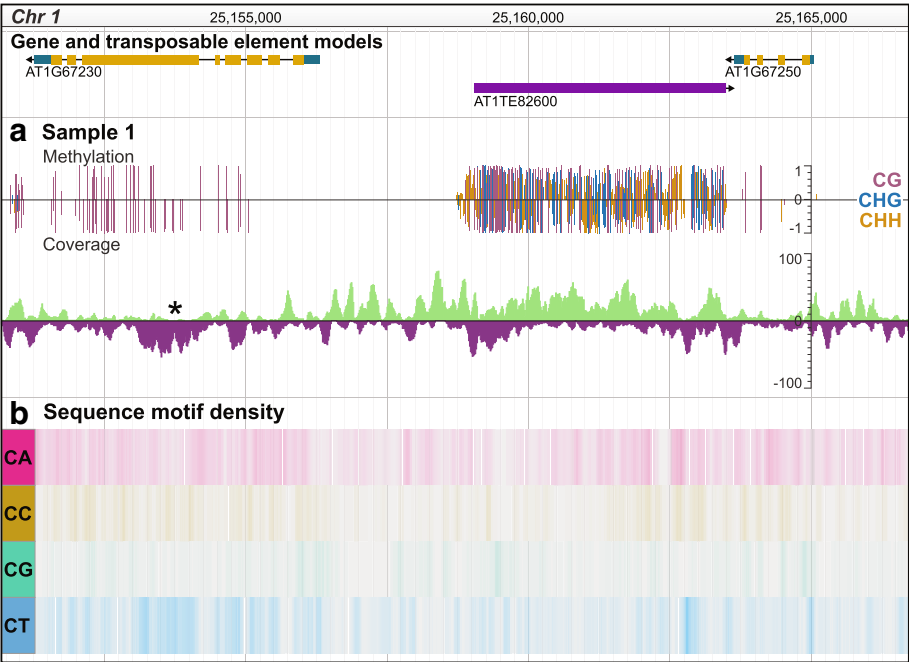


Fig. 3 Visualizing stranded coverage and sequence motif density. Top track shows gene models in gold and transposable element models in purple. **a)** Stranded read coverage for sample used in the methylation track. Asterisk (*) indicates uneven strand coverage which affects the perceived methylation level. **b)** Dinucleotide sequence motif density in *A. thaliana*. Darker color indicates higher density

Stranded read coverage

Quantitative coverage tracks are necessary for any worthwhile genome browser. It is important for visualizing DNA-protein interactions via ChIP-seq and chromatin accessibility via ATAC-seq where coverage is computed in a strand-independent manner. However, for strand-dependent data types, such as 5mC, small RNAs, and mRNAs, read coverage can greatly vary for opposite strands. The default coverage tracks are unable to handle this, thus we developed a plugin which shows stranded read coverage. For example, WGBS can have uneven coverage on both strands which can make only one strand seem methylated (Fig. 3a).

Motif density

Sequence motifs not only have important roles for protein binding, i.e. binding motifs, but can also impact chromatin formation [28] and recombination hotspots [29]. When correlating the frequency of a sequence motif with another characteristic, i.e. 5mC or histone

modification localization, it is preferred to visualize motif density over larger regions compared to single base-pair resolution. To address this, we developed a plugin which visualizes sequence motif density across the genome as a heatmap (Fig. 3b). Users can input multiple motifs in a single track and IUPAC degenerate nucleotides are supported. We also include several options for heatmap coloring and density computation configuration options.

Exporting browser images

One of the most difficult tasks working with any genome browser is obtaining high-quality screenshots for presentations or publications. We have developed a plugin for JBrowse, which allows the user to take high quality and highly configurable screenshots without installing additional software. A dialog window allows users to set general, output, and track-specific configuration options (Fig. 4). Additionally, our plugin is able to create the screenshot with vector graphic objects, which is

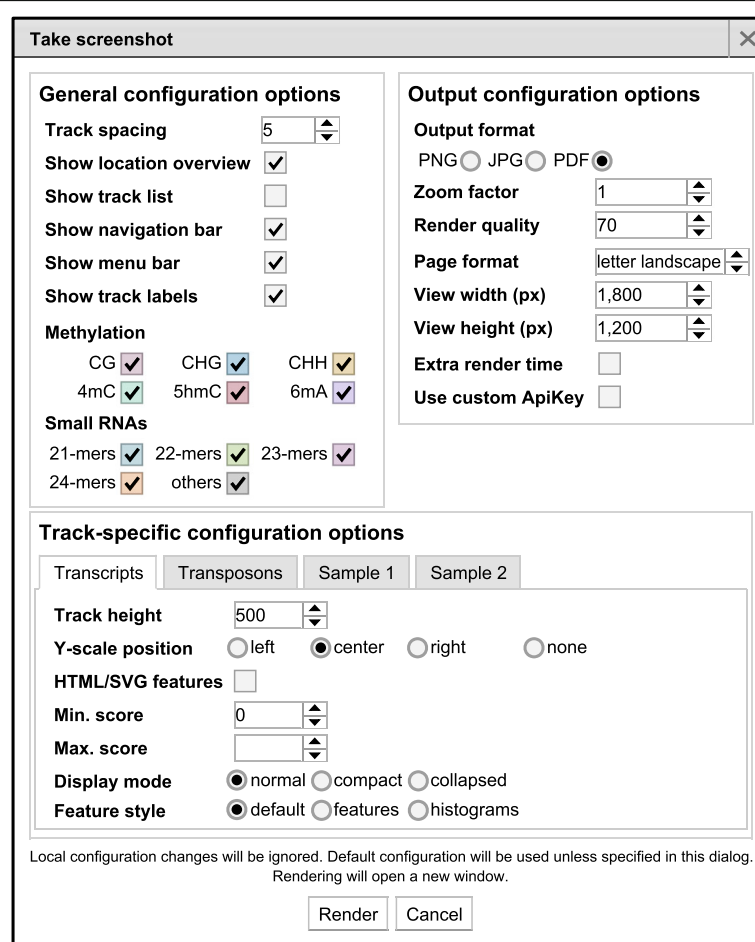


Fig. 4 Screenshot dialog window. The dialog window that opens when taking screenshots with our plugin. There are numerous configuration options for general visualization, image output, and track-specific settings. This includes exporting each track using vector objects

preferred for publication-quality screenshots, without needing to change the underlying track configuration parameters.

Customization

To improve user experience, we have developed several additional JBrowse plugins. These plugins include: (i) Selecting or deselecting all tracks in a category from a hierarchical track list; (ii) An easily customizable y-scale range and location; and (iii) An option to force a track to stay in “feature” view or “histogram” view regardless of the zoom.

Conclusions

With these plugins, we aim to improve epigenomics visualization using JBrowse, a user-friendly genome browser familiar to the research community. All the plugins described can be used together or independently as needed. All plugins are freely available for download and additional customization.

Availability and requirements

Project name: Epigenomics in JBrowse.

Project home page: <http://github.com/bhofmei/bhofmei-jbplugins>

Operating system(s): Platform independent.

Programming language: JavaScript, Python.

Other requirements: JBrowse 1.11.6+.

License: Apache License, Version 2.0.

Any restrictions to use by non-academics: none.

Additional file

Additional file 1: Figure S1. Supplementary methods. (PDF 100 kb)

Abbreviations

4mC: 4-methylcytosine; 5hmC: 5-hydroxymethylcytosine; 5mC: 5-methylcytosine; 6mA: 6-methyladenine; ATAC-seq: Assay for transposase-accessible chromatin sequencing; ChIP-seq: Chromatin immunoprecipitation sequencing; smRNAs: Small non-coding RNAs; SMRT: Single-molecule real-time sequencing; TAB-seq: Tet-assisted bisulfite sequencing; WGBS: Whole-genome bisulfite sequencing

Acknowledgements

We would like to thank Adam Bewick, Lexiang Ji, William Jordan, and Melissa Shockey for comments and discussions. We would like to thank Eric Lyons and Colin Diesh for open-source software code that influenced these plugins early in development. We would like to thank all members of the Schmitz lab for using the plugins during development and suggesting additional features. Additionally, we would like to thank Scott Cain and Mathew Lewsey for being early adopters.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (T32GM007103) to BTH, the National Science Foundation (IOS-1546867) to RJS, and the Office of Research at the University of Georgia to BTH and RJS.

Availability of data and materials

See Additional file 1 for availability and description of data processing for samples used in the figures.

Authors' contributions

Conceptualization and design: BTH and RJS; Implementation and testing: BTH; Writing: BTH; Review and editing: RJS. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 October 2017 Accepted: 19 April 2018

Published online: 25 April 2018

References

- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12:996–1006.
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell.* 2008; <https://doi.org/10.1016/j.cell.2008.03.029>.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29:24–6.
- Zhou X, Maricque B, Xie M, Li D, Sundaram V, Martin EA, et al. The human epigenome browser at Washington University. *Nat Methods.* 2011;8:989.
- Chelaru F, Smith L, Goldstein N, Bravo HC. Epviz: interactive visual analytics for functional genomics data. *Nat Methods.* 2014;11:938.
- Freese NH, Norris DC, Loraine AE. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics.* 2016; <https://doi.org/10.1093/bioinformatics/btw069>.
- Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 2016. <https://doi.org/10.1186/s13059-016-0924-1>.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012;40:D1178–86.
- Lyons E, Freeling M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* 2008; <https://doi.org/10.1111/j.1365-3113.2007.03326.x>.
- Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res.* 2016; <https://doi.org/10.1093/nar/gkv1217>.
- Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M, et al. Araport: the *Arabidopsis* information portal. *Nucleic Acids Res.* 2015; <https://doi.org/10.1093/nar/gku1200>.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature.* 2008;452:215–9.
- Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012; <https://doi.org/10.1016/j.cell.2012.04.027>.
- Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods.* 2010; <https://doi.org/10.1038/nmeth.1459>.
- Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science.* 2007;316:1497–502.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013; <https://doi.org/10.1038/nmeth.2688>.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science.* 2008; <https://doi.org/10.1126/science.1158441>.
- Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods.* 2008; <https://doi.org/10.1038/nmeth.1223>.

19. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; <https://doi.org/10.1038/nmeth.1226>.
20. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res*. 2008; <https://doi.org/10.1101/gr.7179508>.
21. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet*. 2009; <https://doi.org/10.1038/nrg2504>.
22. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010; <https://doi.org/10.1038/nrg2719>.
23. Marinus MG, Løbner-Olesen A. DNA Methylation. *EcoSal Plus*. 2014; <https://doi.org/10.1128/ecosalplus.ESP-0003-2013>.
24. Dojo Toolkit: Reference Guide. <https://dojotoolkit.org/reference-guide/1.10/>. Accessed 15 July 2017.
25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; <https://doi.org/10.1093/bioinformatics/btp352>.
26. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*. 2010; <https://doi.org/10.1093/bioinformatics/btq351>.
27. Finnegan EJ, Matzke MA. The small RNA world. *J Cell Sci*. 2003; <https://doi.org/10.1242/jcs.00838>.
28. Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, et al. A genomic code for nucleosome positioning. *Nature*. 2006; <https://doi.org/10.1038/nature04979>.
29. Myers S, Freeman C, Auton A, Donnelly P, McVean G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet*. 2008; <https://doi.org/10.1038/ng.213>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

