*Bioinformatics*, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Applications Note

*Genome Analysis*

# Ribbon: Intuitive visualization for complex genomic variation

Maria Nattestad[1,*], Robert Aboukhalil[2], Chen-Shan Chin[3] and Michael C. Schatz[1,4]

[1]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA, [2]Invitae, San Francisco, California, USA, [3]DNAnexus, Mountain View, California, USA.,[4]Departments of Computer Science and Biology, Johns Hopkins University, Baltimore, MD, USA.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

## Abstract

**Summary:** Ribbon is an alignment visualization tool that shows how alignments are positioned within both the reference and read contexts, giving an intuitive view that enables a better understanding of structural variants and the read evidence supporting them. Ribbon was born out of a need to curate complex structural variant calls and determine whether each was well supported by long-read evidence, and it uses the same intuitive visualization method to shed light on contig alignments from genome-to-genome comparisons.

**Availability and Implementation:** Ribbon is freely available online at http://genomeribbon.com/ and is open-source at https://github.com/marianattestad/ribbon.

**Contact:** maria.nattestad@gmail.com

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

## 1 Introduction

Visualization has played an important role in the genomic revolution, enabling scientists to investigate variants, expression patterns, evolutionary changes, and a number of other relationships (Kent et al. 2002; Krzywinski et al. 2009; Robinson et al. 2011). Long-standing genome browsers have taken a very useful reference-based perspective, which is largely sufficient for representing short read alignments. However, advances in long-read sequencing (Sedlazeck et al. 2018) have shed light on examples where visualizing the portions of reads that maps to different parts of the reference is crucial. When a read is over 10 kbp long, it can cover large compound structural variations on several loci, but that context is lost using IGV (Robinson et al. 2011) or other reference-centric viewers without adding a read-based perspective and showing all relevant reference loci. In recent years, other visualization tools have tackled some challenging aspects of validating structural variant calls. In particular, SVplaudit (Belyeu et al. 2018) and SVCurator (Chapman et al. 2019) have enabled crowd-sourced curation of structural variants, and svviz (Spies et al. 2015) has used realignment against a putative structural variant allele to better evaluate the evidence. Even with these very useful approaches, more complex patterns of structural variation remain hard to represent. A more flexible approach can be helpful by showing the full picture of alignments from both the reference and read perspective, allowing researchers to explore and communicate about complex variation in a more intuitive way. The true power of long reads is hidden when the tools we use are built for short reads. We have addressed this problem by creating Ribbon (http://genomeribbon.com) an interactive online visualization tool that displays alignments along both reference and query sequences, along with variant calls, genes, and other genomic features.

## 2 Methods

Ribbon can load alignments from SAM (Li et al. 2009) and BAM files with long, short, or paired-end reads. While the main strengths of Ribbon are more apparent with long sequences, short and paired-ends reads are also supported and can benefit from the more structurally focused visualization Ribbon provides (**Supplementary Figure 9**). BAM files can be read from either a local file on the user's computer or from a remote server by entering the file's URL (Miller et al. 2014). To support parsing local BAM files, we compiled the command-line utility samtools (Li et

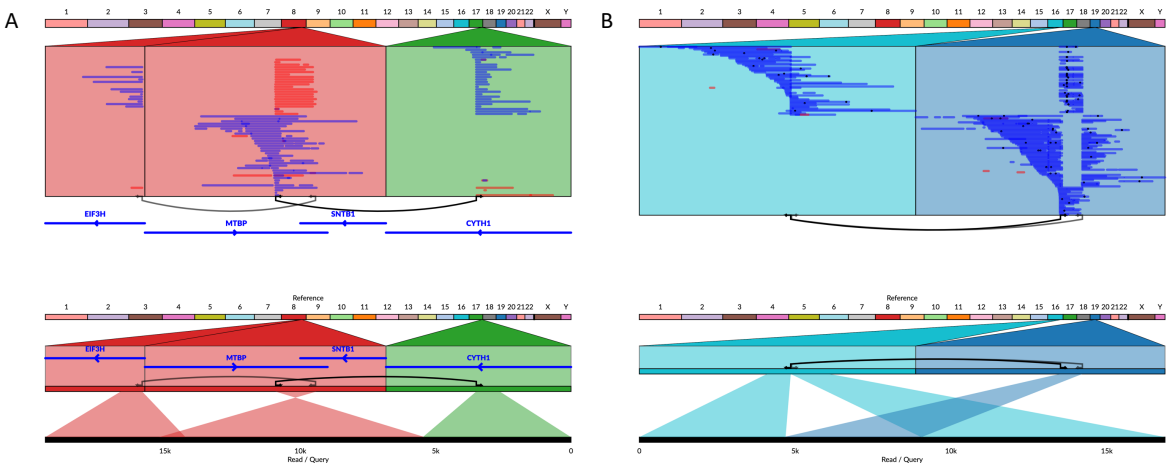


**Figure 1.** Ribbon shows all alignments from the perspective of both the reference and the read or query. Here long reads from SMRT sequencing show evidence for complex variants in the SK-BR-3 breast cancer genome. **(a)** The CYTH1-EIF3H gene fusion takes place through a series of two events that are captured within some of the long reads. **(b)** Long reads from SMRT sequencing show evidence of a full interchromosomal translocation where a sequence is homozygous deleted from chromosome 19 and homozygous inserted into chromosome 16.

al. 2009) from C to WebAssembly (Haas et al. 2017), which allows us to process BAM files efficiently inside a web browser.

In addition to long read alignment, Ribbon also has features for the related field of whole genome alignment visualization. First, we added support in Ribbon for visualizing genome-to-genome alignments, such as for comparing genomes of two related species or comparing a new genome assembly to an existing reference. This is done by supporting a simple tab-delimited, human-readable coordinate file format that can be created by the whole genome aligner MUMmer (Robinson et al. 2011; Kurtz et al. 2004) or by scripting from another file format. Second, we adopted a version of the dot plot visualization as an optional alternative to the classic ribbon visualization, since dot plots are often used in the genome assembly/comparison field. The ribbon and dot plot visualizations show the same coordinates in different ways, but most users have found the ribbon visualization more intuitive, while the dot plot remains a common visualization method in the field of whole genome alignment.

Finally, in addition to alignments with SAM, BAM, and general coordinate files, Ribbon can also show variants in VCF or BEDPE format, and genes or any other features in BED format. The user can select a variant to jump to that locus (or both loci for two-breakpoint variants) in the genome, and all the reads with alignments in the given region(s) will show up, including all their alignments to other places in the genome. This is particularly powerful because it pulls other relevant regions into focus; for example, in **Figure 1a**, the left-most locus is included automatically because it contains alignments of reads also found at the other two loci.

## 3 Results

By showing a synchronized read and reference perspective, Ribbon shows patterns in alignments of many reads across multiple chromosomes, while allowing detailed inspection of individual reads (**Supplementary Note 1**). For example, here we show a gene fusion in the SK-BR-3 breast cancer cell line linking the genes CYTH1 and EIF3H revealed by PacBio long read sequencing (Nattestad et al. 2018). While it has been found in the transcriptome previously (Chen et al. 2013; Kim and Salzberg 2011; Edgren et al. 2011), genome sequencing did not identify a direct chromosomal fusion between these two genes. Using long read sequencing, Ribbon shows that there are indeed reads that span from one gene to the other, going through not one but two variants, for the first time showing the genomic link between these two genes (Nattestad et al. 2018) (**Figure 1a**). More gene fusions of this cancer cell line are investigated with Ribbon in **Supplementary Note 2**. **Figure 1b** shows another complex event in this sample made simple in Ribbon: the translocation of a 4.4 kb sequence deleted from chr19 and inserted into chr16.

With the support for genome-genome alignments, Ribbon can also be used to test assembly algorithms or inspect the similarity between species. **Supplementary Note 4** shows a comparison of gorilla (Gordon et al. 2016) and human genomes using Ribbon, highlighting major structural differences.

## 4 Discussion

Ribbon enables understanding of complex variants, and it may also help in the detection of sequencing and sample preparation issues, testing of aligners and variant-callers, and rapid curation of structural variant candidates (**Supplementary Note 3**). Ribbon is a powerful and versatile visualization tool for investigating complex structural differences between any two genomes, using intuitive methods that take advantage of the rich context of long reads and contigs.

## Funding

This work was supported by NSF[DBI-1350041]; NHGRI[R01-HG006677]

*Conflict of Interest:* During the initial work, M.N. was a contractor and C.-S. C. was an employee and stockholder of Pacific Biosciences, a company commercializing DNA sequencing technologies. M.N. is currently an employee and stockholder of Google. C.-S. C. is currently an employee and stockholder of DNAnexus. R.A. is an employee and stockholder of Invitae.

## References


Belyeu, Jonathan R., Thomas J. Nicholas, Brent S. Pedersen, Thomas A. Sasani, James M. Havrilla, Stephanie N. Kravitz, Megan E. Conway, Brian K. Lohman, Aaron R. Quinlan, and Ryan M. Layer. 2018. "SV-Plaudit: A Cloud-Based Framework for Manually Curating Thousands of Structural Variants." GigaScience 7 (7). https://doi.org/10.1093/gigascience/giy064.

Chapman, Lesley M., Noah Spies, Patrick Pai, Chun Shen Lim, Andrew Carroll, Giuseppe Narzisi, Christopher M. Watson, et al. 2019. "SVCurator: A Crowdsourcing App to Visualize Evidence of Structural Variants for the Human Genome." BioRxiv, July. https://doi.org/10.1101/581264.

Chen, Ken, Nicholas E. Navin, Yong Wang, Heather K. Schmidt, John W. Wallis, Beifang Niu, Xian Fan, et al. 2013. "BreakTrans: Uncovering the Genomic Architecture of Gene Fusions." Genome Biology 14 (8): R87.

Edgren, Henrik, Astrid Murumagi, Sara Kangaspeska, Daniel Nicorici, Vesa Hongisto, Kristine Kleivi, Inga H. Rye, et al. 2011. "Identification of Fusion Genes in Breast Cancer by Paired-End RNA-Sequencing." Genome Biology 12 (1): R6.

Gordon, David, John Huddleston, Mark J. P. Chaisson, Christopher M. Hill, Zev N. Kronenberg, Katherine M. Munson, Maika Malig, et al. 2016. "Long-Read Sequence Assembly of the Gorilla Genome." Science 352 (6281): aae0344.

Haas, Andreas, Andreas Rossberg, Derek L. Schuff, Ben L. Titzer, Michael Holman, Dan Gohman, Luke Wagner, Alon Zakai, and J. F. Bastien. 2017. "Bringing the Web up to Speed with WebAssembly." ACM SIGPLAN Notices. https://doi.org/10.1145/3140587.3062363.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC." Genome Research 12 (6): 996–1006.

Kim, Daehwan, and Steven L. Salzberg. 2011. "TopHat-Fusion: An Algorithm for Discovery of Novel Fusion Transcripts." Genome Biology 12 (8): R72.

Krzywinski, Martin, Jacqueline Schein, Inanç Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. 2009. "Circos: An Information Aesthetic for Comparative Genomics." Genome Research 19 (9): 1639–45.

Kurtz, Stefan, Adam Phillippy, Arthur L. Delcher, Michael Smoot, Martin Shumway, Corina Antonescu, and Steven L. Salzberg. 2004. "Versatile and Open Software for Comparing Large Genomes." Genome Biology 5 (2): R12.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." Bioinformatics. https://doi.org/10.1093/bioinformatics/btp352.

Miller, Chase A., Yi Qiao, Tonya DiSera, Brian D'Astous, and Gabor T. Marth. 2014. "Bam.iobio: A Web-Based, Real-Time, Sequence Alignment File Inspector." Nature Methods 11 (12): 1189.

Nattestad, Maria, Sara Goodwin, Karen Ng, Timour Baslan, Fritz J. Sedlazeck, Philipp Rescheneder, Tyler Garvin, et al. 2018. "Complex Rearrangements and Oncogene Amplifications Revealed by Long-Read DNA and RNA Sequencing of a Breast Cancer Cell Line." Genome Research 28 (8): 1126–35.

Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. "Integrative Genomics Viewer." Nature Biotechnology 29 (1): 24–26.

Sedlazeck, Fritz J., Hayan Lee, Charlotte A. Darby, and Michael C. Schatz. 2018. "Piercing the Dark Matter: Bioinformatics of Long-Range Sequencing and Mapping." Nature Reviews. Genetics 19 (6): 329–46.

Spies, Noah, Justin M. Zook, Marc Salit, and Arend Sidow. 2015. "Svviz: A Read Viewer for Validating Structural Variants." Bioinformatics 31 (24): 3994–96.