

Ribbon: Visualizing complex genome alignments and structural variation

To the Editor: Visualization has played an extremely important role in the current genomic revolution to inspect and understand variants, expression patterns, evolutionary changes, and a number of other relationships¹⁻³. However, most of the information in read-to-reference or genome-genome alignments is lost for structural variations in the one-dimensional views of most genome browsers showing only reference coordinates. Instead, structural variations captured by long reads or assembled contigs often need more context to understand, including alignments and other genomic information from multiple chromosomes.

We have addressed this problem by creating Ribbon (genomeribbon.com) an interactive online visualization tool that displays alignments along both reference and query sequences, along with any associated variant calls in the sample. This way 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*. While it has been found in the transcriptome previously⁴⁻⁶, genome sequencing did not identify a direct chromosomal fusion between these two genes. After SMRT sequencing⁷, Ribbon shows that there are indeed long 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 (**Figure 1a**). More gene fusions of this cancer cell line are investigated 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 (**Figure 1b**). Thus, 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**).

In addition to SAM and BAM^{8,9} files with long, short, or paired-end reads, Ribbon can also load coordinate files from whole genome aligners such as MUMmer¹⁰. Therefore, Ribbon can be used to test assembly algorithms or inspect the similarity between species. **Supplementary Note 4** shows a comparison of gorilla¹¹ and human genomes using Ribbon, highlighting major structural differences. In conclusion, Ribbon is a powerful interactive web tool for viewing complex genomic alignments.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

M.N. created Ribbon with input from all authors. M.N. and M.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

C.-S. C. is an employee and stockholder of Pacific Biosciences, a company commercializing DNA sequencing technologies. M.N. was a contractor of Pacific Biosciences.

Maria Nattestad¹, Chen-Shan Chin², Michael C. Schatz^{1,3}

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. ²Pacific Biosciences, Menlo Park, California, USA. ³Departments of Computer Science and Biology, Johns Hopkins University, Baltimore, MD, USA. e-mail: mnattest@cshl.edu

1. Robinson, J.T., Thorvaldsdóttir, H. & Winckler, W. *Nature* (2011).
2. Kent, W.J., Sugnet, C.W., Furey, T.S. & Roskin, K.M. ... *research* (2002).
3. Krzywinski, M. et al. *Genome Research* **19**, 1639–1645 (2009).
4. Chen, K. et al. *Genome Biol.* **14**, R87 (2013).
5. Kim, D. & Salzberg, S.L. *Genome Biol.* **12**, R72 (2011).
6. Edgren, H. et al. *Genome Biol.* **12**, R6 (2011).
7. Eid, J. et al. *Science* **323**, 133–138 (2009).
8. Miller, C.A., Qiao, Y., DiSera, T., D'Astous, B. & Marth, G.T. *Nat Meth* **11**, 1189 (2014).
9. Down, T.A., Piipari, M. & Hubbard, T.J.P. *Bioinformatics* **27**, 889–890 (2011).
10. Kurtz, S. et al. *Genome Biol.* **5**, R12 (2004).
11. Gordon, D. et al. *Science* **352**, aae0344–aae0344 (2016).

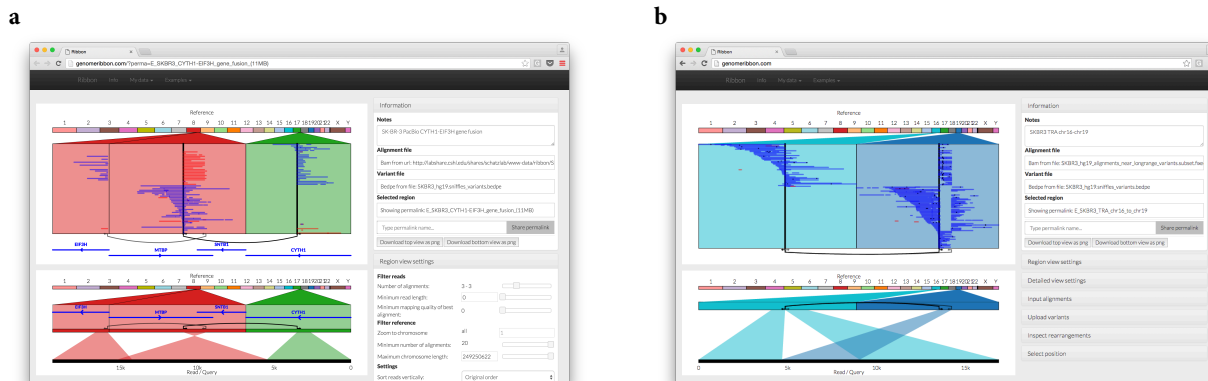


Figure 1 | Ribbon shows all alignments from the perspective of both the reference and the read or query. Here long reads from SMRT sequencing show evidence for complex variants in the SK-BR-3 breast cancer genome. (a) The *CYTH1-EIF3H* gene fusion takes place through a series of two events that are captured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