BrowserGenome.org: web-based RNA-seq data analysis and visualization

To the Editor: Applications of deep-sequencing technologies in life science research and clinical diagnostics are rapidly expanding. Although fast data-processing algorithms exist¹, intuitive, portable data-evaluation solutions are still needed. Web tools have a history in bioinformatics of providing platform-independent, intuitive, barrier-free software solutions. Whereas in most scientific web tools a server performs intense calculations, the new HTML5 standard and the competition between web browser platforms have recently opened access to computational resources for web apps. However, so far web apps have been used only to visualize existing genome annotations or alignment data^{2,3}. Here we describe BrowserGenome (<http://www.BrowserGenome.org>), a web-based deep-sequencing data-analysis platform offering barcode deconvolution, read mapping, real-time data visualization, transcript-count analysis and data normalization. BrowserGenome is specifically focused on the evaluation of mRNA-seq data, but it can easily be extended to other applications. BrowserGenome matches the speed and memory footprint of state-of-the-art software while being visually driven and intuitive to use.

Read-mapping, visualization and transcript-counting algorithms were implemented in JavaScript through adaptation of a non-overlapping *q*-gram indexing algorithm⁴, sorted data structures and random sampling³ (Supplementary Note 1 and Supplementary Figs. 1 and 2). The read-mapping strategy was specifically designed to allow quantification of gene expression in the limited web browser environment, without aims of splice-variant detection, calling of single-nucleotide polymorphisms or the evaluation of paired-end sequencing data, as offered by other software⁵. BrowserGenome uses raw sequencing data in FASTQ format or imports mapping results from other software in SAM format. It outputs binary or SAM-format mapping results or transcript-count tables. The graphical user interface displays the genome as a dynamic circle, with the mapping density displayed eccentrically (Fig. 1). The user navigates through

the data using a mouse, with gestures similar to those used in web applications such as Google Maps. Reference gene names and exons are displayed at high zoom levels. Up to six hit-density tracks can be loaded in parallel. Wizard menus guide users through the read-mapping and transcript-counting processes (Supplementary Note 2).

To validate the performance of BrowserGenome, we analyzed a publically available mRNA-seq data set from the ENCODE database⁶ (human HepG2 cells; data set ENCF000DPK) on a standard laptop computer. We observed that 59.2% of 26.6 million raw reads were mapped to the human genome at a rate of 18 million reads per hour. The hit-density map could be navigated in real time, and normalized transcript counts were calculated in less than two seconds (Supplementary Table 1). Despite BrowserGenome's simple read-mapping algorithm, analyzing the same data with the established STAR⁵ software produced highly correlated transcript-count data (Pearson $R = 0.974$; Supplementary Fig. 3) and near-equal correlation coefficients between gene expression results and sequencing-independent gene expression data (Supplementary Fig. 4).

BrowserGenome's usability and accessibility compare favorably with those of other graphics-based RNA-seq evaluation tools (Supplementary Fig. 5). The core functions can be easily extended or incorporated into other web apps through a library interface (Supplementary Note 3). The platform-independent web app does not transfer any scientific data via the Internet and is open-source software under the terms of GNU General Public License version 2 without depending on third-party code.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.3615).

ACKNOWLEDGMENTS

We thank H. Hauswedell (Institut für Informatik, Freie Universität Berlin, Berlin, Germany) for critical comments. This work was supported by grants from the German Research Foundation (EXC1023) and the European Research Council to V.H. and from the German National Academic Foundation to J.L.S.-B.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Jonathan L Schmid-Burgk & Veit Hornung

Institute of Molecular Medicine, University Hospital, University of Bonn, Bonn, Germany.

e-mail: jschmid@uni-bonn.de

Figure 1 | The BrowserGenome.org web application. The circular representation of the genome can be intuitively moved and zoomed with mouse or track pad gestures. Up to six tracks of deep-sequencing data can be displayed as concentric circles, and even large data sets can be visualized in real time.

- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Skinner, M.E., Uzulov, A.V., Stein, L.D., Mungall, C.J. & Holmes, I.H. JBrowse: a next-generation genome browser. *Genome Res.* **19**, 1630–1638 (2009).
- Miller, C.A., Qiao, Y., DiSera, T., D'Astous, B. & Marth, G.T. bam.iobio: a web-based, real-time, sequence alignment file inspector. *Nat. Methods* **11**, 1189 (2014).
- Ukkonen, E. Approximate string-matching with *q*-grams and maximal matches. *Theor. Comput. Sci.* **92**, 191–211 (1992).
- Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).