

SOFTWARE

Open Access



A graphical, interactive and GPU-enabled workflow to process long-read sequencing data

Shishir Reddy¹, Ling-Hong Hung², Olga Sala-Torra³, Jerald P. Radich^{3,4,5}, Cecilia CS Yeung^{3,6} and Ka Yee Yeung^{2*}

Abstract

Background: Long-read sequencing has great promise in enabling portable, rapid molecular-assisted cancer diagnoses. A key challenge in democratizing long-read sequencing technology in the biomedical and clinical community is the lack of *graphical* bioinformatics software tools which can efficiently process the raw nanopore reads, support graphical output and *interactive* visualizations for interpretations of results. Another obstacle is that high performance software tools for long-read sequencing data analyses often leverage graphics processing units (GPU), which is challenging and time-consuming to configure, especially on the cloud.

Results: We present a graphical cloud-enabled workflow for fast, interactive analysis of nanopore sequencing data using GPUs. Users customize parameters, monitor execution and visualize results through an accessible graphical interface. The workflow and its components are completely containerized to ensure reproducibility and facilitate installation of the GPU-enabled software. We also provide an Amazon Machine Image (AMI) with all software and drivers pre-installed for GPU computing on the cloud. Most importantly, we demonstrate the potential of applying our software tools to reduce the turnaround time of cancer diagnostics by generating blood cancer (NB4, K562, ME1, 238 MV4;11) cell line Nanopore data using the Flongle adapter. We observe a 29x speedup and a 93x reduction in costs for the rate-limiting basecalling step in the analysis of blood cancer cell line data.

Conclusions: Our interactive and efficient software tools will make analyses of Nanopore data using GPU and cloud computing accessible to biomedical and clinical scientists, thus facilitating the adoption of cost effective, fast, portable and real-time long-read sequencing.

Keywords: Cancer diagnostics, Workflows, Cloud computing, Nanopore, GPU, FAIR, Long-read sequencing, Leukemia

Background

Cancer is a leading cause of mortality worldwide with the highest burden of death affecting lower- and middle-income countries [1]. Delays in medical care from the inability to detect cancer earlier are a key component contributing to higher morbidity, poor response to treatments and lower survival [2]. Advances in molecular

diagnosis have enabled detection of specific driver mutations that can be essential for prognosis, monitoring, and targeted therapy [3–5]. Examples of such “precision medicine” include the BCR-ABL fusion gene in chronic myeloid leukemia (CML), PML-RARA fusions in acute promyelocytic leukemia (APL) and FLT3 mutations in acute myeloid leukemia (AML) [6–9]. Potentially targetable mutations are also found in solid tumors, such as renal cell carcinoma [10–12]. Currently, detection of fusion genes by chromosomal analysis requires highly specialized laboratories. Chromosomal analyses may not

* Correspondence: kayee@uw.edu

²School of Engineering and Technology, University of Washington, 98402 Tacoma, WA, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. 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 in a credit line to the data.

have the precision necessary to identify the specific breakpoint in a patient or provide the sequence of the flanking segments around the fusion gene to allow for downstream development of patient specific monitoring assays. Routine PCR based assays can be rapid, but a priori knowledge of the fusion breakpoints is required and when fusions involve large intronic regions RNA input is generally needed. Thus, turn-around times for these methods can take three days to two weeks [13, 14]. To capitalize on the potential of precision medicine, faster analysis of sequencing data is needed to improve the potential of molecular-assisted cancer diagnoses [15].

For cancer management, next generation sequencing (NGS) has many limitations, such as phasing errors, mis-mapping from short reads, strand bias and amplification errors causing irregular variant allele frequency, and bioinformatically-challenging repetitive sequences [16, 17]. In contrast, long-read sequencing technology, such as Oxford Nanopore Technologies (ONT), generates continuous sequences up to a few megabases in length at this time [18]. Nanopore sequencing provides both sequencing and phasing information because the read lengths are generally very long in comparison to NGS (NGS = 150 to 250 bp, nanopore = generally > 2 kb to 200,000 kb or longer), therefore it will not incur the same artifacts and mapping errors [19]. Unlike NGS, which takes an average of three days to complete sample processing and library preparation plus one additional day for sequencing, nanopore sequencing can directly sequence DNA, resulting in much shorter turnaround times [20, 21]. Thus, long-read sequencing technologies hold promise in overcoming the current diagnostic gap in cancer research.

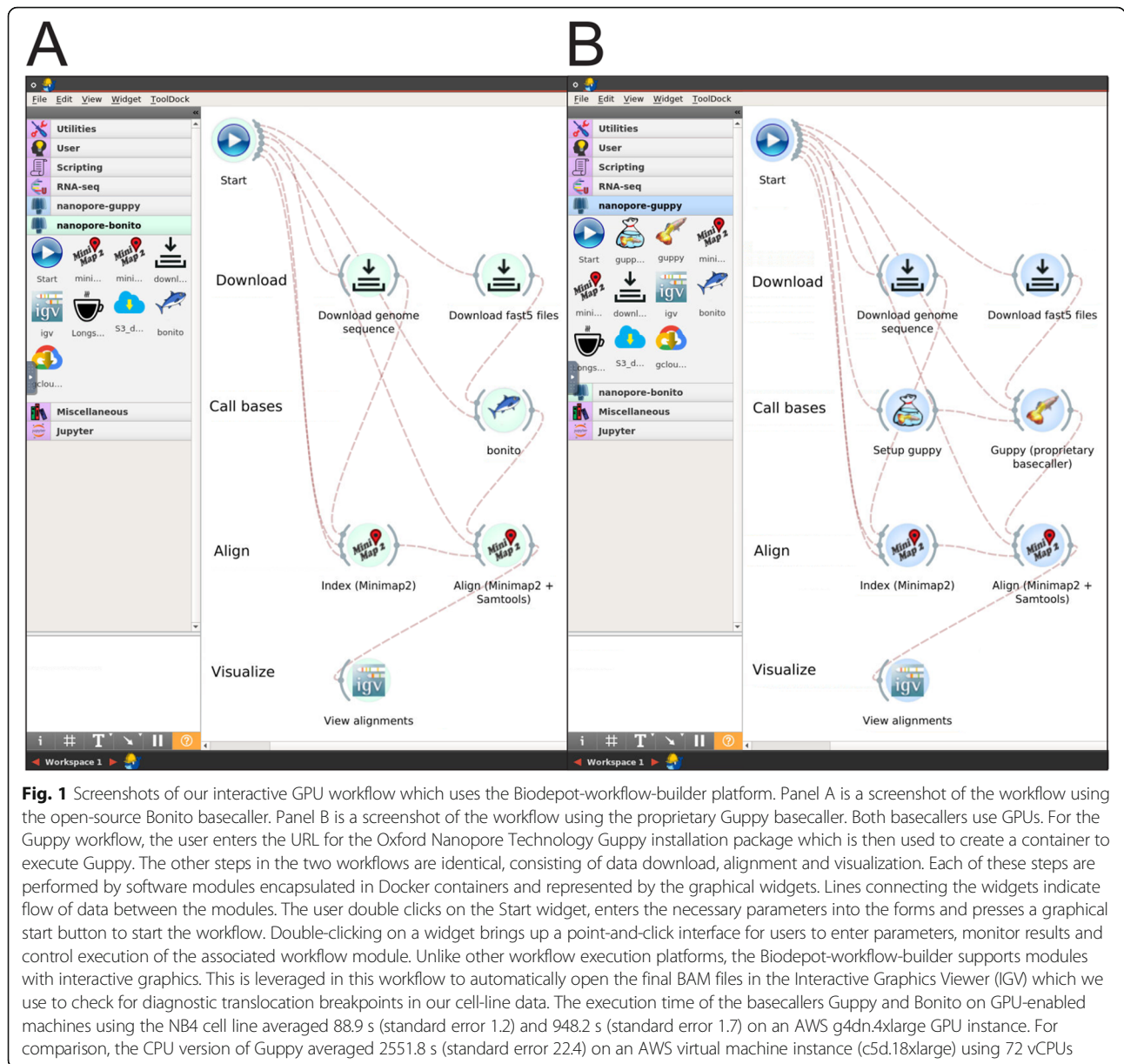
Computational methods and software tools tailored for long-read sequencing data are essential to enable use of this emerging and promising technology [22, 23]. In nanopore sequencing, electrical current alterations are recorded as different bases traverse the pore opening. Basecalling, which translates the signal (stored as fast5 files) into a sequence of base pairs is the key step determining accuracy of the sequencing experiment. The initial conversion is followed by error correction and data polishing to obtain the final sequence [23]. Basecalling is computationally expensive and a rate-limiting step in the analysis of nanopore data. In contrast, for NGS data, aligning reads is the rate limiting step due to the greater number of reads and the absence of a separate basecalling step. Deep learning neural network models have been applied to basecalling to increase the accuracy [24]. With standard CPU processing, these methods are prohibitively slow and require large numbers of computational cores operating in parallel to be practical. Graphics processing units (GPUs) can be used to accelerate the analysis but require specialized hardware and

software. Hardware in the form of GPU instances are available on public cloud services such as Amazon Web Services (AWS). However, virtual machine instances do not come with the drivers or Compute Unified Device Architecture (CUDA) libraries installed. In addition, the versions of these drivers and libraries must be carefully matched to the software being executed.

Nanopore sequencing is fast and cost effective in terms of data collection, with a \$90 USB Flongle attached to a laptop supporting real-time sequencing. Sequencing is potentially accessible to a broad range of biomedical scientists who do not have access to a traditional sequencer. A challenge in democratizing long-read sequencing technology is the difficulty of the processing of the data which requires command line tools and technically difficult installation of software, libraries and drivers. There is a lack of *graphical* bioinformatics software tools which can efficiently process the raw nanopore reads, and *interactive* visualizations for interpretations of results. An example of a command-line workflow is the MasterOfPores pipeline that performs pre-processing and analysis (prediction of RNA modifications and estimation of polyA tail lengths) of long-read data [25]. MasterOfPores is a workflow using the NextFlow framework [26], a script-based engine that requires programming experience to deploy and modify. While MasterOfPores uses software containers for most of its components, it does not provide a container for the key basecalling step which requires the most setup and configuration to operate with GPU computing. In addition, MasterOfPores does not include the product-grade basecaller Guppy [27], which is available to ONT customers via their community site [28] and cannot be distributed in a container.

Implementation

We present a graphical cloud-enabled containerized workflow for fast, interactive analysis of nanopore data using GPUs. Specifically, we extended the Biodepot-workflow-builder (Bwb) [29] to provide a modular and easy-to-use graphical interface that allows users to create, customize, execute, and monitor bioinformatics workflows. Figure 1 shows screenshots of the platform. The workflow consists of modules to download the data and genome files, basecallers Guppy [27] or Bonito [30], minimap2 [31] for sequence alignment, and the Integrated Genome Viewer (IGV) [32, 33] for visualization of the BAM files. We provide containers for both the ONT proprietary Guppy [27] and the ONT open-source Bonito [30] basecallers. In accordance with the licensing constraints of Guppy, we provide a containerized setup module that creates the Guppy container locally when the user provides the download URL from the ONT community site. To facilitate deployment, we use Docker



containers for all modules (including the Bwb platform) and provide an Amazon Machine Image (AMI) with all software and drivers pre-installed for GPU computing on the cloud.

Integrated and interactive workflow in the Biodepot-workflow-builder (Bwb)

Biodepot-workflow-builder (Bwb) [29] provides a modular and easy-to-use graphical interface for reproducible execution, customization and interactive visualization of the nanopore pipeline. Graphical widgets representing Docker containers that execute modular tasks are graphically linked to define bioinformatics workflows that can then be reproducibly deployed across different local and cloud

platforms. New widgets (modules) can be added without writing code. In this work, we created new widgets in the Bwb for basecalling using Guppy [27] and Bonito [30], alignment using minimap2 [31], visualization of the resulting BAM files using the Integrated Genome Viewer (IGV) [32, 33]. Each of these widgets call a Docker container in the backend. Users can adjust input parameters of each widget using an intuitive form-based user interface, and check intermediate results using a console. A key characteristic of Bwb is integrated support for graphical output, enabling *interactive* tools such as Jupyter notebooks, spreadsheets, and visualization tools to be included in the workflow. We leverage this graphical output support feature of the Bwb to create an integrated, interactive

workflow by integrating the IGV to visualize resulting BAM files. Our workflows can be shared using a Bwb native format or exported as shell scripts. In addition, the Bwb is distributed as a Docker container that can be easily deployed on *any local or cloud platform*.

GPU software versioning and compatibility ensured by using containers and providing an AMI with drivers and software

GPU-enabled executables require additional software layers to use the GPU hardware. For NVIDIA hardware, CUDA software provided by the manufacturer is necessary to perform general computations on the GPU. AMD GPUs use a different and incompatible set of software for their hardware. In addition, low level libraries (drivers) are required to communicate with the GPU card. Additional language and operating system dependent libraries and headers may also be required to integrate the CUDA software. All these layers of software interact with each other and as a result, compatibility is version dependent and even sensitive to the method of installation. Components installed using scripts may be incompatible with components installed using package managers.

An example of the complexities involved in deploying GPU software is the open-source Bonito basecaller. The current version of the Bonito caller will not install if one follows the instructions on the Github [30]. This is due to dependency incompatibilities in the current Python PyTorch [34] packages and CuPy [35] libraries. We were able to install Bonito v0.38 by downgrading from CUDA 11.3 to CUDA 10.2, cuDNN 7, CuPy 10.2. and downgrading PyTorch from 1.8.0 to 1.7.1. By providing a Docker container we can ensure that Bonito is deployed in this compatible environment without the need for the user to install the exact versions of the software. However, this is not sufficient, the user still needs to install the compatible drivers and CUDA software on their host machine. This is true even for GPU instances on Amazon cloud. AWS does provide the basic Linux operating systems but requires that users install their own drivers and libraries. There are commercial distributions on the AWS marketplace that provide support, but we could find nothing among the free community offerings. As a resource for the nanopore research community, we therefore have provided a public freely available AMI for use with AWS GPU virtual machine instances with versions of drivers and libraries that we have tested with our workflows. The combination of AMI and containerized modules eliminates the arcane installation steps and makes the nanopore GPU software accessible to a broad audience in the biomedical and clinical community.

Results

Data generation using cell lines

5,000 ng of DNA from four cell lines (NB4, K562, ME1, MV4;11) was dephosphorylated and cut with Cas9 enzyme complexed with RNA guides designed to target the genes involved in the translocations present in each line (BCR and ABL1 for K562, PML and RARA for NB4, CBFB and MYH11 for ME1, KMT2A and AFF1 for MV4;11). Nanopore adapters were then ligated to the newly created DNA ends and the library was loaded onto a flow cell or Flongle and sequenced using a MinION Mk1B.

Detection of fusion genes in blood cancer cell lines

Applications of nanopore sequencing coupled with our workflows in cancer diagnostics are shown in Fig. 2. We can reliably detect fusion genes from DNA sequences using cell lines (NB4, K562, ME1, and MV411) with known fusion genes. This provides an advancement to molecular diagnostics by its ability to detect specific breakpoints even if there is a large intronic region between the fusion genes of interest. DNA sequencing of NB4 on a comparatively low cost (\$90) Flongle nanopore device, (Oxford, UK), confirmed fusion gene sequencing spanning the *PML* and *RARA* genes (Fig. 2a). DNA sequencing of K562 on a Flongle was able to capture the fusion gene spanning the *BCR* and *ABL1* genes, as well as capturing the intervening large *ABL1* intronic region 1 where the specific breakpoint occurs (Fig. 2b). Figure S1 in Additional file 1 presents additional results for the detection of fusion genes CBFB and MYH11 for the cell line ME1 (Figure S1A), KMT2A and AFF1 for MV4;11 (Figure S1B).

Benchmarking basecallers

We compared the execution time of the basecallers Guppy [27] and Bonito [30] on GPU-enabled machines using the NB4 cell line data. We also measured the execution time of running Guppy using CPU and on a local host. The runtime of Guppy is reduced from over 42 min to just over 1 min using GPU computing, representing a *29x speedup and a 93x reduction in cloud computing costs*. The results are summarized in Table 1.

Experimental setup for benchmarking

Guppy GPU was benchmarked on an AWS g4dn.4xlarge virtual machine instance with a NVIDIA Tesla T4 GPU with the template_r9.4.1_450bps_hac.jsn model file provided by ONT. Bonito GPU was also benchmarked on the same instance using the provided dna_r9.4.1 model file and the default settings (chunk size of 4000 and batch size of 32). Guppy CPU was benchmarked on a c5d18xlarge instance with 72 vCPUs, 72 threads/basecaller, and 1 basecaller. The Guppy GPU experiments on a

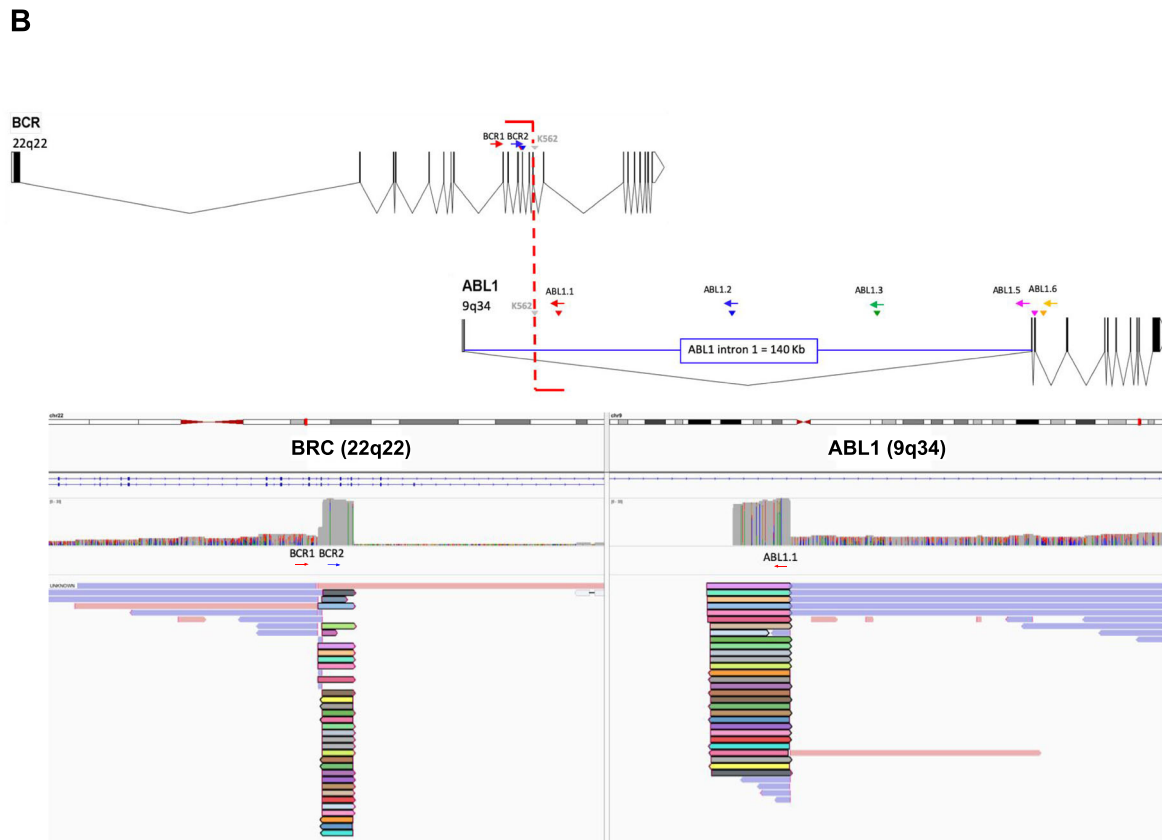
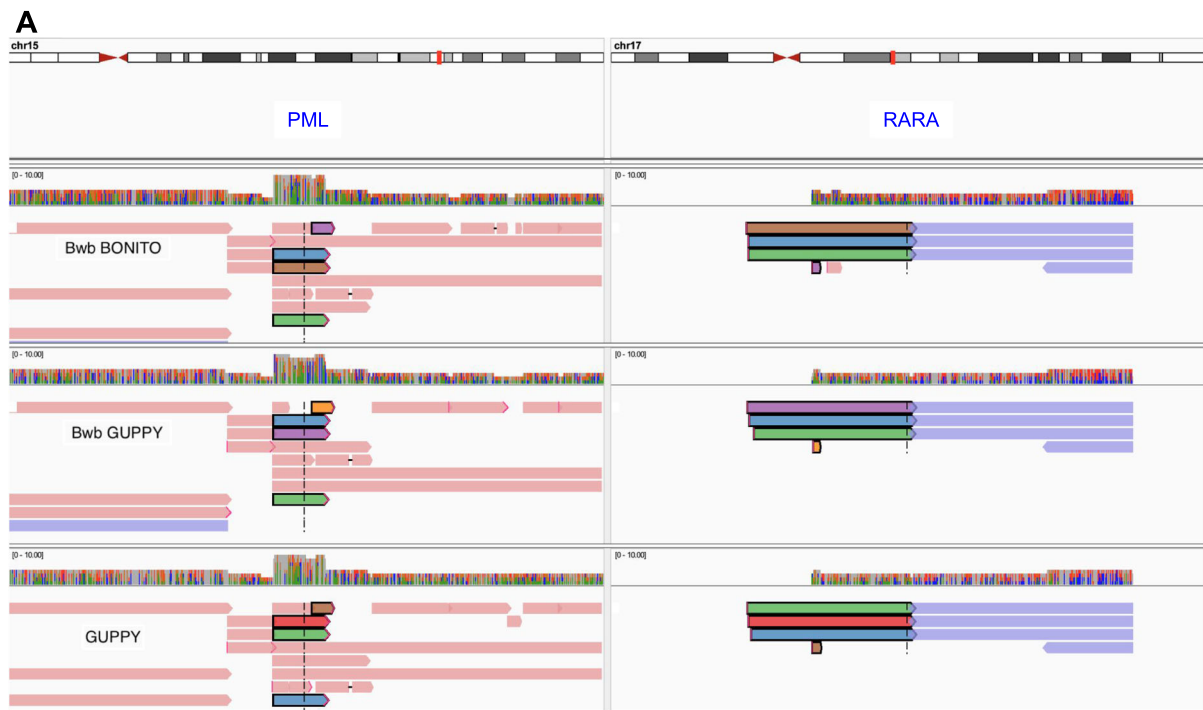


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 A. IGV viewer alignment on the PML and RARA genes of a Flongle Nanopore generated sequences for the NB4 cell line. Library was generated from DNA with a PCR-free enrichment protocol using CRISPR guides targeting PML and RARA genes. The top panel shows the alignment for the reads processed with the Bonito basecaller and minimap2 aligner in the Bwb. The middle panel shows the alignment of reads processed with the Guppy basecaller and minimap2 aligner in the Bwb workflow. The bottom panel shows reads processed in a manual step-by-step workflow using the Guppy flipflop basecaller and minimap2 aligner. Reads with PML-RARA breakpoint are colored to highlight the fragment aligned to PML and RARA. **B.** Genomic BCR-ABL1 breakpoint identified in the K562 cell line by long-read sequencing. Schematic representation (generated with <http://wormweb.org/exonintron>) shows the breakpoint captured with our amplification-free enrichment protocol and long-read sequencing. The breakpoint is represented in the upper graphic by the red vertical line, and the location of the sequence specific guides is marked by colored arrows. ABL1 intron 1 spans 140Kbs. In the lower panel, nanopore sequence alignments in IGV show sequences partially aligned to BCR and ABL1. Reads with BCR-ABL1 breakpoint are colored to highlight the same read is partially aligned to BCR and ABL1.

local host were performed on a laptop with a GeForce RTX 2060 GPU. All benchmark experiments on AWS were based on 4 runs. We observed that Guppy GPU achieved the fastest average time at 88.9 s (1.5 min) with standard error of 1.2 s. Bonito GPU finished basecalling in 948.2 s (15.8 min) on average with standard error of 1.7 s. Guppy CPU achieved the slowest average time at 2551.8 s (42.5 min) with standard error of 22.4 s. The 29x speedup is computed by comparing the average run-time (in seconds) of Guppy CPU to Guppy GPU ($2551.8/88.9 = 28.7$).

Estimation of costs of basecalling step

A large selection of virtual machine instance types with different pricing structures are available on AWS [36]. We conducted our empirical experiments using the C5 and G4 EC2 (Elastic Compute Cloud) instances, designed for compute-intensive workloads and GPU computing respectively. In the us-east-2 region, the on-demand pricing of the AWS c5d.18xlarge EC2 instance, with 72 vCPUs and 144GB memory, is \$3.888 per hour. The pricing for a g4dn.4xlarge, with single GPU, 16 vCPUs and 64GB memory, is \$1.204 per hour [36]. The ratio of the costs of CPU vs. GPU is the time ratio 2551.8/88.9 multiplied by the pricing ratio 3.888/1.204 which works out to be 92.7-fold cheaper when the GPU instance is used for basecalling. These cost estimates are based on single samples.

Conclusions

A potential advantage of long-read sequencing is lower costs in comparison to standard NGS. We used low-cost Flongles to detect the PML-RARA and BCR-ABL1

fusions. In this work, we demonstrate the ability to detect fusions on nanopores devices that read DNA at speeds faster than 1 nt/ μ s. We present interactive software tools that not only make analyses of Nanopore data accessible to biomedical and clinical scientists, but also efficient and economical through the use of GPU computing. Most importantly, we have illustrated the applicability of our workflow for the analysis of cell line data as part of a rapid, cost effective assay to detect fusions. Using an intuitive graphical interface, our workflow integrates the processing of raw nanopore reads, with the visualization steps that are used to interpret the results. This provides the capability to identify and confirm pathognomonic fusion genes such as BCR-ABL1 in CML or PML-RARA in APL. This methodology can potentially both define specific breakpoints important in treating and tracking a patients' specific disease, but also allow multiplex phasing to identify and follow multiple mutations in the same patient. Future directions include optimizing the assay for faster library preparation, sequencing times, and analysis to enable fusion detection to less than a day. Improvements to turnaround time in the laboratory combined with an accessible, efficient informatics workflow that enables most molecular technologists and pathologists to implement long-read sequencing into current clinical pathology workflows will advance the field of molecular pathology beyond what is currently possible with NGS. These small portable, low-cost devices, together with integrated bioinformatics support, will allow for rapid diagnostics to assist in point-of-care clinical decision making.

Availability and requirements

Project name: Nanopore-GPU.

Table 1 Comparison of runtime from different basecallers (Guppy and Bonito) using the NB4 cell line. The AWS results were averaged over 4 runs. The local host results were averaged over 5 runs

Basecaller	cloud/local	average runtime (seconds)	standard error (seconds)
Guppy CPU	AWS c5d18xlarge	2551.8	22.4
Guppy GPU	AWS g4dn.4xlarge	88.9	1.2
Guppy GPU	Laptop	135.3	0.6
Bonito GPU	AWS g4dn.4xlarge	948.2	1.7

Project home page: All code is publicly available and distributed under a custom academic license at <https://github.com/BioDepot/nanopore-gpu>. A public AMI (ami-0ecb1effab7fcfaa3) is provided with all necessary drivers and software pre-installed.

Operating system(s): Platform independent.

Programming language: Python.

Other requirements: Docker.

License: Custom non-commercial license.

Any restrictions to use by non-academics: The software is used solely for noncommercial purposes.

Abbreviations

AMI: Amazon Machine Image; AML: acute myeloid leukemia; APL: acute promyelocytic leukemia; AWS: Amazon Web Services; Bwb: Biodepot-workflow-builder; CUDA: Compute Unified Device Architecture; EC2: Elastic Compute Cloud; GPU: Graphics Processing Unit; IGV: Integrated Genome Viewer; NGS: Next-generation sequencing; ONT: Oxford Nanopore Technologies

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07927-1>.

Additional file 1.

Acknowledgements

We would like to acknowledge Katherine Melville from Oxford Nanopore Technologies and cloud credits from Amazon Web Services.

Authors' contributions

Conceptualization: J.R., C.Y. and K.Y.Y. Software development and testing: L.H.H. and S.R. Code documentation: S.R. Cloud resources: L.H.H. Computational benchmarking: L.H.H. and S.R. Data generation: O.S. and C.Y. Data analysis: O.S. and S.R. Figure preparation: L.H.H., O.S. and S.R. Writing - original draft preparation: K.Y.Y. and C.Y. Writing - reviewing and editing: All authors. Funding acquisition: J.R., C.Y. and K.Y.Y. The authors read and approved the final manuscript.

Funding

LHH and KYY are supported by NIH grant R01GM126019. JR is supported by NIH grants R01 CA175008-06 and UG1 CA233338-02. CY is supported by NCCN Young Investigator Award and Hyundai Hope on Wheel Scholars Hope Grant. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

The content is solely the responsibility of the authors, and does not necessarily represent the official views of the National Institutes of Health.

Availability of data and materials

Code and data for the workflows are available in the Github repository <https://github.com/BioDepot/nanopore-gpu>. The cell line Nanopore data are included in the workflow, with direct links from the README page. We have also created a demo video on Youtube <https://youtu.be/yPhBkjd8iGy>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

LHH and KYY also have equity interest in and employed by Biodepot LLC, which receives compensation from NCI SBIR contract number

75N91020C00009. The terms of this arrangement have been reviewed and approved by the University of Washington in accordance with its policies governing outside work and financial conflicts of interest in research.

Author details

¹University of California, 92697 Irvine, CA, USA. ²School of Engineering and Technology, University of Washington, 98402 Tacoma, WA, USA. ³Clinical Research Division, Fred Hutchinson Cancer Research Center, 98109 Seattle, WA, USA. ⁴Clinical Research Division, Kurt Enslin Endowed Chair, Fred Hutchinson Cancer Research Center, 98109 Seattle, WA, USA. ⁵Department of Medicine, University of Washington, 98109 Seattle, WA, USA. ⁶Department of Laboratory Medicine and Pathology, University of Washington, 98109 Seattle, WA, USA.

Received: 27 May 2021 Accepted: 10 August 2021

Published online: 23 August 2021

References

1. The Lancet. GLOBOCAN 2018: counting the toll of cancer. *Lancet*. 2018; 392(10152):985.
2. Milner DA Jr, Holladay EB. Laboratories as the Core for Health Systems Building. *Clin Lab Med*. 2018;38(1):1–9.
3. Mehta S, Shelling A, Muthukaruppan A, Lasham A, Blenkiron C, Laking G, Print C. Predictive and prognostic molecular markers for cancer medicine. *Ther Adv Med Oncol*. 2010;2(2):125–48.
4. Yeung CCS, Radich J. Predicting Chemotherapy Resistance in AML. *Curr Hematol Malig Rep*. 2017;12(6):530–6.
5. Yaghmaie M, Yeung CC. Molecular Mechanisms of Resistance to Tyrosine Kinase Inhibitors. *Curr Hematol Malig Rep*. 2019;14(5):395–404.
6. Radich J, Yeung C, Wu D. New approaches to molecular monitoring in CML (and other diseases). *Blood*. 2019;134(19):1578–84.
7. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010;115(3):453–74.
8. O'Donnell MR, Tallman MS, Abboud CN, Altman JK, Appelbaum FR, Arber DA, Attar E, Borate U, Coutre SE, Damon LE, et al. Acute myeloid leukemia, version 2.2013. *J Natl Compr Canc Netw*. 2013;11(9):1047–55.
9. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391–405.
10. Xiao X, Garbutt CC, Hornicek F, Guo Z, Duan Z. Advances in chromosomal translocations and fusion genes in sarcomas and potential therapeutic applications. *Cancer Treat Rev*. 2018;63:61–70.
11. Tretiakova MS, Wang W, Wu Y, Tykodi SS, True L, Liu YJ. Gene fusion analysis in renal cell carcinoma by FusionPlex RNA-sequencing and correlations of molecular findings with clinicopathological features. *Genes Chromosomes Cancer* 2019.
12. Parker BC, Zhang W. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chin J Cancer*. 2013;32(11):594–603.
13. Tsongalis GJ, Al Turkmani MR, Suriawinata M, Babcock MJ, Mitchell K, Ding Y, Scicchitano L, Tira A, Buckingham L, Atkinson S, et al. Comparison of Tissue Molecular Biomarker Testing Turnaround Times and Concordance Between Standard of Care and the Biocartis Idylla Platform in Patients With Colorectal Cancer. *Am J Clin Pathol*. 2020;154(2):266–76.
14. Dawson AJ, McGowan-Jordan J, Chernos J, Xu J, Lavoie J, Wang JC, Steinrath M, Shetty S. Canadian College of Medical Geneticists guidelines for the indications, analysis, and reporting of cancer specimens. *Curr Oncol*. 2011;18(5):e250–5.
15. VanderLaan PA, Chen Y, DiStasio M, Rangachari D, Costa DB, Heher YK. Molecular Testing Turnaround Time in Non-Small-Cell Lung Cancer: Monitoring a Moving Target. *Clin Lung Cancer*. 2018;19(5):e589–90.
16. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. 2017;19(3):341–65.
17. Strom SP, Lee H, Das K, Vilain E, Nelson SF, Grody WW, Deignan JL. Assessing the necessity of confirmatory testing for exome-sequencing

- results in a clinical molecular diagnostic laboratory. *Genet Med.* 2014;16(7):510–5.
18. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif.* 2015;3:1–8.
 19. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol.* 2018;36(4):338–45.
 20. Helmersen K, Aamot HV. DNA extraction of microbial DNA directly from infected tissue: an optimized protocol for use in nanopore sequencing. *Sci Rep.* 2020;10(1):2985.
 21. Cumbo C, Minervini CF, Orsini P, Anelli L, Zagaria A, Minervini A, Coccaro N, Impera L, Tota G, Parciante E, et al: Nanopore Targeted Sequencing for Rapid Gene Mutations Detection in Acute Myeloid Leukemia. *Genes (Basel)* 2019, 10(12).
 22. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet.* 2018;19(6):329–46.
 23. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 2020; 21(1):30.
 24. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 2019;20(1):129.
 25. Cozzuto L, Liu H, Prysycz LP, Pulido TH, Delgado-Tejedor A, Ponomarenko J, Novoa EM. MasterOfPores: A Workflow for the Analysis of Oxford Nanopore Direct RNA Sequencing Datasets. *Front Genet.* 2020;11:211.
 26. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol.* 2017;35(4):316–9.
 27. Oxford Nanopore Technologies GitHub: Guppy [<https://github.com/nanoporetech>].
 28. The Nanopore Community [<https://nanoporetech.com/community>].
 29. Hung LH, Hu J, Meiss T, Ingersoll A, Lloyd W, Kristiyanto D, Xiong Y, Sobie E, Yeung KY. Building Containerized Workflows Using the BioDepot-Workflow-Builder. *Cell Syst.* 2019;9(5):508–14 e503.
 30. Bonito. A PyTorch Basecaller for Oxford Nanopore Reads [<https://github.com/nanoporetech/bonito>].
 31. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100.
 32. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24–6.
 33. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178–92.
 34. PyTorch [<https://pytorch.org/>].
 35. CuPy. A NumPy-compatible array library accelerated by CUDA [<https://cupy.dev/>].
 36. Amazon. EC2 pricing <https://aws.amazon.com/ec2/pricing/>.

Publisher's Note

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

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

