Malays J Pathol 2025; 47(1) : 135 – 138

LETTER TO EDITOR

Evaluation of variant calling methods of sequencing data for *BCR::ABL1* kinase domain mutation detection

Zahidah ABU SEMAN[1,3], Fadly AHID[3], Nor Rizan KAMALUDDIN[1], Ermi Neiza MOHD SAHID[1], Ezalia Esa[1], Siti Shahrum MUHAMED SAID[2], Norazlina AZMAN[2], Wan Khairull Dhalila WAN MAT[1], Julia ABDULLAH[1], Nurul Aqilah ALI[1] and Yuslina MAT YUSOFF[1]*

[1]Haematology Unit, Cancer Research Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, 40170 Shah Alam, Selangor, Malaysia; [2]Department of Pathology, Hospital Tunku Azizah, Ministry of Health Malaysia, 50300 Kuala Lumpur, WP Kuala Lumpur, Malaysia; [3]Centre for Medical Laboratory Technology Studies, Faculty of Health Sciences, Universiti Teknologi MARA, 42300 Puncak Alam, Selangor, Malaysia

Dear Editors,

Despite remarkable achievement of tyrosine kinase inhibitors (TKIs) therapy, 20–30% of chronic myeloid leukaemia patients developed primary or secondary resistance to treatment during the disease course.[1] The acquisition of a point mutation in the *BCR::ABL1* kinase domain (KD) is the most common cause of treatment failure.[2-4] The use of next-generation sequencing (NGS) is advised for accurately determining the mutation status of *BCR::ABL1* KD, particularly in cases where the allele frequency is low, and for identifying mutations across multiple exons simultaneously compared to Sanger Sequencing (SS).[5-8] An increasing number of software tools or algorithms for the detection of somatic mutations have been developed for the analysis of sequencing data. Scientific reports have indicated that somatic variant callers have different levels of sensitivity.[9,10]

In the present study, three variants caller, BCFtools, VarScan 2, and GATK Mutect2 were applied on the sequencing data from 85 TKI-resistant patients. These tools were selected for their demonstrated performance, accuracy, and support from previous studies.[11-18] Each variant caller has distinct strengths regarding sensitivity and specificity. The GATK Mutect2 is designed for somatic variants, employing local assembly and realignment techniques to detect single nucleotide variants (SNVs) and insertions/deletions (indels).[11] It is highly regarded for its best practices in variant discovery from high-throughput sequencing data.[12-13] Varscan 2 is a computational method used to detect somatic mutations, loss of heterozygosity (LOH), and copy number alteration (CNA) events in pairs of tumour and normal samples.[14-16] BCFtools comprises a collection of tools designed to manipulate variant calls within the Variant Call Format (VCF) and its binary counterpart BCF.[17] It is widely cited and used within the scientific community, provides a level of validation and reproducibility.[18] To detect point mutations in *BCR::ABL1* KD, a sequential process was developed as illustrated in FIG. 1, encompassing the subsequent stages: i) assessment of sequencing reads quality; ii) alignment to Reference Genome; iii) identification of mutations using three-variant caller; and iv) annotation of variants. Single-nucleotide variants (SNVs) with variant allele frequency (VAF) greater than 3% were identified for further analysis.

The variants detected through the three-algorithm strategy were compared with those reported in the COSMIC (Catalogue Of Somatic Mutations In Cancer) database for *BCR::ABL1* gene (https://cancer.sanger.ac.uk/cosmic).[19] Synonymous variants listed in the COSMIC database were excluded, as these variations are not pertinent to TKI resistance. In total, the present study identified 10 missense kinase domain mutations relevant to TKI resistance (FIG. 2) and these mutations were confirmed using SS. Interestingly, NGS analysis revealed four additional low-frequency mutations (3% to 15% VAF) in four TKI resistant patients. Moreover, eight mutations: Y253H, E255K, T315I, F317L, K357T, F359V, F359C and E453K, were identified by more than one variant calling tool.

The results of this study show that BCFtools detected all variants identified by SS. VarScan 2

**Address for correspondence:* Yuslina Mat Yusoff, Haematology Unit, Cancer Research Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, 40170 Shah Alam, Selangor, Malaysia. Tel: +603-3627872; Email: *yuslina.my@moh.gov.my* (Mat Yusoff Y)

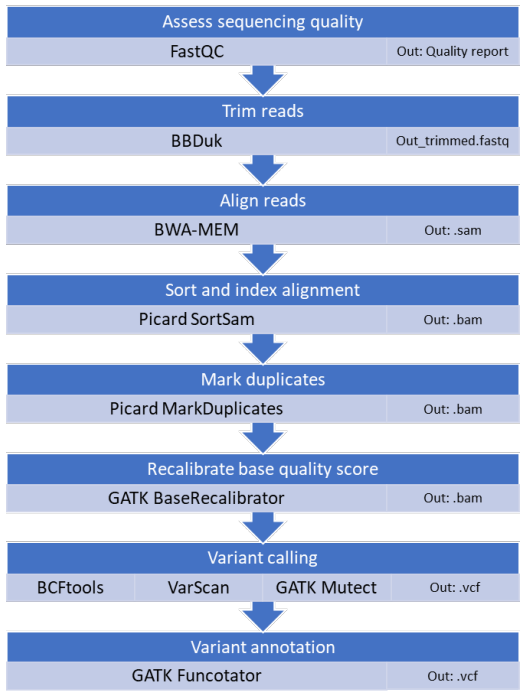

FIG. 1: Bioinformatics workflow for analysis of sequence data

Mutation	Sample ID	NGS detection			SS detection
		Bcftools	Varscan 2	GATK Mutect2	
G250E	CML024				
Y253H	CML014				
	CML020				
	CML023				
	CML066				
	CML081				
E255K	CML012				
	CML014				
	CML037				
	CML081				
T315I	CML004				
	CML017				
	CML024				
	CML032				
	CML068				
F317L	CML032				
K357T	CML070				
F359V	CML017				
	CML078				
F359C	CML070				
E453K	CML077				
E459K	CML024				

NGS detection
VAF>15%
VAF>3% and ≤15%
No mutation detected

SS detection
Mutation detected
No mutation detected

FIG. 2: Identification of mutation by next-generation sequencing (NGS) and Sanger Sequencing (SS).

detected most of the variants identified by SS but did not detect the Y253H mutation in patient CML081. This oversight may be attributed to the limitations of VarScan 2 in detecting low-frequency variants. The presence of the Y253H mutation at a low frequency might not have met the specific thresholds for variant allele frequency and sequencing depth set by VarScan 2, resulting in it not being flagged by the detection algorithms.

GATK Mutect2 identified additional low-frequency mutations that were missed by BCFtools, VarScan 2, and SS. However, GATK Mutect2 did not detect the F359V mutation in patients CML017 and CML078. This could be due to the sensitivity settings and filtering criteria of GATK Mutect2, which are optimised to reduce false positives but can occasionally lead to missed mutations, especially in regions of high complexity or in proximity to other variants.

Therefore, relying on a single variant caller could lead to missed variants that other tools might detect. The findings suggest that combining the strengths of BCFtools, VarScan 2, and GATK Mutect2 would be advantageous for maximising the detection of genetic variations. Using multiple variant callers can improve both sensitivity and specificity. It is generally recommended to use multiple

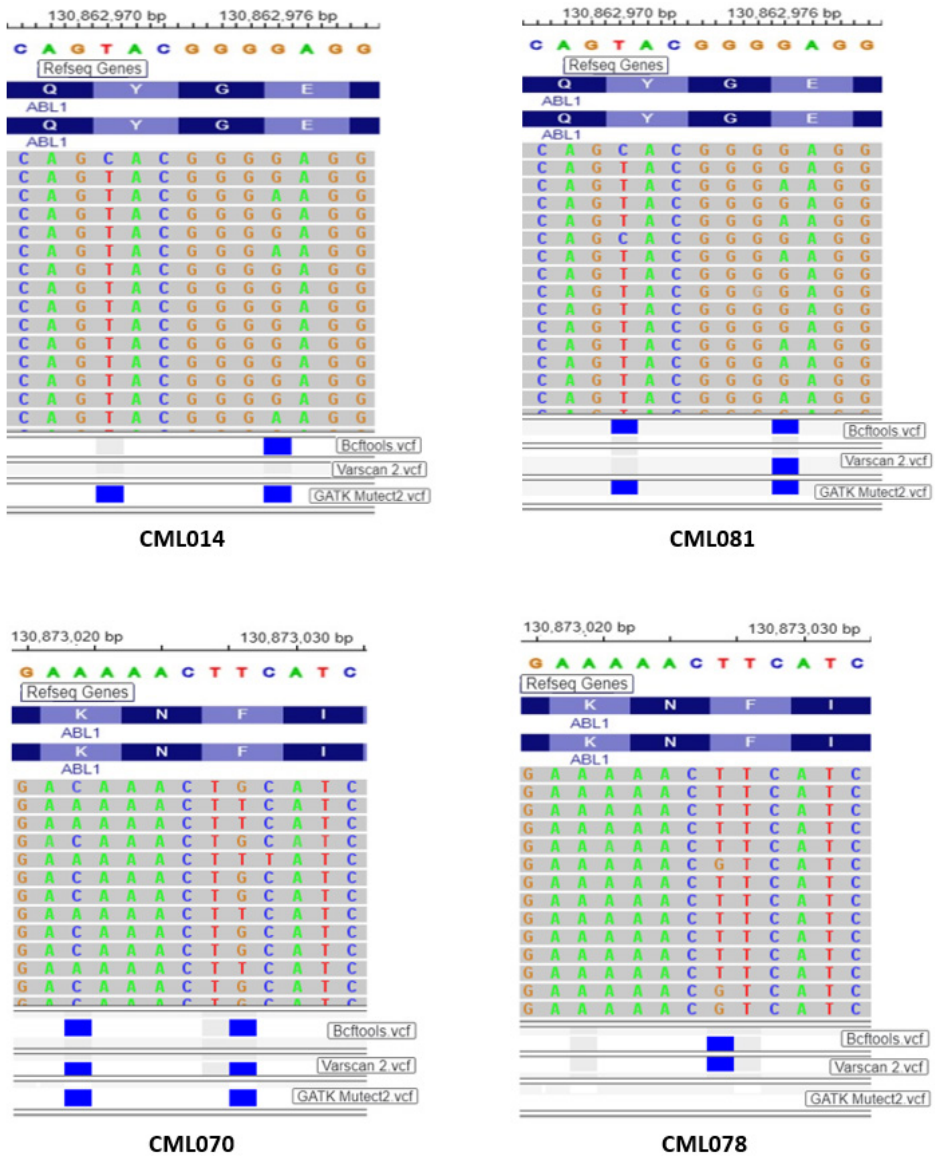

Figure 3. Integrative Genomics Viewer (IGV) images of mutations detected in four patients, CML014, CML081, CML070 and CML078.

tools to cross-validate findings, creating a more robust and comprehensive variant-calling pipeline. Thus, combining all three tools is likely the most effective approach. Our results provide valuable guidance for researchers in choosing appropriate workflows for extracting SNVs from NGS data in *BCR::ABL1* KD mutation screening.

Keywords: BCR::ABL1 kinase domain mutations, Variant calling methods, next-generation sequencing (NGS), Tyrosine kinase inhibitors (TKIs) resistance, Somatic mutations detection

Acknowledgements: The authors would like to thank the Director General of Health Malaysia for approval to publish this scientific paper. We would also like to thank the Deputy Director General of Health (Research and Technical Support) and the Director of the Institute for Medical Research (IMR) for their support.

Authors' contributions: All authors contributed equally.

Conflicts of interest: The authors declare no conflict of interest.

REFERENCES

1. Quintás-Cardama A, Kantarjian H, Cortes J. Imatinib and beyond--exploring the full potential of targeted therapy for CML. Nat Rev Clin Oncol. 2009;6(9):535-43.
2. Nicolini FE, Ibrahim AR, Soverini S, *et al*. The BCR-ABLT315I mutation compromises survival in chronic phase chronic myelogenous leukemia patients resistant to tyrosine kinase inhibitors, in a matched pair analysis. Haematologica. 2013;98(10):1510-6.
3. Soverini S, Hochhaus A, Nicolini FE, *et al*. BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet. Blood. 2011;118(5):1208-15.
4. Braun TP, Eide CA, Druker BJ. Response and Resistance to BCR-ABL1-Targeted Therapies. Cancer Cell. 2020;37(4):530-42.
5. Erbilgin Y, Eskazan AE, Hatirnaz Ng O, *et al*. Deep sequencing of BCR-ABL1 kinase domain mutations in chronic myeloid leukemia patients with resistance to tyrosine kinase inhibitors. Leuk Lymphoma. 2019;60(1):200-7.
6. Kizilors A, Crisà E, Lea N, *et al*. Effect of low-level BCR-ABL1 kinase domain mutations identified by next-generation sequencing in patients with chronic myeloid leukaemia: a population-based study. Lancet Haematol. 2019;6(5):e276-e84.
7. Soverini S, De Benedittis C, Polakova KM, *et al*. Next-generation sequencing for sensitive detection of BCR-ABL1 mutations relevant to tyrosine kinase inhibitor choice in imatinib-resistant patients. Oncotarget. 2016;7(16):21982-90.
8. Benjamin ESB, Ravindra N, Rajamani BM, *et al*. BCR-ABL1 kinase domain mutation analysis by next generation sequencing detected additional mutations in chronic myeloid leukemia patients with suboptimal response to imatinib. Leuk Lymphoma. 2021;62(6):1528-31.
9. Spencer DH, Tyagi M, Vallania F, *et al*. Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data. J Mol Diagn. 2014;16(1):75-88.
10. Krøigård AB, Thomassen M, Lænkholm A-V, Kruse TA, Larsen MJ. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. PLOS ONE. 2016;11(3):e0151664.
11. David B, Takuto S, Kristian C, *et al*. Calling Somatic SNVs and Indels with Mutect2. BioRxiv. 2019:861054.
12. Chen J, Kao YR, Sun D, *et al*. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. Nat Med. 2019;25(1):103-10.
13. Cai L, Yuan W, Zhang Z, He L, Chou K-C. In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. Scientific Reports. 2016;6(1):36540
14. Koboldt DC, Larson DE, Wilson RK. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. Curr Protoc Bioinformatics. 2013 Dec;44:15.4.1-17.
15. Churpek JE, Pyrtel K, Kanchi KL, *et al*. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. Blood. 2015;126(22):2484-90.
16. Roberts ND, Kortschak RD, Parker WT, *et al*. A comparative analysis of algorithms for somatic SNV detection in cancer. Bioinformatics. 2013;29(18):2223-30.
17. Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics. 2017;33(13):2037-9.
18. Ura H, Togi S, Niida Y. Dual Deep Sequencing Improves the Accuracy of Low-Frequency Somatic Mutation Detection in Cancer Gene Panel Testing. Int J Mol Sci. 2020 May 16;21(10):3530.
19. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ, Forbes SA. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-D947.