

REVIEWER'S REPORT

Manuscript No.: IJAR-59358

Title: OBJECT-ORIENTED DESIGN AND UI PROTOTYPING FOR A BIOINFORMATICS TOOL TO DETECT AND ANNOTATE NGS VARIANTS OF PLASMODIUM FALCIPARUM IN INDIVIDUALS WITH SICKLE CELL DISEASE

Recommendation:

Accept as it is

Accept after minor revision.....

Accept after major revision

Do not accept (*Reasons below*)

Rating	Excel.	Good	Fair	Poor
Originality		x		
Techn. Quality		x		
Clarity			X	
Significance			X	

Reviewer's ID: JPR-142

Detailed Reviewer's Report

The manuscript presents the object-oriented design, UML-based conceptual modeling, deployment architecture, and GUI prototypes of IPCINGSTOOLKIT, a proposed platform intended to facilitate NGS-based *Plasmodium falciparum* sequence extraction, variant calling, and annotation in the context of sickle cell disease. The manuscript has an interdisciplinary focus combining bioinformatics, software engineering, and user-interface design. The use-case, activity, and deployment diagrams are useful for communicating the proposed architecture, while the GUI mockups provide a practical visualization of the intended workflow. However, the manuscript currently describes largely a conceptual design and prototype, while some statements imply that a functional and validated bioinformatics platform already exists. The distinction between software design/prototyping, implementation, and empirical validation needs to be made much clearer. In particular, no real NGS dataset, benchmark analysis, runtime assessment, variant-calling accuracy, host-read removal efficiency, annotation accuracy, or usability evaluation is presented. Therefore, the current manuscript does not yet demonstrate that the proposed platform performs reliably in practice.

1. Clarify the actual status of IPCINGSTOOLKIT

The most important issue is the distinction between a **prototype/design** and a **fully implemented bioinformatics platform**.

REVIEWER'S REPORT

The manuscript describes GUI prototypes and states that future work will involve actual implementation using Snakemake or Nextflow and completion of the web frontend. Therefore, the manuscript should consistently describe IPCINGSTOOLKIT as a proposed architecture/prototype unless the authors can demonstrate that the computational pipeline has already been implemented and executed.

Statements such as:

"Such a modular architecture enables secure, high-performance, and scalable remote execution..."

should be reconsidered because performance, scalability, and security are not experimentally evaluated in the current study

Recommendation: Clearly distinguish:

- conceptual design;
- GUI prototype;
- implemented software components;
- computational pipeline;
- experimentally validated components.

2. The manuscript lacks empirical validation

This is the principal technical weakness.

The manuscript describes FastQC, MultiQC, Trim Galore, BWA, SAMtools, GATK, FreeBayes, BCFtools and VEP, but no actual sequencing dataset appears to have been processed and no analytical performance is reported.

The authors should provide, if available:

- number of NGS samples analyzed;
- sequencing platform;
- read length;
- paired-end/single-end status;
- sequencing depth;
- parasite parasitemia or parasite read fraction;
- percentage of human reads removed;
- percentage of parasite reads recovered;
- mapping rate;
- coverage depth;
- breadth of genome coverage;
- number of variants identified;
- number of variants retained after filtering;
- concordance between variant callers;
- annotation results.

If no real dataset has yet been processed, this should be explicitly stated as a **prototype/design study**, and claims concerning performance should be removed or qualified.

REVIEWER'S REPORT

3. The title currently overstates the scope

The title is:

“OBJECT-ORIENTED DESIGN AND UI PROTOTYPING FOR A BIOINFORMATICS TOOL TO DETECT AND ANNOTATE NGS VARIANTS OF *PLASMODIUM FALCIPARUM* IN INDIVIDUALS WITH SICKLE CELL DISEASE.”

The manuscript does not actually present experimental variant detection and annotation in individuals with sickle cell disease. Instead, it presents the design and prototype of a proposed tool.

4. The NGS workflow needs substantially more technical detail

The proposed workflow includes quality control, human-read removal, parasite extraction, variant calling, filtering, and VEP annotation. However, important computational parameters are insufficiently specified.

The authors should provide a reproducible workflow including:

- FastQC version;-
- MultiQC version;
- Trim Galore version;
- BWA version and algorithm;
- SAMtools version;
- reference genome accession/version;
- *P. falciparum* 3D7 reference version;
- annotation database/version;
- GATK version;
- FreeBayes version;
- BCFtools version;
- VEP version;
- Python version;
- Streamlit version;
- Bioconda environment specification.

The current manuscript states that GRCh38/hg38 and *P. falciparum* 3D7 are used as references, but accession/version information should be provided to ensure reproducibility.

5. The “double subtraction” strategy requires stronger justification

The manuscript describes a two-stage process:

1. removal of human-origin reads;
2. extraction of parasite reads by alignment against the *P. falciparum* reference. This is reasonable conceptually, but the authors need to explain why this particular strategy was selected over alternative approaches.

REVIEWER'S REPORT

More importantly, alignment against a single 3D7 reference may potentially reduce sensitivity for genetically divergent parasite isolates.

The authors should discuss:

- reference bias;
- genomic diversity of *P. falciparum*;
- mixed infections;
- repetitive genomic regions;
- var/rifin/stevor or other highly variable regions;
- potential loss of non-3D7 parasite reads;
- whether alternate or population references could be supported.

This is particularly important because the stated scientific objective concerns parasite adaptation and population-level variation.

6. Variant-calling strategy needs methodological justification

The manuscript states that the user selects two variant callers from GATK HaplotypeCaller, BCFtools mpileup/call, and FreeBayes, and that only variants present in the intersection of the two callers are retained. The statement that this approach “maximizes[s] reliability” is too strong without validation.

Intersection-based calling may improve specificity but can also reduce sensitivity by eliminating true variants detected by only one caller.

The authors should therefore:

- justify the two-caller strategy;
- explain whether the intersection is mandatory;
- distinguish sensitivity from specificity;
- define how discordant calls are handled;
- specify whether the approach is appropriate for haploid *P. falciparum* genomes;
- Provide benchmarking against a truth set, if available.

A statement such as “the intersection was designed to prioritize concordant calls and potentially reduce false-positive variants” would be more scientifically appropriate than claiming that it maximizes reliability without supporting data.

7. Variant filtering thresholds require justification

The manuscript uses examples such as:

- DP > 10;
- QUAL > 20.
- These thresholds appear arbitrary unless supported by literature or benchmarking.

The authors should explain:-

- why DP >10 was selected;

REVIEWER'S REPORT

- why QUAL >20 was selected;
- whether thresholds are configurable;
- whether different thresholds are applied to SNPs and INDELs;
- whether allele frequency/allele balance is considered;
- whether mapping quality is considered;
- whether strand bias is evaluated.

If these are merely example values in the prototype, the text should explicitly say so.

8. Annotation methodology needs clarification

The manuscript states that VEP is used to annotate variants and assess transcript consequences, protein impact, and potential resistance mechanisms. However, the exact annotation resource is not sufficiently described.

The authors should specify:

- VEP version;
- *P. falciparum* annotation release;
- GFF/GTF source;
- transcript database;
- reference genome build;
- whether resistance-associated mutations are obtained directly from VEP or from an additional resistance database;
- which resistance database(s) are used;
- How annotation conflicts are handled.

The phrase “potential resistance mechanisms” requires particular caution unless an explicit resistance database or curated evidence source is incorporated.

9. No usability evaluation has been conducted

However, no actual usability assessment is presented.

If the authors want to claim improved usability, they should ideally conduct a preliminary evaluation involving representative users, for example:

- bioinformaticians;
- biologists;
- clinicians/researchers.

Possible measures could include:

- System Usability Scale (SUS);
- task completion rate;
- completion time;
- user errors;

REVIEWER'S REPORT

- satisfaction;
- learnability.

10. Security claims need substantial clarification

This is an important issue because the manuscript describes clinical/biological data and user authentication.

The authors should discuss:

- password hashing;
- authentication;
- authorization/role-based access;
- session management;
- encryption;
- audit logging;
- protection of human genomic data;
- secure data transfer;
- server isolation.

11. The deployment architecture should match the actual implementation

However, Docker/Singularity/Apptainer implementation details are not actually presented.

If containerization has not yet been implemented, the authors should describe it as a planned deployment architecture, rather than an established feature.

Similarly, claims of:

- portability;
- scalability;
- high-performance execution;
- reproducibility

should either be demonstrated or appropriately qualified.

12. The Results section is largely a design description

The Results section primarily reports:

- actors;
- use cases;
- activity diagrams;
- deployment architecture;
- GUI mockups.

I recommend reorganizing the Results into clearly defined subsections such as:

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

1. Requirements analysis
2. System architecture
3. Bioinformatics workflow
4. GUI prototype
5. Computational implementation status
6. Validation/benchmarking
7. Usability evaluation

If no validation has yet been performed, this limitation should be explicitly acknowledged.