

TransQSAR-pf: A Bio-Informed QSAR Framework Using *Plasmodium falciparum* Stress Signatures for Enhanced Antiplasmodial Activity Prediction [†]

Favour O. Igwezeke ^{1,*}  and Charles O. Nnadi ² ¹ Faculty of Pharmaceutical Sciences, University of Nigeria, Nsukka 410001, Enugu, Nigeria² Department of Pharmaceutical and Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Nigeria, Nsukka 410001, Enugu, Nigeria; charles.nnadi@unn.edu.ng

* Correspondence: favour.igwezeke.249461@unn.edu.ng

[†] Presented at the 6th International Electronic Conference on Applied Sciences, 9–11 December 2025; Available online: <https://sciforum.net/event/ASEC2025>.

Abstract

Traditional QSAR modeling relies solely on molecular descriptors, neglecting the biological state of target organisms. While prior approaches have integrated biological data with molecular features for activity prediction, we developed TransQSAR-pf, a methodological framework that integrates *Plasmodium falciparum* transcriptomic stress signatures with molecular descriptors to construct biologically informed activity prediction models. Applied to 125 triazolopyrimidine derivatives, the framework distilled 764 transcriptomic features into 13 key predictors through Boruta selection, constructing an interpretable model ($R^2 = 0.762$, RMSE = 0.470) that demonstrated improved performance over the baseline QSAR-only model ($R^2 = 0.719$, RMSE = 0.529). Biological mapping revealed that 71.2% of feature importance derived from conserved unknown-function genes, representing largely uncharacterized stress response pathways that correlate with compound efficacy and warrant experimental characterization, demonstrating the framework's utility for generating mechanistic hypotheses. This work presents a novel computational pipeline for building biology-aware QSAR models that prioritize experimental targets for antimalarial discovery.

Keywords: QSAR; transcriptomics; *Plasmodium falciparum*; machine learning; antiplasmodial activity; feature selection; drug discovery; malaria

1. Introduction

Malaria remains a major global health threat, with the World Health Organization reporting an estimated 282 million cases and 610,000 deaths worldwide in 2024, exceeding pre-pandemic levels [1]. The burden is disproportionately high in the WHO African Region, which accounted for approximately 94% of all cases and deaths. Despite ongoing control efforts, emerging challenges, such as drug and insecticide resistance, and the growing threats posed by climate change and humanitarian crises underscore the urgent need for novel therapeutic strategies and improved discovery pipelines against *Plasmodium falciparum*.

Quantitative structure–activity relationship (QSAR) models are widely used for virtual screening and prioritization of candidate antiplasmodial compounds because they relate molecular descriptors to measured activity [2]. For example, a recent application of multiple machine learning algorithms to build a QSAR model for triazolopyrimidine analogs against



Academic Editor: Roger Narayan

Published: 13 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

P. falciparum [3]. However, traditional QSAR approaches ignore the parasite's biological state [2,3]. The integration of biological data with chemical structure information to enhance activity prediction has been explored in multiple contexts. Compound-induced transcriptomic profiling has shown promise for target prediction in drug discovery [4,5], with studies demonstrating that gene expression signatures can overcome chemical space limitations inherent in traditional QSAR [6]. Furthermore, transcriptomics-guided approaches have successfully prioritized compounds in pharmaceutical projects [7], and multi-modal integration of cellular data with chemical features has improved mechanism of action clustering [8]. While these pioneering studies have established the value of incorporating biological information into predictive models, they have predominantly focused on human cell lines or cancer models. For parasitic organisms like *P. falciparum*, high-throughput transcriptomic studies have mapped dynamic gene expression during the intraerythrocytic cycle, revealing stage-specific programs and conserved stress responses [9,10] not captured by molecular descriptors alone, suggesting that integrating transcriptomic signatures could enhance compound efficacy prediction specifically for antimalarial applications.

To address this limitation, we present TransQSAR-pf, a methodological framework for constructing biologically informed QSAR models by integrating *P. falciparum* transcriptomic stress signatures with classical molecular descriptors. The framework uses differential expression analysis (limma) [11], pathway enrichment (GSEA) [12], and Boruta feature selection [13] to extract biologically meaningful features from public transcriptomic data. We demonstrate the framework's application to a triazolopyrimidine library, showing how it constructs interpretable models and identifies prioritized biological targets for mechanistic validation.

2. Materials and Methods

2.1. Transcriptomic Data Acquisition and Processing

Public microarray data (GSE10022) comprising 24 samples from chloroquine-resistant *P. falciparum* strains were obtained from the Gene Expression Omnibus [14]. The dataset included three genotypes (106/1, 106/1 + 76I, 106/1 + 76I + 352K) under control and chloroquine treatment conditions, with 24,563 probes representing the parasite transcriptome. Raw CEL files were processed in R (v4.5.1) using the oligo and affy packages. The data underwent Robust Multiarray Average (RMA) background correction, quantile normalization, and log₂ transformation [15]. RMA (Robust Multiarray Average) is a preprocessing method combining background adjustment, quantile normalization, and summarization; MA (Minus-Average) plots are quality control visualizations displaying log-ratio versus average intensity to identify systematic biases. To obtain gene-level expression values, probes were mapped to genes using PlasmoDB annotations (release 68), retaining only the probe with the highest mean expression for genes with multiple probes. Probes with low detection rates (<20% of samples) were filtered out. Quality control was performed using MA plots, hierarchical clustering, and principal component analysis. No significant batch effects were observed; samples clustered by biological condition (genotype and treatment) rather than by technical artifacts. Detailed data processing and analysis code are provided in the Supplementary Materials.

2.2. Differential Expression Analysis

Differential gene expression was performed using the limma package (version 3.56.2) in R [11]. A linear model was fitted with the design matrix:

$$\text{design} = \text{model.matrix}(\sim 0 + \text{group_factor})$$

where *group_factor* represented combinations of genotype and treatment conditions. Contrasts were defined to separately analyze chloroquine responses in each parasite genotype to capture genotype-specific effects: CQ_WT (chloroquine response in wild-type (106/1) parasites), CQ_76I (chloroquine response in 76I mutant parasites) and Genotype_76I (baseline transcriptional differences of 76I mutation versus wild-type). Moderated t-statistics were computed using empirical Bayes methods, with Benjamini–Hochberg false discovery rate (FDR) correction applied (significance threshold: adjusted $p < 0.05$).

2.3. Gene Set Enrichment Analysis (GSEA) and Gene Set Curation

Pathway enrichment was performed using pre-ranked GSEA via the *fgsea* R package version 1.27.0 [16]. Genes were ranked by $\log_2$ fold-change from our differential expression analysis for each contrast. We performed 10,000 permutations to compute normalized enrichment scores (NES) and p -values, setting significance at an adjusted p -value < 0.05 . Gene set annotations were sourced from PlasmoDB (release 68) [17], which provided ~18,000 Gene Ontology (GO) and Reactome terms. Where specific biological categories of interest (e.g., conserved proteins of unknown function, PfEMP1 virulence factors) were not adequately represented by a single standard GO term, we created custom curated gene sets directly from PlasmoDB annotations.

2.4. Transcriptomic Feature Engineering

We engineered 764 transcriptomic features in four categories: differential expression signatures (600), pathway enrichment scores (3), expression variability (100), and functional group profiles (61). These features represent baseline parasite stress states and genotype profiles.

2.5. QSAR Dataset and Molecular Descriptors

The QSAR dataset consisted of 125 triazolopyrimidine derivatives with experimentally determined pIC_{50} values against the chloroquine-sensitive *P. falciparum* 3D7 strain. The chemical structures and activity data were sourced from a previous work [3]; this published study serves as our baseline comparator for model performance. Molecular descriptors were calculated using the RDKit cheminformatics toolkit [18]. We focused on fifteen descriptors previously identified as most significant for modeling antiparasitic activity in this specific chemical scaffold: the five features from the published regression equation (npr1 , pmi3 , slogP , vsurf_W2 , vsurf_CW2) [3], supplemented with ten additional standard descriptors relevant to drug-like properties (HBD, HBA, TPSA, nRotB , MW, nAromRing , logP_ow , mr , apol , and density).

2.6. Feature Integration Strategy

We integrated transcriptomic features (baseline parasite stress state) with compound molecular descriptors. Since transcriptomic signatures derived from chloroquine experiments, we implemented a controlled simulation to test whether they provide biologically informed priors. Specifically, each baseline transcriptomic feature was modulated by two key compound physicochemical properties, which are lipophilicity (slogP) and molecular weight (MW), which were first standardized (mean-centered and scaled to unit variance). The modulation formula for a given feature j was:

$$\text{transcriptomic}_{\text{adjusted}j} = \text{transcriptomic}_{\text{baseline}j} + (0.1 \times \text{slogP}_{\text{scaled}}) + (0.05 \times \text{MW}_{\text{scaled}}) + \varepsilon_j$$

where $\varepsilon_j \sim N(0, 0.03)$ represents a small Gaussian noise term. A fixed random seed (42) ensured full reproducibility.

Small coefficients (0.1, 0.05) modeled subtle physicochemical influence while preserving baseline biological signal. This proof-of-concept simulation tests whether biologically derived features yield interpretable models. The final, combined matrix for each compound contained 779 features (764 modulated transcriptomic + 15 QSAR descriptors) for downstream analysis.

2.7. Boruta Feature Selection

Boruta feature selection [13] was applied to the 779-feature matrix using Random Forest with 200 iterations. The algorithm compares real feature importance scores against “shadow features” (permuted copies) to identify truly informative predictors. Features were classified as “confirmed,” “tentative,” or “rejected.” Only confirmed features were retained for final modeling. Boruta identified 13 confirmed transcriptomic features from the original 764 (a 98.3% reduction), which were combined with the 15 QSAR descriptors to create the final feature set (28 features total) used in the machine learning models.

2.8. Machine Learning Modeling

The 125 compounds were randomly split into training (80%, $n = 100$) and held-out test (20%, $n = 25$) sets using a fixed random seed to ensure reproducibility. We verified that the distribution of pIC_{50} values was comparable across the two sets. We evaluated three algorithms: Random Forest (RF) [19], Support Vector Machine with radial basis function kernel (SVM-RBF) [20], and Elastic Net regularized regression [21]. Hyperparameters were optimized via 5-fold cross-validation on the training set only. The best model from each algorithm was selected based on the lowest mean cross-validated Root Mean Square Error (RMSE). To contextualize our TransQSAR-pf framework, we established two key baselines: (1) a QSAR-only RF model using only the 15 molecular descriptors, and (2) a naive integration RF model using all 779 features without Boruta selection, to illustrate the overfitting risk. Final model performance was evaluated on the held-out test set using the coefficient of determination (R^2) and RMSE.

2.9. Feature Importance Aggregation and Reporting

For the final Random Forest model, feature importance was computed as mean decrease in accuracy and normalized to sum = 1. We grouped importance by feature category (Conserved unknown-function genes; Genotype_DE; CQ_DE; QSAR descriptors; Variability features; Functional groups). The reported percentages represent the sum of normalized importances for features assigned to each category.

3. Results

3.1. Transcriptomic Landscape of Chloroquine Stress

Transcriptomic analysis revealed distinct patterns: chloroquine treatment induced expression changes in 1246 genes in wild-type parasites at nominal significance ($p < 0.05$; 711 upregulated, 535 downregulated), though these did not survive strict multiple testing correction. The strongest responding genes showed substantial fold-changes (median $|\log_2\text{FC}| = 0.388$ for top 200 genes). In contrast, substantial baseline differences existed between genotypes, with 337 and 1494 genes differentially expressed between mutant and wild-type strains at $\text{FDR} < 0.05$.

GSEA revealed three significantly enriched pathways ($p < 0.05$; Figure 1): conserved Plasmodium proteins of unknown function ($p = 0.0038$, $\text{NES} = +1.58$)—upregulated under chloroquine stress, RNA-binding proteins ($p = 0.034$, $\text{NES} = +1.58$)—upregulated under chloroquine stress, and PfEMP1 virulence factors ($p = 0.036$, $\text{NES} = -1.52$)—downregulated under chloroquine stress.

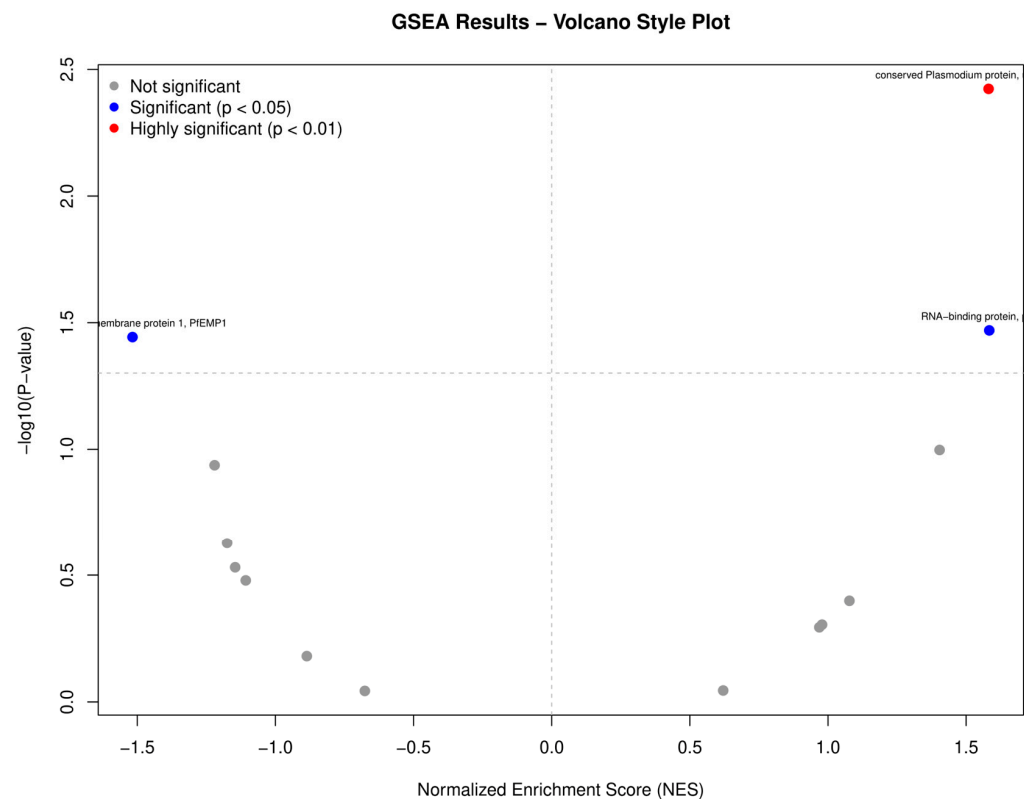


Figure 1. GSEA volcano plot of pathway enrichment under chloroquine stress. Pathways are plotted by Normalized Enrichment Score (NES) versus statistical significance ($-\log_{10} p\text{-value}$). Positive NES indicates up-regulation under drug stress. The three significantly enriched pathways ($p < 0.05$) are labeled.

3.2. Boruta Feature Selection Identifies 13 Critical Predictors

Boruta selection identified 13 confirmed transcriptomic predictors from the original 764 features, yielding a 98.3% reduction (Table 1). These represent three biological categories: conserved unknown-function proteins (69.2%), genotype differences (23.1%), and direct drug response (7.7%). The most important feature (Importance = 9.43) was CQ_WT_DE_40, corresponding to chloroquine response in the wild-type strain.

Table 1. Top 13 Boruta-selected transcriptomic features for antiplasmodial activity prediction. Features represent Boruta-confirmed predictors from the TransQSAR-pf framework applied to 125 triazolopyrimidines.

Rank	Feature	Importance	Category	Context
1	CQ_WT_DE_40	9.43	CUF	CQ response in wild-type
2	CQ_WT_DE_128	5.31	DR	CQ response (logFC = −0.304)
3	Variability_Pf.12.198.0	4.61	CUF	Strain variability
4	CQ_WT_DE_169	4.54	CUF	CQ response
5	Genotype_76I_DE_88	4.17	U	Genotype difference (logFC = 2.179)
6	Variability_Pf.13.1.443.0	4.11	CUF	Strain variability
7	CQ_WT_DE_79	3.29	CUF	CQ response
8	Variability_Pf.2.13.0	3.23	CUF	Strain variability
9	Genotype_76I_DE_92	2.33	U	Genotype difference (logFC = 0.748)
10	Genotype_76I_DE_184	1.97	CUF	Genotype difference
11	Genotype_76I_DE_98	1.97	U	Genotype difference (logFC = 1.096)
12	Variability_Pf.5.281.0	1.83	CUF	Strain variability
13	CQ_WT_DE_135	1.00	CUF	CQ response

Conserved Unknown Function (CUF), chloroquine (CQ), drug response (DR), Others/unclassified (U).

3.3. TransQSAR-pf Framework Generates an Interpretable Bio-Informed Model

Application of the TransQSAR-pf framework to the triazolopyrimidine dataset produced models with varying characteristics depending on the feature selection stage (Table 2). In addition to the summary metrics, predicted versus actual pIC_{50} values for all evaluated models are visualized in Figure 2, allowing for a direct comparison of model generalization behavior.

Table 2. Model performance at key framework stages.

Model Stage	R ² Train	R ² Test	RMSE	Features	Notes
QSAR-only RF	0.812	0.719	0.529	15	Baseline (modest fit, good generalization)
RF Tuned	0.974	0.602	0.653	779	Naive Integration (poor generalization)
RF Boruta-Selected	0.962	0.762	0.470	28	TransQSAR-pf (best model)

Random Forest (RF), Root Mean Square Error (RMSE), coefficient of determination (R²).

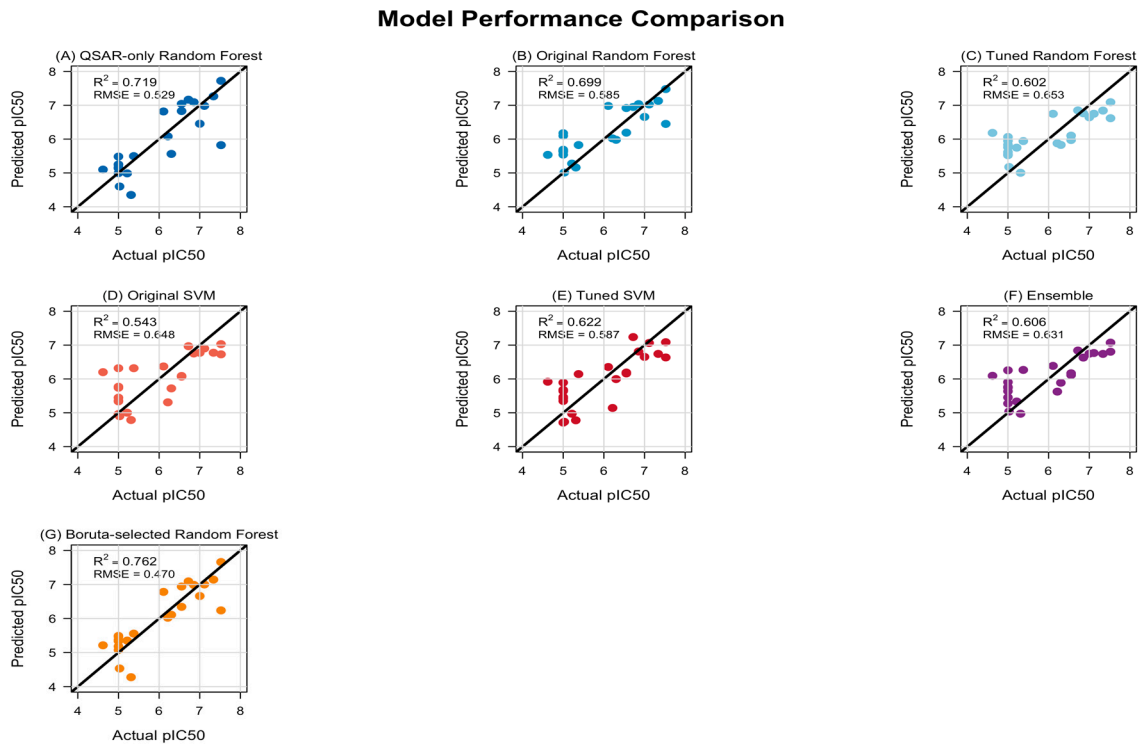


Figure 2. Model performance comparison showing predicted versus actual pIC_{50} values for: (A) QSAR-only Random Forest, (B) Original Random Forest, (C) Tuned Random Forest, (D) Original SVM, (E) Tuned SVM, (F) Ensemble, and (G) Boruta-selected testing Random Forest. The solid line represents perfect prediction ($y = x$). Each model’s R^2 value from testing is indicated in the respective panel.

Naive integration of all 779 features resulted in poor generalization (test $R^2 = 0.602$), underscoring the critical role of Boruta feature selection. In contrast, the Boruta-selected Random Forest achieved the best balance between fit and generalization (test $R^2 = 0.762$, $RMSE = 0.470$), with predictions closely aligned to the ideal $y = x$ relationship shown in Figure 2G.

The framework’s value lies not primarily in predictive metrics, but in its ability to construct interpretable models where biological features can be mapped to mechanistic hypotheses. The 28-feature final model represents a biologically informed architecture suitable for understanding compound–parasite interactions.

3.4. Biological Feature Importance Distribution

The TransQSAR-pf model allocated feature importance unevenly across biological categories, as summarized in Figure 3. Conserved proteins of unknown function accounted for 71.2% of the total importance, far exceeding contributions from genotype differences (17.7%) and direct drug-response signatures (11.1%). These conserved unknown-function genes, while not yet fully characterized, represent approximately 30% of the *P. falciparum* proteome (~1600 of 5389 proteins) and include many essential genes identified through saturation mutagenesis [22–24]. Recent functional annotation studies have revealed that subsets of these hypothetical proteins contain conserved domains involved in stress response, DNA repair, and metabolic regulation [25,26]. This disproportionate weighting toward unknown-function genes, particularly those showing high expression variability across strains, suggests that inherent parasite stress states offer more robust predictors of compound activity than acute transcriptional responses. The finding corroborates our pathway analysis and nominates conserved unknown-function pathways as priority targets for mechanistic investigation.

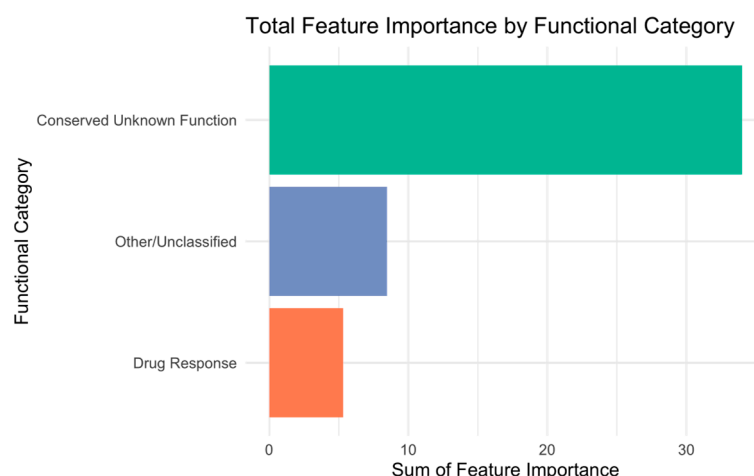


Figure 3. Distribution of feature importance by biological category. Horizontal bars represent normalized importance scores (0–30 scale) from the Boruta-selected Random Forest model, categorized as: Conserved Unknown Function (71.2%), Other/Unclassified (17.7%), and Drug Response (11.1%). Importance scores represent mean decrease in accuracy.

4. Discussion

TransQSAR-pf integrates *P. falciparum* transcriptomic signatures with molecular descriptors to construct biologically informed QSAR models. Applied to 125 triazolo-pyrimidines [3], Boruta selection distilled 764 features to 13 critical predictors ($R^2 = 0.762$, RMSE = 0.470), representing a ~6% improvement in R^2 over the baseline QSAR-only model ($R^2 = 0.719$, RMSE = 0.529) while substantially reducing prediction error (RMSE improvement of 11.2%). More importantly, with 71.2% importance from conserved unknown-function genes, demonstrating utility for hypothesis generation and identifying previously uncharacterized biological pathways that may mediate compound activity.

The framework demonstrates three key insights. First, while the quantitative performance improvement is modest (~6% in R^2), the primary value lies in model interpretability and biological insight generation. Boruta successfully identified a sparse set of 13 biologically meaningful features while discarding noise, producing a model that generalizes better than either the baseline QSAR-only approach or the overfitted naive integration ($R^2 = 0.602$). Second, the predominance of conserved unknown-function genes (71.2% importance) aligns with recent findings that many essential Plasmodium genes remain uncharacterized [22–24]. Rather than suggesting that these genes directly mediate

drug action, their high importance indicates that they represent stress response pathways that correlate with compound efficacy. This identifies them as high-priority targets for experimental characterization, transforming the framework from a purely predictive tool into a hypothesis generator. Third, the PfEMP1 pathway enrichment identified by GSEA ($p = 0.036$) suggests a testable biological hypothesis. PfEMP1 surface antigens mediate cytoadherence and antigenic variation [27], and their downregulation under chloroquine stress may indicate that surface antigen dynamics influence compound activity through effects on membrane properties or cellular stress tolerance. This represents a specific, experimentally addressable hypothesis generated by the framework.

We acknowledge key limitations. The simulation approach, modulating baseline parasite signatures by compound physicochemical properties, serves as a proof-of-concept, as compound-specific expression profiles were unavailable. This methodological choice was necessitated by the absence of experimentally measured compound-induced transcriptomic data for our specific triazolopyrimidine library. Future applications should incorporate experimentally measured compound-induced expression [28] using methods like DRUG-seq [29] or, when available, leverage expression databases such as L1000 or comparable parasite-specific resources. The modest performance improvement ($\sim 6\%$ R^2) reflects the exploratory nature of this proof-of-concept framework; the greater value lies in identifying which biological features correlate with activity, thereby guiding experimental validation efforts.

The TransQSAR-pf framework enables practical antimalarial applications: interpretable models where predictions trace to specific biological features, identification of high-priority genes for validation, and integration into virtual screening workflows. Critically, it addresses a fundamental QSAR limitation by incorporating the target organism's biological state into the prediction of activity.

5. Conclusions

We developed TransQSAR-pf, a computational framework for constructing biologically informed QSAR models by integrating pathogen transcriptomic stress signatures with molecular descriptors. Applied to a triazolopyrimidine library as a proof-of-concept, the framework distilled 764 transcriptomic features into 13 critical predictors through Boruta selection, generating an interpretable model ($R^2 = 0.762$, $RMSE = 0.470$) that demonstrated improved performance and interpretability over the baseline QSAR-only approach. Biological mapping revealed that conserved unknown-function genes account for 71.2% of feature importance, identifying unexplored essential pathways as high-priority targets for mechanistic validation. This work presents a methodological pipeline for building biology-aware QSAR models that prioritize experimental targets and generate testable hypotheses for antimalarial drug discovery. All code, processed data, and analysis scripts are available in the Supplementary Materials.

Supplementary Materials: All code, processed data, and supplementary figures/tables are available in the TransQSAR-pf GitHub repository: <https://github.com/yanny-alt/TransQSAR-pf> (accessed on 19 October 2025).

Author Contributions: Conceptualization, F.O.I. and C.O.N.; methodology, F.O.I.; software, F.O.I.; validation, F.O.I.; formal analysis, F.O.I.; investigation, F.O.I.; resources, C.O.N.; data curation, F.O.I.; writing—original draft preparation, F.O.I.; writing—review and editing, F.O.I. and C.O.N.; visualization, F.O.I.; supervision, C.O.N.; project administration, C.O.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this study are publicly available: Microarray data (GSE10022) from Gene Expression Omnibus <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10022> (accessed on 19 October 2025); PlasmoDB annotations (version 68) from VEuPathDB <https://veupathdb.org> (accessed on 19 October 2025); QSAR compound data from Apeh et al. [3]. Complete analysis code and processed datasets are available at <https://github.com/yanny-alt/TransQSAR-pf> (accessed on 19 October 2025).

Acknowledgments: The authors acknowledge the Gene Expression Omnibus and VEuPathDB for providing public access to transcriptomic data and genome annotations.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript: QSAR (Quantitative Structure-Activity Relationship); GSEA (Gene Set Enrichment Analysis); GO (Gene Ontology); FDR (False Discovery Rate); NES (Normalized Enrichment Score); RMSE (Root Mean Square Error); SVM (Support Vector Machine); RF (Random Forest).

References

1. World Malaria Report 2023. Available online: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023> (accessed on 19 October 2025).
2. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inform.* **2010**, *29*, 476–488. [CrossRef] [PubMed]
3. Apeh, I.S.; Ayoka, T.O.; Nnadi, C.O.; Obonga, W.O. Modeling the Quantitative Structure–Activity Relationships of 1,2,4-Triazolo[1,5-a]Pyrimidin-7-Amine Analogs in the Inhibition of Plasmodium Falciparum. *Eng. Proc.* **2025**, *87*, 52. [CrossRef]
4. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437–1452.e17. [CrossRef]
5. Pabon, N.A.; Xia, Y.; Estabrooks, S.K.; Ye, Z.; Herbrand, A.K.; Süß, E.; Biondi, R.M.; Assimon, V.A.; Gestwicki, J.E.; Brodsky, J.L.; et al. Predicting Protein Targets for Drug-like Compounds Using Transcriptomics. *PLoS Comput. Biol.* **2018**, *14*, e1006651. [CrossRef]
6. Baillif, B.; Wichard, J.; Méndez-Lucio, O.; Rouquié, D. Exploring the Use of Compound-Induced Transcriptomic Data Generated from Cell Lines to Predict Compound Activity Toward Molecular Targets. *Front. Chem.* **2020**, *8*, 296. [CrossRef]
7. Verbist, B.; Klambauer, G.; Vervoort, L.; Talloen, W.; Shkedy, Z.; Thas, O.; Bender, A.; Göhlmann, H.W.H.; Hochreiter, S. Using Transcriptomics to Guide Lead Optimization in Drug Discovery Projects: Lessons Learned from the QSTAR Project. *Drug Discov. Today* **2015**, *20*, 505–513. [CrossRef]
8. Ha, S.V.; Jaensch, S.; Kańduła, M.M.; Herman, D.; Czodrowski, P.; Ceulemans, H. Cross Modality Learning of Cell Painting and Transcriptomics Data Improves Mechanism of Action Clustering and Bioactivity Modelling. *Sci. Rep.* **2025**, *15*, 23010. [CrossRef]
9. The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium Falciparum. *PLoS Biol.* **2003**, *1*, e5. Available online: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0000005> (accessed on 19 October 2025).
10. Le Roch, K.G.; Zhou, Y.; Blair, P.L.; Grainger, M.; Moch, J.K.; Haynes, J.D.; De La Vega, P.; Holder, A.A.; Batalov, S.; Carucci, D.J.; et al. Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. *Science* **2003**, *301*, 1503–1508. [CrossRef]
11. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef] [PubMed]
12. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. Available online: <https://www.pnas.org/doi/10.1073/pnas.0506580102> (accessed on 19 October 2025). [CrossRef] [PubMed]
13. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Soft.* **2010**, *36*, 1–13. [CrossRef]
14. Jiang, H.; Patel, J.J.; Yi, M.; Mu, J.; Ding, J.; Stephens, R.; Cooper, R.A.; Ferdig, M.T.; Su, X.Z. Genome-Wide Compensatory Changes Accompany Drug-Selected Mutations in the Plasmodium Falciparum Crt Gene. *PLoS ONE* **2008**, *3*, e2484. [CrossRef]
15. Irizarry, R.A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y.D.; Antonellis, K.J.; Scherf, U.; Speed, T.P. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics* **2003**, *4*, 249–264. [CrossRef]

16. Korotkevich, G.; Sukhov, V.; Budin, N.; Shpak, B.; Artyomov, M.N.; Sergushichev, A. Fast Gene Set Enrichment Analysis. *bioRxiv* **2019**. [CrossRef]
17. Amos, B.; Aurrecochea, C.; Barba, M.; Barreto, A.; Basenko, E.Y.; Bazant, W.; Belnap, R.; Blevins, A.S.; Böhme, U.; Brestelli, J.; et al. VEuPathDB: The Eukaryotic Pathogen, Vector and Host Bioinformatics Resource Center. *Nucleic Acids Res.* **2022**, *50*, D898–D911. [CrossRef]
18. Landrum, G. RDKit: Open-Source Cheminformatics Software, Version 2024.03.5; RDKit Contributors. Available online: <https://www.rdkit.org> (accessed on 19 October 2025).
19. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
20. Scikit-Learn: Machine Learning in Python. Available online: <https://www.jmlr.org/papers/v12/pedregosa11a.html> (accessed on 19 October 2025).
21. Zou, H.; Hastie, T. Regularization and Variable Selection Via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
22. Zhang, M.; Wang, C.; Otto, T.D.; Oberstaller, J.; Liao, X.; Adapa, S.R.; Udenze, K.; Bronner, I.F.; Cassandra, D.; Mayho, M.; et al. Uncovering the Essential Genome of the Human Malaria Parasite *Plasmodium Falciparum* by Saturation Mutagenesis. *Science* **2018**, *360*, eaap7847. [CrossRef]
23. Ali, F.; Wali, H.; Jan, S.; Zia, A.; Aslam, M.; Ahmad, I.; Afridi, S.G.; Shams, S.; Khan, A. Analysing the Essential Proteins Set of *Plasmodium Falciparum* PF3D7 for Novel Drug Targets Identification against Malaria. *Malar. J.* **2021**, *20*, 335. [CrossRef]
24. Singh, G.; Gupta, D. In-Silico Functional Annotation of *Plasmodium Falciparum* Hypothetical Proteins to Identify Novel Drug Targets. *Front. Genet.* **2022**, *13*, 821516. [CrossRef] [PubMed]
25. Hillier, C.; Pardo, M.; Yu, L.; Bushell, E.; Sanderson, T.; Metcalf, T.; Herd, C.; Anar, B.; Rayner, J.C.; Billker, O.; et al. Landscape of the *Plasmodium* Interactome Reveals Both Conserved and Species-Specific Functionality. *Cell Rep.* **2019**, *28*, 1635–1647.e5. [CrossRef] [PubMed]
26. Panda, M.; Srivastava, V.; Singh, S.; Prusty, D. Unveiling Prospective Therapeutic Potential of Conserved Hypothetical *Plasmodium Falciparum* Proteins by Using Integrated Proteo Genomic Annotation and In-Silico Therapeutic Discovery Approach. *Protein J.* **2025**, *44*, 437–463. [CrossRef]
27. Hadjimichael, E.; Deitsch, K.W. Variable Surface Antigen Expression, Virulence, and Persistent Infection by *Plasmodium Falciparum* Malaria Parasites. *Microbiol. Mol. Biol. Rev.* **2025**, *89*, e00114-23. [CrossRef]
28. Silva, M.; Malmberg, M.; Otienoburu, S.D.; Björkman, A.; Ngasala, B.; Mårtensson, A.; Gil, J.P.; Veiga, M.I. *Plasmodium Falciparum* Drug Resistance Genes *Pfmdr1* and *Pfcr1* In Vivo Co-Expression During Artemether-Lumefantrine Therapy. *Front. Pharmacol.* **2022**, *13*, 868723. [CrossRef] [PubMed]
29. Ye, C.; Ho, D.J.; Neri, M.; Yang, C.; Kulkarni, T.; Randhawa, R.; Henault, M.; Mostacci, N.; Farmer, P.; Renner, S.; et al. DRUG-Seq for Miniaturized High-Throughput Transcriptome Profiling in Drug Discovery. *Nat. Commun.* **2018**, *9*, 4307. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.