P falciparum Artemisinin Resistance and Clearance

Natalie Thornton

P falciparum and Malaria

- *P falciparum* is one of four malaria-causing parasites. It is responsible for nearly all malaria fatalities

- WHO estimates 210 million cases of falciparum malaria in 2016, with 435,000 deaths

- Artemisinin drugs are part of standard combination therapy used to treat malaria globally.

- Artemisinin resistant *P falciparum* populations suspected in parts of East Africa and Southeast Asia

World Health Organization (2017). World Malaria Report 2018.
Retrieved from https://www.who.int/malaria/publications/world-malaria-report-2017/report/en/

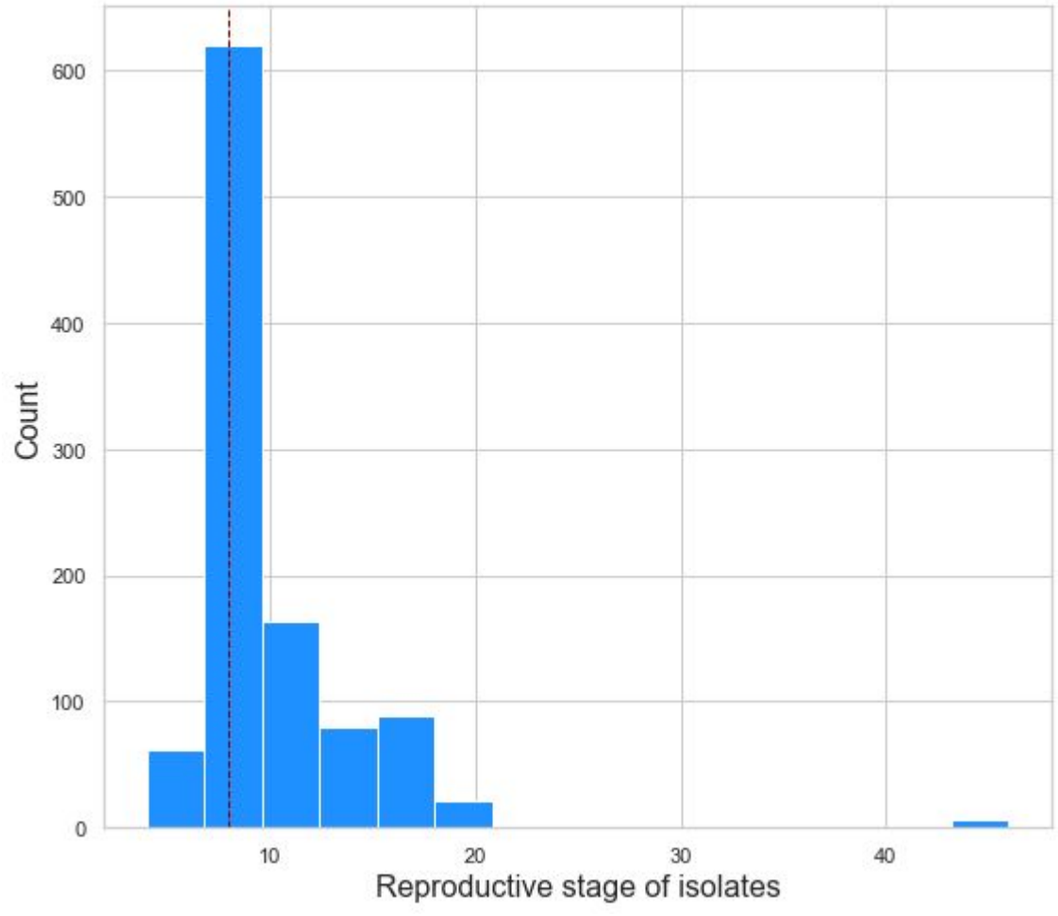
Datasets:

Dataset 1:

- 30 parasite isolates with DHA IC50 dose
- Microarray data covering 5540 genes
- Conditions: treated and untreated, at 6 hours and 24 hours after estimated time of infection. 2 to 6 biological replicates per isolate and condition

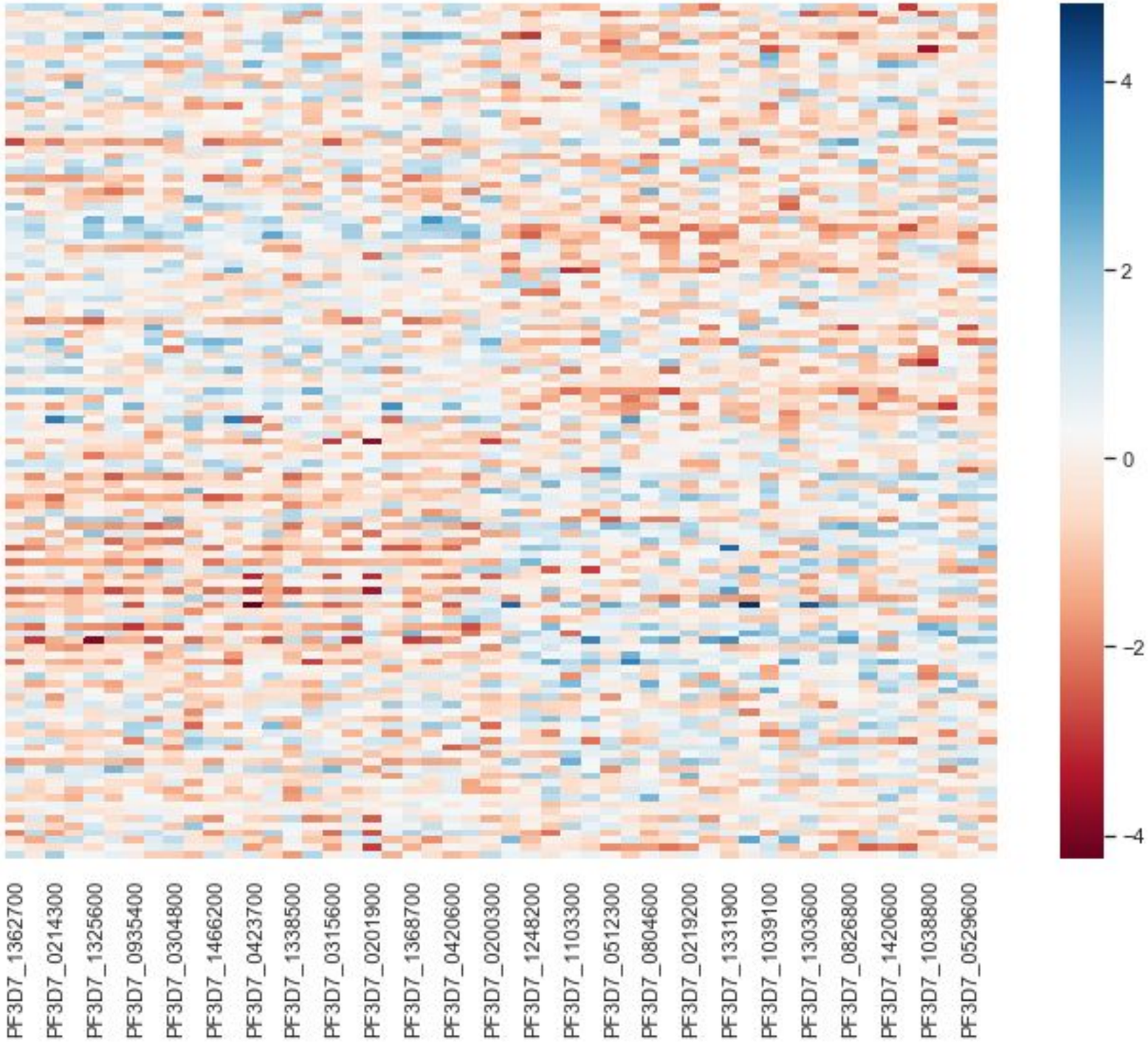
Dataset 2:

- 1043 parasite isolates, with country of infection, reproductive stage of isolate, categorical clearance rate
- Microarray data covering 4950 genes per isolate
- No biological replicates



In Vitro Microarray Data:

Top 50 differentially expressed genes under DHA perturbation or control



Artemisinin IC50 Regression Model:

- Tested Ridge, Lasso, Bayesian Ridge, SVM, random forest, KNN, with poor results

- Tested models again using feature selection:

    - Random forest feature importance used to select top 50 features. Tested random forest, SVM, and logistic regressions. Still generated no positive R2 scores

    - Recursive feature selection used to select 50 and 100 features. Best model performance yielded by SVM linear kernel applied to top 100 RFE selected features

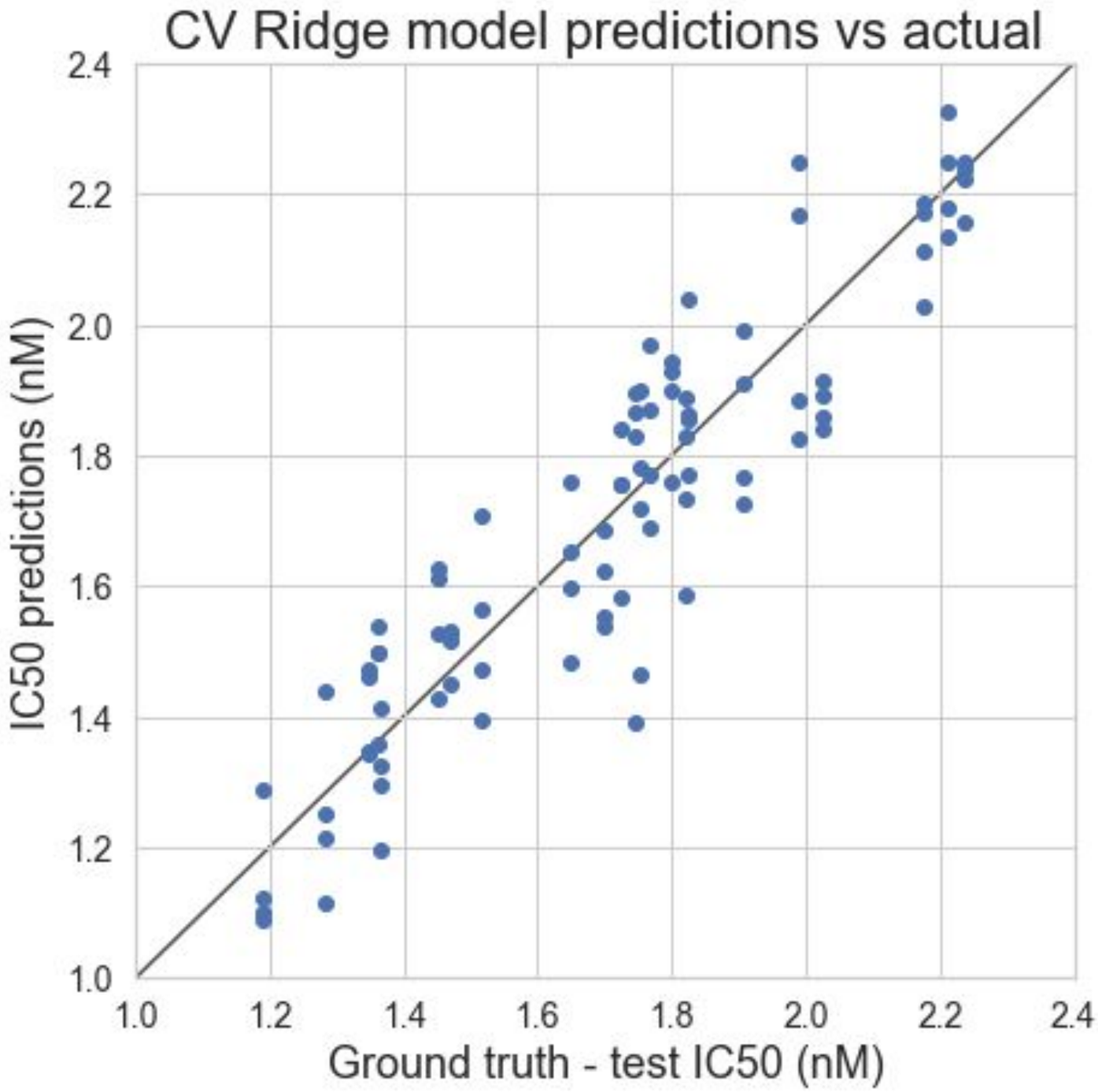
- Hyperparameter tuning performed on most promising models

In Vitro IC50 prediction model

- Most successful model: cross validated ridge regression model trained on 100 features selected with RFE

5-fold CV scores:

- R2 score: 0.90
- MAE: 0.10

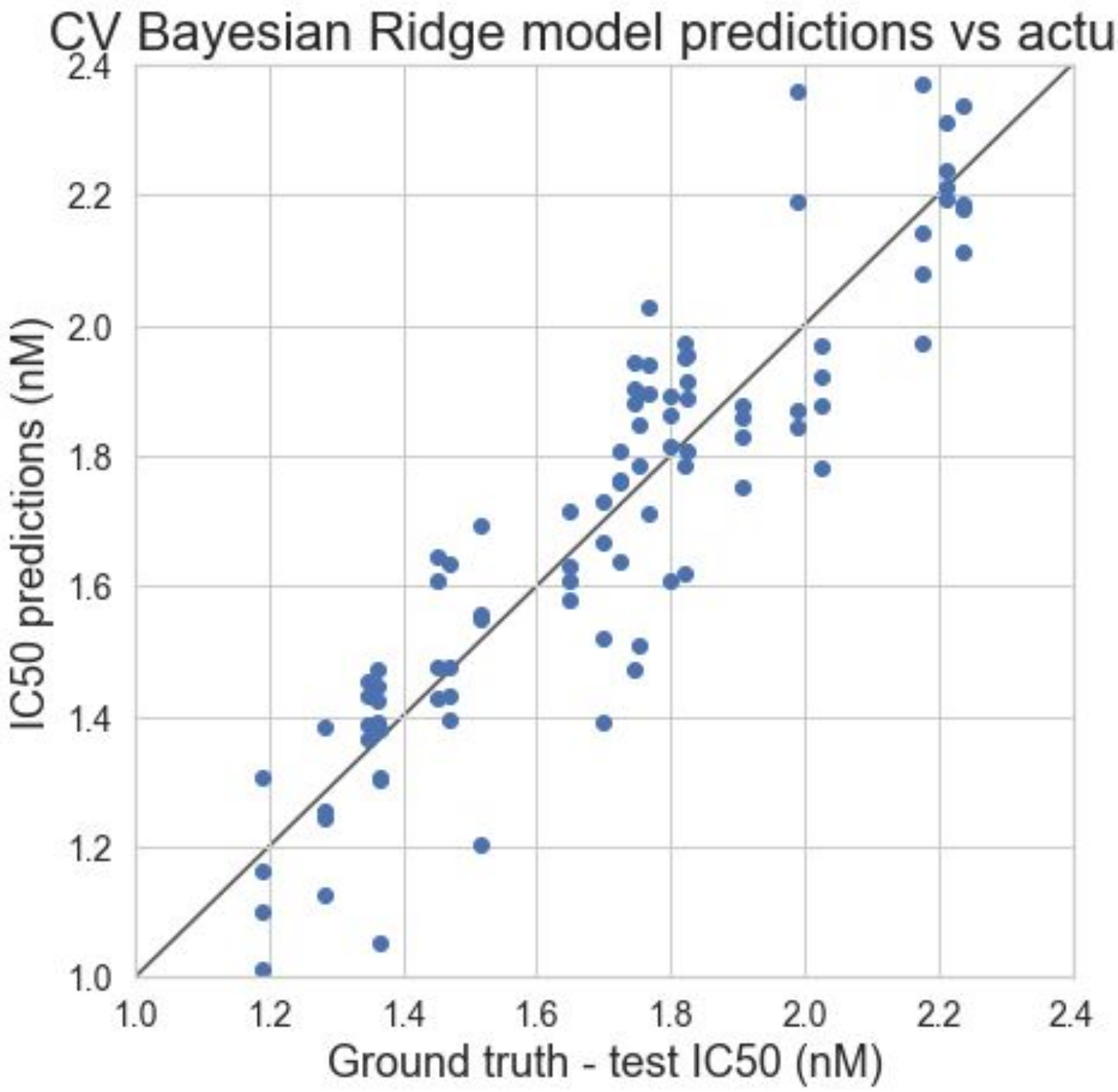


In Vitro IC50 prediction model

- Next best model: Bayesian ridge regression model trained on 100 features selected with RFE

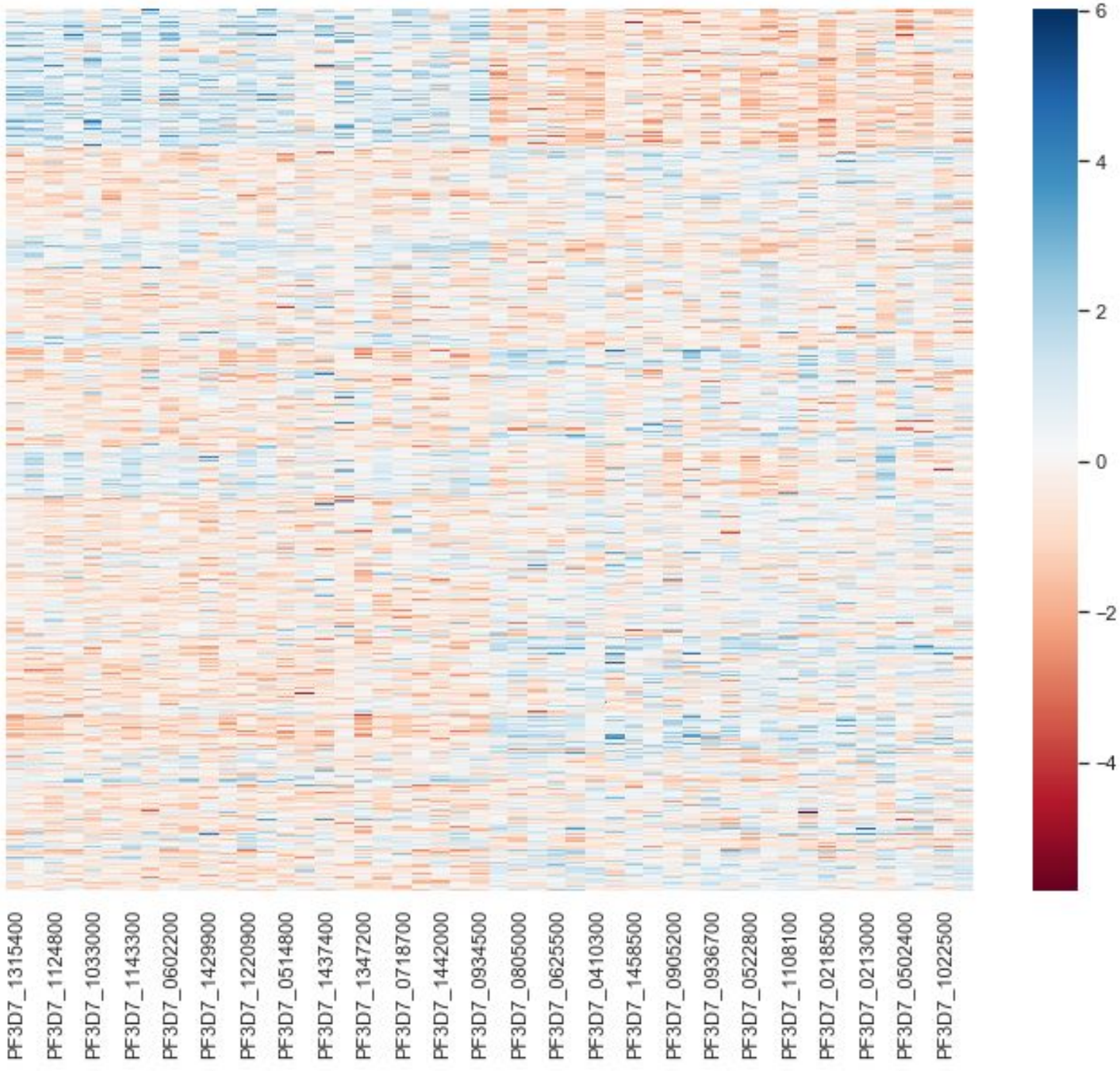
5 fold CV scores:

- R2 score: 0.90
- MAE: 0.11



In Vivo Microarray Data:

Top 50 differentially expressed genes between slow and fast parasite clearance



Artemisinin IC50 and Parasite Clearance Rates

- There is no proven relationship between *P falciparum* DHA IC50 (*in vitro*), and rate of parasite clearance (*in vivo*).
- Theorized reasons for this include the challenge of estimating parasite life cycle stage *in vivo* and inability to control for perturbing variables that could affect artemisinin sensitivity

Clearance Rate Classification Model:

- Tested SVM, logistic regression, random forest, LDA, extra trees, KNN

- Tested models again using feature selection:

  - SVM feature importance used to train random forest, SVM, and logistic regressions.

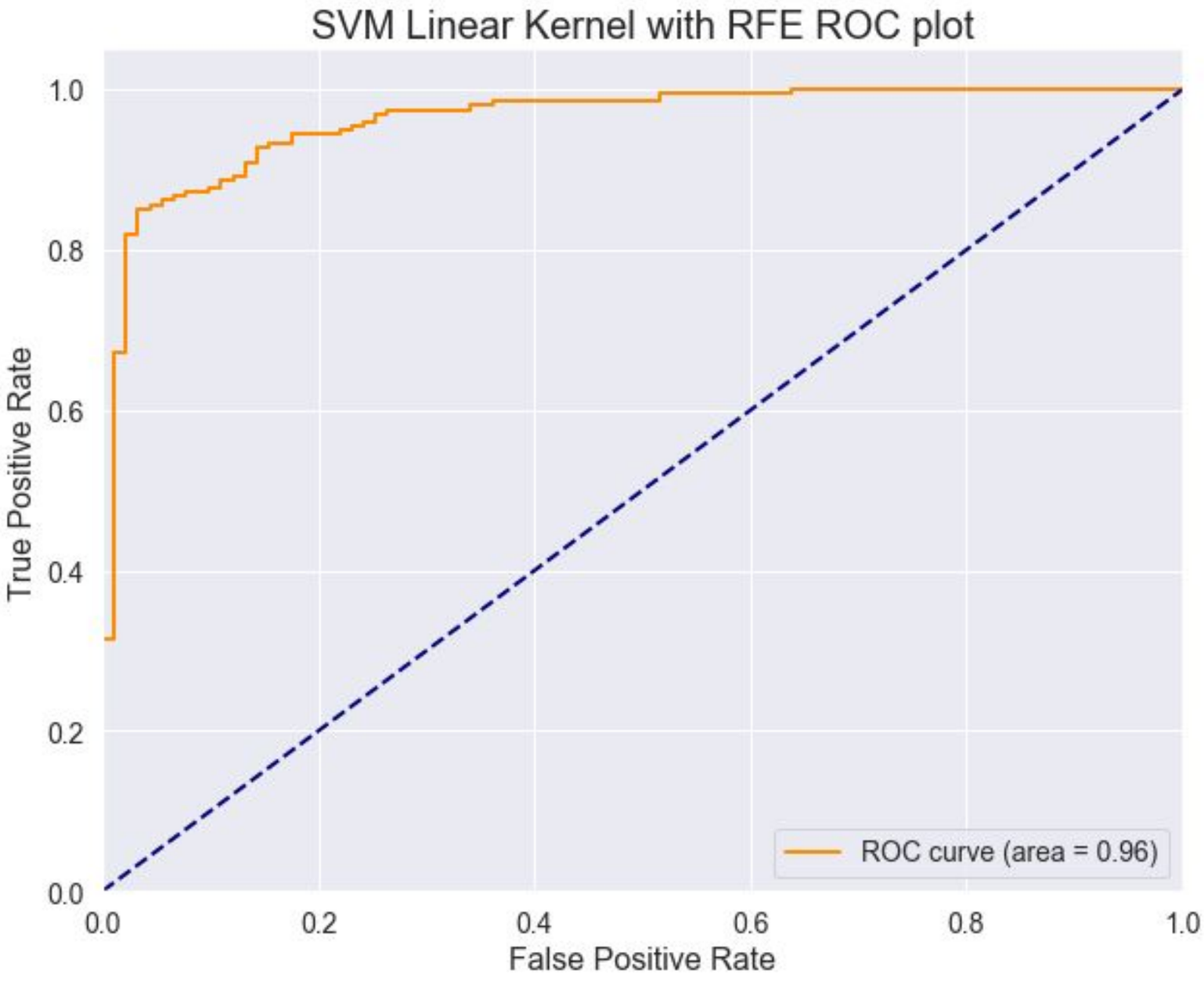
  - Recursive feature selection used to select 50 and 100 features. Best model performance yielded by SVM linear kernel applied to top 100 RFE selected features

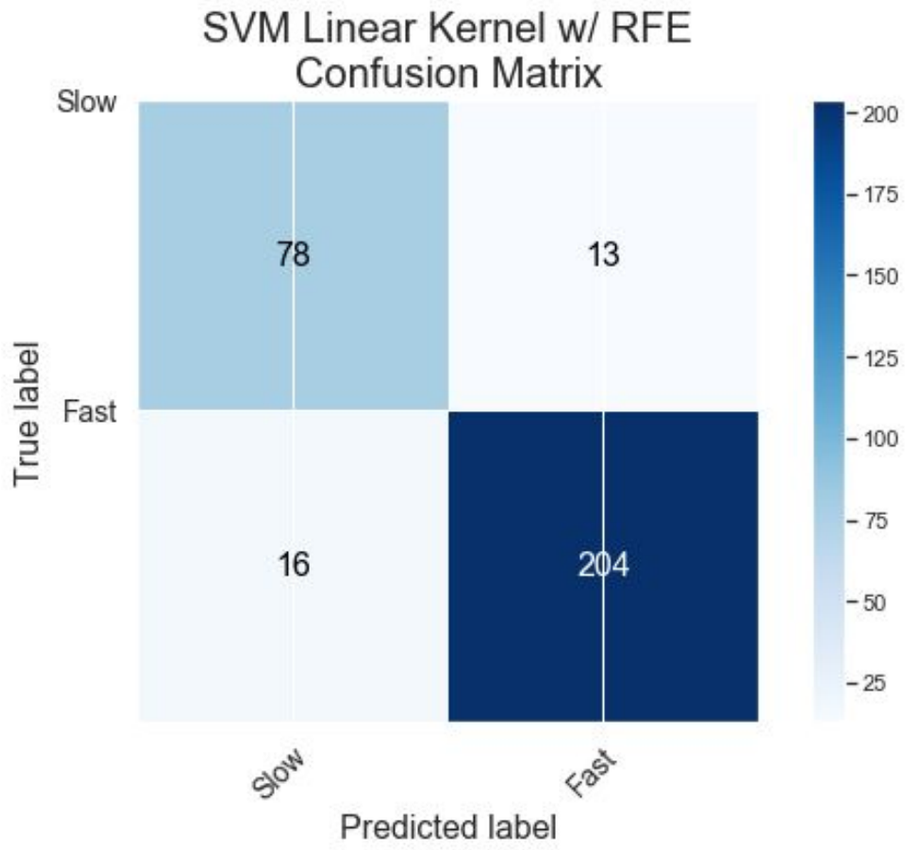
- Hyperparameter tuning performed on most promising models

Clearance Rate Classification Models:

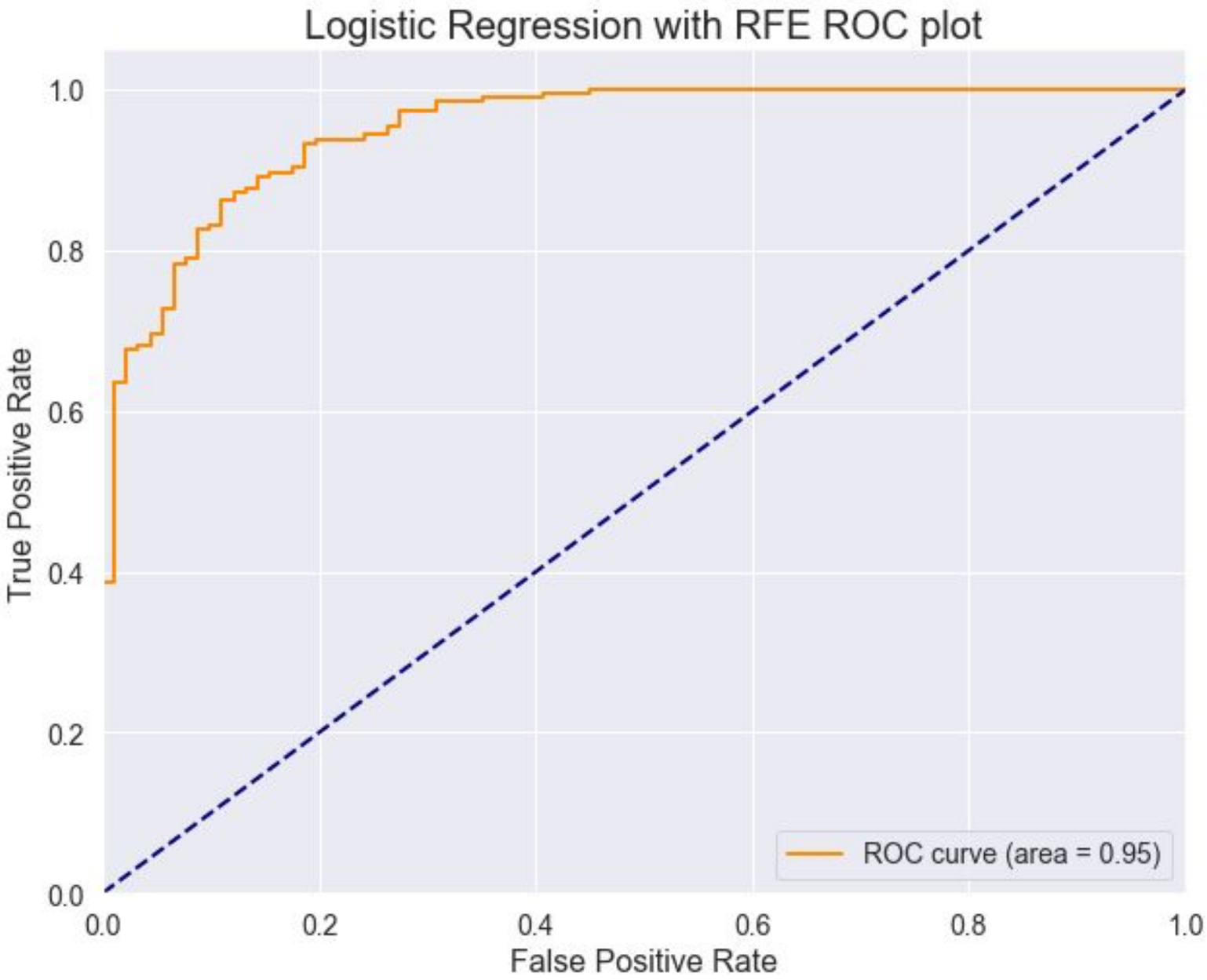
MODEL:	AUC:	Accuracy:
Random Forest	0.765	0.736
SVM Linear Kernel	0.855	0.810
Logistic Regression	0.868	0.820
Random Forest (SVM feature selection)	0.734	0.723
SVM (SVM feature selection)	0.706	0.707
Logistic Regression (SVM feature selection)	0.751	0.751
Random Forest (RFE 100 features)	0.805	0.742
SVM (RFE 100 features	0.961	0.907
Logistic Regression (RFE 100 features)	0.951	0.894

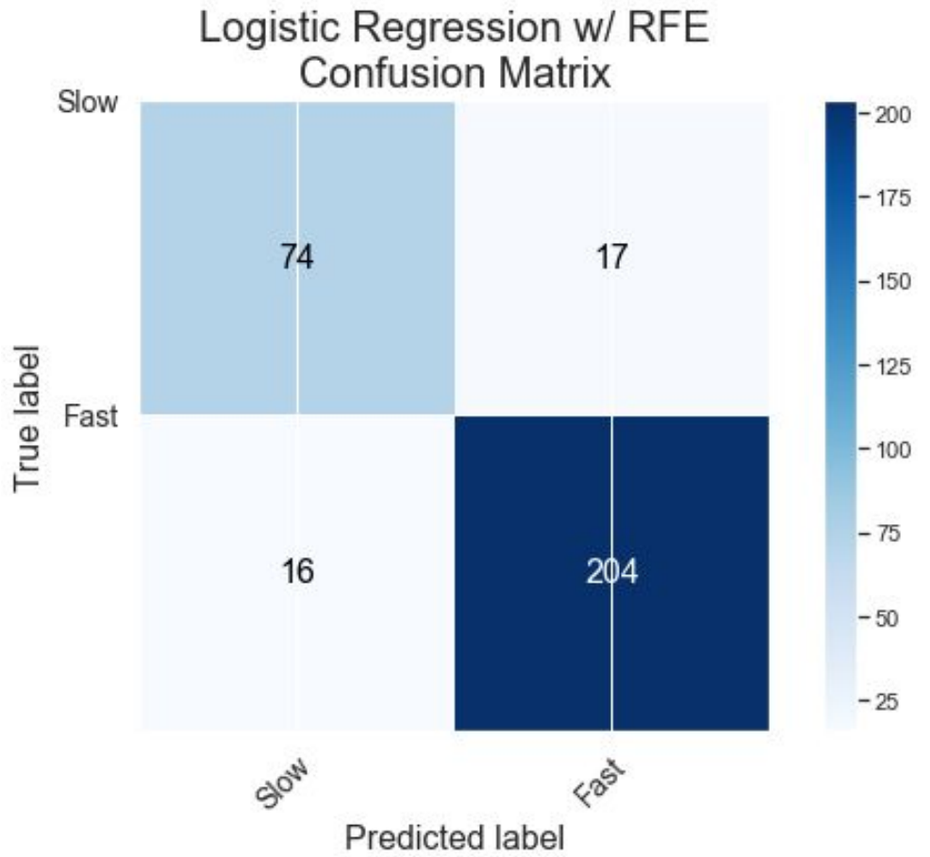
Clearance Rate Classification Model:





Clearance Rate Classification Model:





Selected features *in vitro* vs *in vivo* models

Microarrays used to generate *in vitro* and *in vivo* datasets were slightly different

	***In Vivo*:**	***In Vitro*:**
Array	**Bozdech**	**Agilent HD Exon Array**
Platform	Printed	Agilent
Plexes	1	8
Unique Probes	10159	62976
Range of Probes Per Exon	N/A	1–52
Average Probes Per Gene	2	12
Genes Represented	5363	5440
Transcript Isoform Profiling	No	Yes
ncRNAs	No	Yes
Channel Detection Method	Two Color	Single Color
Scanner	PowerScanner	Agilent
Data Extraction	GenePix Pro	Agilent

Adapted with permission from Turnbull et al (2017) PLOS ONE

Intersection of features between *in vitro* and *in vivo* models

Compared 100 RFE selected features for both models; only 4 features intersected:

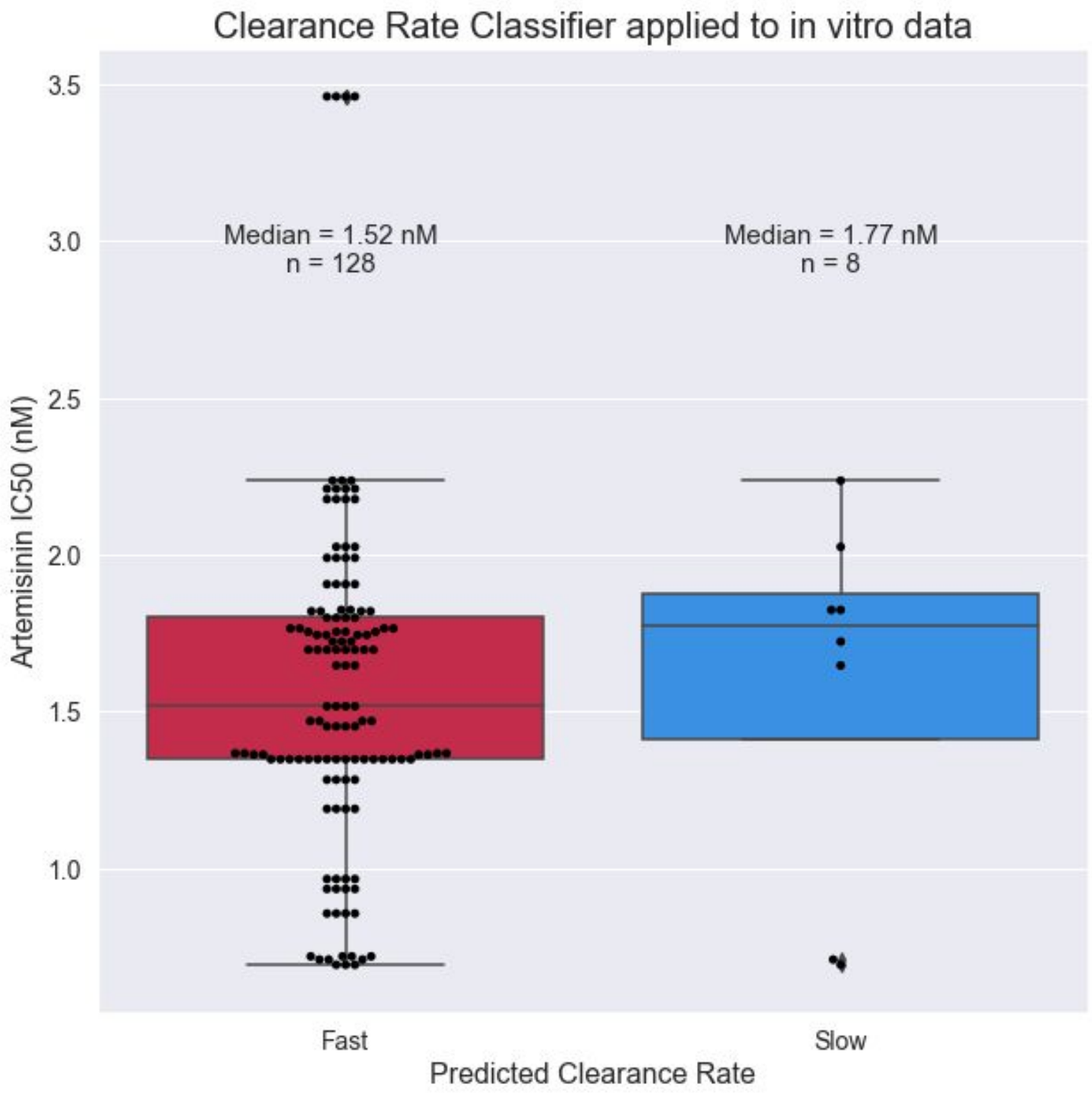
PfEMP1: (PF3D7_1240400)
A surface protein expressed by a VAR gene. This family of genes encodes a variety of antigenically variant adhesion proteins, necessary for *P falciparum* infection of RBCs (Su et al.,1995)

FIKK9.7 serine/threonine protein kinase: (PF3D7_0902600)
Potential function in regulating antigenic and cytoadhesion variation (Nunes et al., 2007)

Merozoite Surface Protein 8: (PF3D7_0502400)
A surface protein expressed during the ring-stage of P falciparum infection (early in the blood infection stage). Interuption of this protein does not inhibit parasite replication (Drew et al., 2005))

HYP12: (PF3D7_1001000)
An unknown exported protein (Vincensini et al., 2005)

Cross-applied classifier:



Sources

Antinori, S., Galimberti, L., Milazzo, L., & Corbellino, M. (2012). Biology of human malaria plasmodia including Plasmodium knowlesi. *Mediterranean journal of hematology and infectious diseases*, *4*(1), e2012013. doi:10.4084/MJHID.2012.013

“Malaria Biology.” CDC. Retrieved from https://www.cdc.gov/malaria/about/biology

Dong, Y., Wang, J., Sun, A., Deng, Y., Chen, M., Xu, Y., & Xue, J. (2018). Genetic association between the Pfk13 gene mutation and artemisinin resistance phenotype in Plasmodium falciparum isolates from Yunnan Province, China. *Malaria journal*, *17*(1), 478.

Drew, D. R., Sanders, P. R., & Crabb, B. S. (2005). Plasmodium falciparum merozoite surface protein 8 is a ring-stage membrane protein that localizes to the parasitophorous vacuole of infected erythrocytes. *Infection and immunity*, *73*(7), 3912-3922.

Mok, S., Ashley, E. A., Ferreira, P. E., Zhu, L., Lin, Z., Yeo, T., ... & Nguon, C. (2015). Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science*, *347*(6220), 431-435.

Nunes, M. C., Goldring, J. D., Doerig, C., & Scherf, A. (2007). A novel protein kinase family in Plasmodium falciparum is differentially transcribed and secreted to various cellular compartments of the host cell. *Molecular microbiology*, *63*(2), 391-403.

Su, X. Z., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., ... & Wellems, T. E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. *Cell*, *82*(1), 89-100.

Vincensini, L., Richert, S., Blisnick, T., Van Dorsselaer, A., Leize-Wagner, E., Rabilloud, T., & Breton, C. B. (2005). Proteomic analysis identifies novel proteins of the Maurer’s clefts, a secretory compartment delivering Plasmodium falciparum proteins to the surface of its host cell. *Molecular & cellular proteomics*, *4*(4), 582-593.

World Health Organization (2017). *World Malaria Report 2018*. Retrieved from https://www.who.int/malaria/publications/world-malaria-report-2017/report/en/