The Malaria DREAM challenge: team TPOT's sub-challenge 2 write-up

This manuscript ([permalink](#)) was automatically generated from [trang1618/plasmodium-falciparum@98e0f2c](#) on August 13, 2019.

Authors

- **Alena Orlenko**
  ⓘ 0000-0003-1757-293X · ⓞ desmidium · 🐦 desmidium
  Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104

- **Trang T. Le**
  ⓘ 0000-0003-3737-6565 · ⓞ trang1618 · 🐦 trang1618
  Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104

- **Weixuan Fu**
  ⓘ 0000-0002-6434-5468 · ⓞ weixuanfu · 🐦 weixuanfu
  Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104

- **Jason H. Moore**
  ⓘ 0000-0002-5015-1099 · ⓞ EpistasisLab · 🐦 moorejh
  Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104 · Funded by National Institutes of Health Grant Nos. LM010098, LM012601, AI116794

Abstract

In the subchallenge 2, the training set were collected very differently in comparison with the test set. Especially, training set is *in vivo* transcription data set described in [1] while test set is *in vitro* transcription data set. To adjust know batch effects, ComBat algorithm[2] in sva package[doi:10.1093/bioinformatics/bts034] was applied on the high-throughput transcription data and principal component analysis plots on the raw data and adjusted data were made for checking the performance of adjustment (Fig. 1).
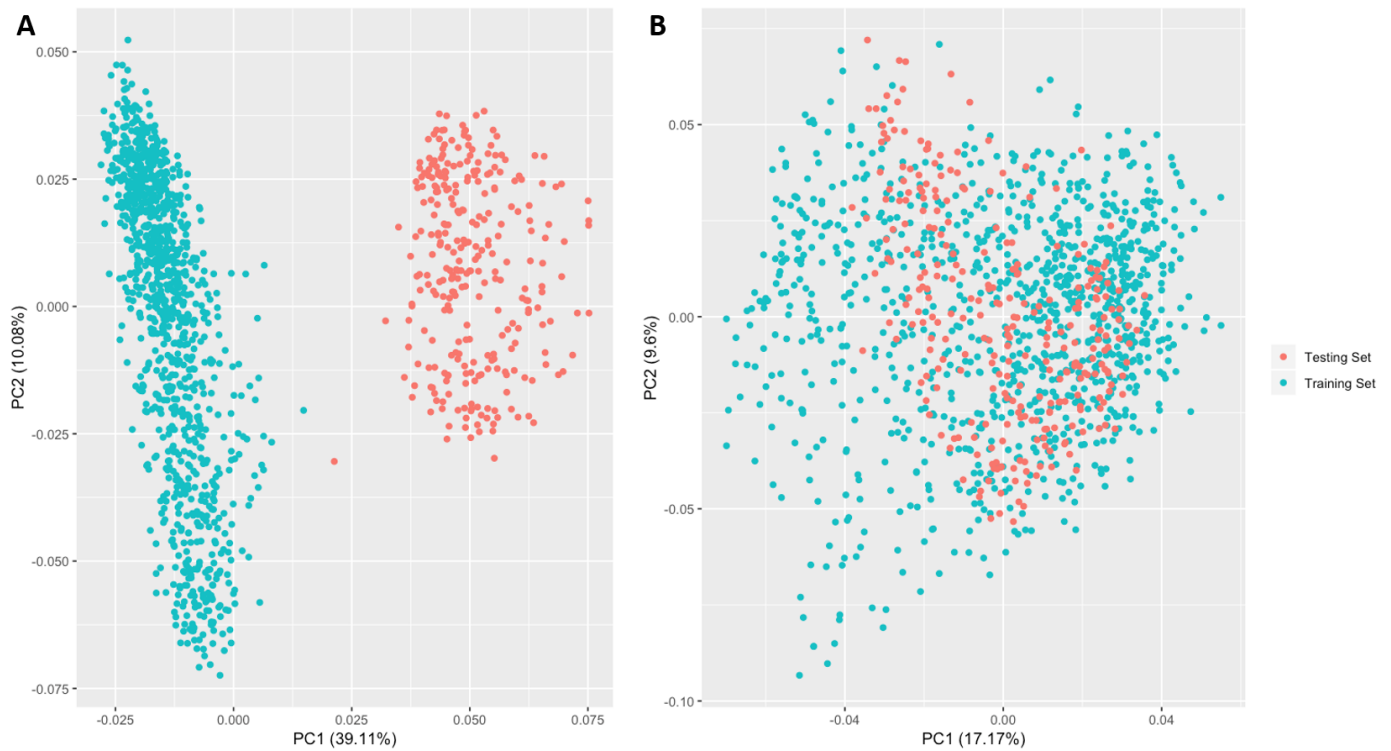
Figure 1: Principal component analysis plots before (A) and after (B) adjusting batch effects

References

1. **Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance**
S. Mok, E. A. Ashley, P. E. Ferreira, L. Zhu, Z. Lin, T. Yeo, K. Chotivanich, M. Imwong, S. Pukrittayakamee, M. Dhorda, … Z. Bozdech
Science (2014-12-11) https://doi.org/f3ph2f
DOI: 10.1126/science.1260403 · PMID: 25502316 · PMCID: PMC5642863

2. **Adjusting batch effects in microarray expression data using empirical Bayes methods**
W. Evan Johnson, Cheng Li, Ariel Rabinovic
Biostatistics (2006-04-21) https://doi.org/dsf386
DOI: 10.1093/biostatistics/kxj037 · PMID: 16632515