RECOMMENDED CITATION
Fukutani E, Rodrigues M, Kasprzykowski JI, de Araujo CF, Paschoal AR, Ramos PIP, et al. Follow up of a robust meta-signature to identify Zika virus infection in *Aedes aegypti*: Another brick in the wall [Submitted]. Mem Inst Oswaldo Cruz E-pub: 31 Jan 2018. doi: 10.1590/0074-02760180053.

Running Title: Meta-signature in Aedes aegypti.

Title: Follow up of a robust meta-signature to identify Zika virus infection in *Aedes aegypti*: Another brick in the wall.

Author's names: Eduardo Fukutani[1,£], Moreno Rodrigues[1,£], José Irahe Kasprzykowski[1,2], Cintia Figueiredo de Araujo[3], Alexandre Rossi Paschoal[4], Pablo Ivan Pereira Ramos[1,&], Kiyoshi Ferreira Fukutani[5,a,&] and Artur Trancoso Lopo de Queiroz[1,2,&]

Institutional affiliations: [1] *Instituto Gonçalo Moniz, Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Brazil;* [2] *Post-Graduation Program in Biotechnology in Health and Investigative Medicine, Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Brazil;* [3] *Serviço de Imunologia. Federal University of Bahia. Salvador, Brazil;* [4] *Federal University of Technology — Paraná, UTFPR, Campus Cornélio Procópio, Cornélio Procópio, Brazil;* [5] *Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil.*

[a] corresponding author: ferreirafk@gmail.com
[£] equal contribution
[&] equal contribution

Abstract.

The mosquito *Aedes aegypti* is the main vector of several arthropod-borne diseases with global impact. In a previous meta-analysis, our group has identified a vector gene set comprised by 110 genes strongly associated to infection by Dengue, West Nile and Yellow Fever viruses, of which four genes allowed highly accurate classification of infected status. More recently, a new study of *A. aegypti* infected with

Zika virus (ZIKV) was published, providing new data to investigate whether this "infection" gene set is also altered during ZIKV infection. Our hypothesis is that the infection-associated signature may serve as a proxy also to classify zika virus infection in the vector. Raw data associated to the NCBI/BioProject was downloaded and re-analysed. A total of 18 paired-end replicates corresponding to 3 ZIKV-infected and 3 controls were included in this study. The nMDS technique with a logistic regression was used to obtain the probabilities of belonging to a given class. Thus, to compare both gene sets we used the area under the curve and performed a comparison using the bootstrap method. Our meta-signature was able to separate the infected mosquitoes from the controls, with good predictive power to classificate the Zika-infected mosquitoes.

Keywords: RNA-seq; Signature; Transcriptome; Zika Virus.

The mosquito *Aedes aegypti* (L.) is the main vector of several globally distributed diseases (Lorenzo et al., 2014). One of these diseases is dengue (DENV), affecting more than 2.5 billion people (WHO, 2015). Moreover, other illnesses like yellow fever (YFV) is endemic in tropical regions (Bae et al., 2005), and has recently reemerged in close proximity to major urban centers in Brazil (Paules and Fauci, 2017), while West Nile fever (WNF), which usually associates to small outbreaks, presents high mortality rates (Pradier et al., 2012). Chikungunya virus (CHIKV), once localized to parts of Africa, is now globally spread (Cauchemez et al., 2014). Zika virus (ZIKV) has emerged in 2015 in the Americas (Zanluca et al., 2015; Faria et al., 2016; Slavov et al., 2016) following sporadic outbreaks in the Pacific in 2007 (Micronesian island Yap) and 2013-14 (French Polynesia). Initially considered an uncomplicated, self-limited disease, Zika infection was later associated to the 'microcephaly outbreak' that followed (Oliveira Melo et al., 2016; Tang et al., 2016), leading Brazilian authorities to declare a state of national health emergency and the World Health Organization to designate the Zika epidemic as a public health emergency of international concern (WHO, 2016).

Previous studies were performed to elucidate altered pathways in *A. aegypti* in response to viral infection, and our group recently identified a viral infection meta-signature for DENV, WNV and DENV by investigating the relationship between alimentation and infection (Fukutani et al., 2018). We identified a set of 110 genes highly correlated to viral infection, of which 4 genes (AAEL012128, AAEL014210, AAEL002477, and AAEL005350) were highly informative to infection-status classification in the vector.

The datasets used in this previous study lacks in ZIKV-infected samples, unavailable at the time. Recently, a next generation sequencing study with *A. aegypti* infected with ZIKV was published (Etebari et al., 2017) with a new data, allowing the test of our signature in this infection. The hypothesis is that the expression of signature genes, as we have demonstrated for other viruses, also plays an important role in the classification of ZIKV infection in the mosquito. To achieve this study, we assessed the NCBI/BioProject PRJNA399504 (https://www.ncbi.nlm.nih.gov/bioproject/399504) and downloaded all raw data. In total there are 18 paired-end replicates corresponding to six samples (3 ZIKV-infected and 3 controls) (Etebari et al. 2017). The fastq-dump from the SRA toolkit (NCBI, 2011) was used to obtain sequence files in FASTQ format. Sequences were filtered for low quality reads and adapters using Trimmomatic (Bolger et al., 2014) and transcripts were quantified using the *A. aegypti* reference transcriptome AaegL3.5 as reference within Salmon v0.9.1 (Patro et al, 2017). Transcripts were summarized at the gene-level using the R package *tximport* (Soneson et al, 2015), yielding a count table. The count table was filtered using the *edgeR* package (Robinson et al, 2010) and only genes consistently expressed (counts per million [cpm] greater than 0.5) were kept. Each sample-expression values were determined by its sample-replicates mean, as specified by the BioProject metadata. To test our signature, a nonmetric multidimensional scaling (NMDS) was performed using the *metaMDS* function within R package *vegan* v2.4.5 (Oksanen et al 2017). *metaMDS* aims to represent the position of samples in a multidimensional space, as accurately as possible, using a reduced number of dimensions (axes). The Axes resulted from metaMDS were used in a logistic regression model to access the predicted probability of a sample belonging to a given class (ie. ZIKV- or Mock-infected). To compare both models (the complete set of 110 genes or the restricted set of 4 genes) the area under the Receiver operating characteristic (ROC) curve (AUC) was calculated using the *pROC* package (Robin et al, 2011).

First, we tested our previous identified signature composed by 110 genes. This set consists of the previously correlated genes with infection without the blood-feeding influence (Fukutani et al., 2018). It was able to discriminate the groups (infected mosquitoes with ZIKV and uninfected samples) by nMDS model (Figure 1A). The same approach was used with the smaller gene set (4 genes) (Figure 1B). Both 110 and 4 gene sets were able to discriminate the ZIKV-infected from uninfected samples with $7x10^{-2}$ and $2x10^{-4}$ stress values respectively (lower stress values indicate more reliable ordination of the dataset). To measure the accuracy of each gene set, we calculate the area under curve (AUC); using 110 gene dataset an AUC of 0.94 was found whereas using a 4 gene dataset this value was 0.83 (Figure 2). However, there is no statistical difference between both gene set AUC (D = -1.48, boot.n = 2000, p-value = 0.13).

Despite our original meta-signature was identified for the mosquitoes classification infected with YFV, WNF and DENV, when we applied in ZIKV-infected mosquitos, these genes were able to discriminate the infected and healthy samples. In addition, irrespective of the set of genes can chosen (110 or 4 geneset), the signature remains robust with good predictive power for the classification of the Zika infection.

The limitation of the current undertaking is the relatively low number of samples re-analysed (*n*=6). However, there is few dataset available of mosquitoes infected with Zika virus. Another dataset available of *A. aegypti* (GSE96605) does not have adequate numbers of sample (3 samples: 1 DENV, 1 ZIKV and 1 Control with 3 technical replicates)(Angleró-Rodríguez et al, 2017). Despite this limitation, our results showed that these gene sets are a powerful framework for future studies of vector infection. Applications of this meta-signature are promising and suggests a process that could be similar in others vector infections such as CHIKV and Oropouche virus. However, as of yet there are no publicly available datasets of vectors infected with these virus. These results together update the prediction power of our previously identified meta-signature as a consistent signature that is able to identify mosquitoes infected by DENV, YFV, WNF and now ZIKV. This suggests common processes involved in all infections. Moreover, they impact in vector pathways related to maintenance of virus replication, such as host protein machinery and aminoácids transportation processes (Mosso et al., 2008).

Acknowledgements

We thank Dr. Cleyson Barros and Dr. Herculano da Silva for their insights that greatly improved the manuscript. The assistance of Ms. Raíza Tourinho and Mr. Olival Rocha.

Author's contribution:

Eduardo Fukutani, Moreno Rodrigues, José Irahe Kasprzykowski and Cintia Figueiredo de Araujo made the data analysis and download the data. Alexandre Rossi Paschoal, Pablo Ivan Pereira Ramos, Kiyoshi Ferreira Fukutani and Artur Trancoso Lopo de Queiroz wrote the Manuscript.

Sponsorships:

The financial supports of Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB process no. JCB0004/2013), Fundação de Amparo à Pesquisa do Estado de

São Paulo (FAPESP no. 2017/03491-6) and CNPq Universal MCTI/CNPQ/Universal14/2014 (no. 454505/2014-0).

Reference

1. Lorenzo, M. G., Vidal, D. M., and Zarbin, P. H. G. Control of neglected disease insect vectors: future prospects for the use of tools based on behavior manipulation-interference. J. Braz. Chem. Soc. 2014; 25: 1799–1809.

2. WHO (World Health Organization) [homepage on the Internet]. Global Strategy for Dengue Prevention and Control. [updated 2015; cited 2018] Available from: http://www.who.int/denguecontrol/9789241504034/en/.

3. Bae, H.-G., Drosten, C., Emmerich, P., Colebunders, R., Hantson, P., Pest, S. Analysis of two imported cases of yellow fever infection from Ivory Coast and the Gambia to Germany and Belgium. J. Clin. Virol. 2005. 33: 274–280.

4. Pradier, S., Lecollinet, S., and Leblond, A. West Nile virus epidemiology and factors triggering change in its distribution in Europe. Rev. Sci. Tech. 2012. 31: 829–844.

5. Cauchemez, S., Ledrans, M., Poletto, C., Quenel, P., de Valk, H., Colizza, V. Local and regional spread of chikungunya fever in the Americas. Euro Surveill. 2014. 19.

6. Zanluca, C., Melo, V. C., Mosimann, A. L., Santos, G. I., Santos, C. N., and Luz, K. First report of autochthonous transmission of Zika virus in Brazil. Mem. Inst. Oswaldo Cruz 2015. 110: 569–572.

7. Slavov, S. N., Otaguiri, K. K., Kashima, S., and Covas, D. T. Overview of Zika virus (ZIKV) infection in regards to the Brazilian epidemic. Braz. J. Med.Biol. Res. 2016. 49.
8. Fukutani K.F., Kasprzykowski J.I., Paschoal A.R., Gomes M.d.S., Barral A., de Oliveira C.I., Ramos P.I.P. and Queiroz A.T.L. Meta-Analysis of *Aedes aegypti* Expression Datasets: Comparing Virus Infection and Blood-Fed Transcriptomes to Identify Markers of Virus Presence. *Front. Bioeng. Biotechnol.* 2018. 5(84).

9. Etebari K., Hegde S., Saldaña M.A. Global Transcriptome Analysis of *Aedes aegypti* Mosquitoes in Response to Zika Virus Infection. Fernandez-Sesma A, ed. *mSphere*. 2017; 2(6): 456-17.

10. National Center for Biotechnology Information (US) [homepage on the Internet]. Sequence Read Archive Submissions Staff. Using the SRA Toolkit to convert .sra files into other formats. In: SRA Knowledge Base. Bethesda (MD): [updated 2011; cited 2018]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK158900/.

11. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017.

12. Soneson C, Love MI and Robinson MD. “Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.” *F1000Research*, 2015. 4.

13. Robinson MD, McCarthy DJ and Smyth GK. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.” Bioinformatics. 2010. 26(1): 139-140.

14. Oksanen, Jari, F Blanchet Guillaume, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, R. Minchin, et al. “Vegan : Community Ecology Package”. 2017. 1(2): 1–12.

15. Robin X, Turck N, HainardA, Tiberti N, Lisacek F, Sanchez J and Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011. 12: 77.

16. Angleró-Rodríguez YI, MacLeod HJ, Kang S, Carlson JS, Jupatanakul N and Dimopoulos G. Aedes aegypti Molecular Responses to Zika Virus: Modulation of Infection by the Toll and Jak/Stat Immune Pathways and Virus Host Factors. *Front. Microbiol.* 2017. 8:2050.

17. Mosso, C., Galván-Mendoza, I. J., Ludert, J. E., and del Angel, R. M. Endocytic pathway followed by dengue virus to infect the mosquito cell line C6/36 HT. Virology. 2010. 378: 193–199.

18. Paules and Fauci. N Engl J Med 2017; 376:1397-1399 DOI: 10.1056/NEJMp1702172

19. Faria, Nuno Rodrigues, et al. "Zika virus in the Americas: early epidemiological and genetic findings." Science 352.6283 (2016): 345-349.

20. WHO (World Health Organization). 2016. Available from: http://www.who.int/mediacentre/news/statements/2016/emergency-committee-zika-microcephaly/en/

21. Tang, Hengli, et al. "Zika virus infects human cortical neural progenitors and attenuates their growth." Cell stem cell 18.5 (2016): 587-590.

22. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

Figure Legend

Figure 1: Nonmetric multidimensional scaling analysis (NMDS) based on Bray dissimilarity index from 110 gene set (A) and 4 gene set (B) to discriminate infected (blue triangles) and uninfected mosquitoes (red circles).

Figura 2. Receiver operating characteristic (ROC) curve for the gene sets. The area under curve (AUC) for predicting ZIKV-infection was 0.94 for 110 gene (blue) set and 0.83 for 4 gene (red) set.

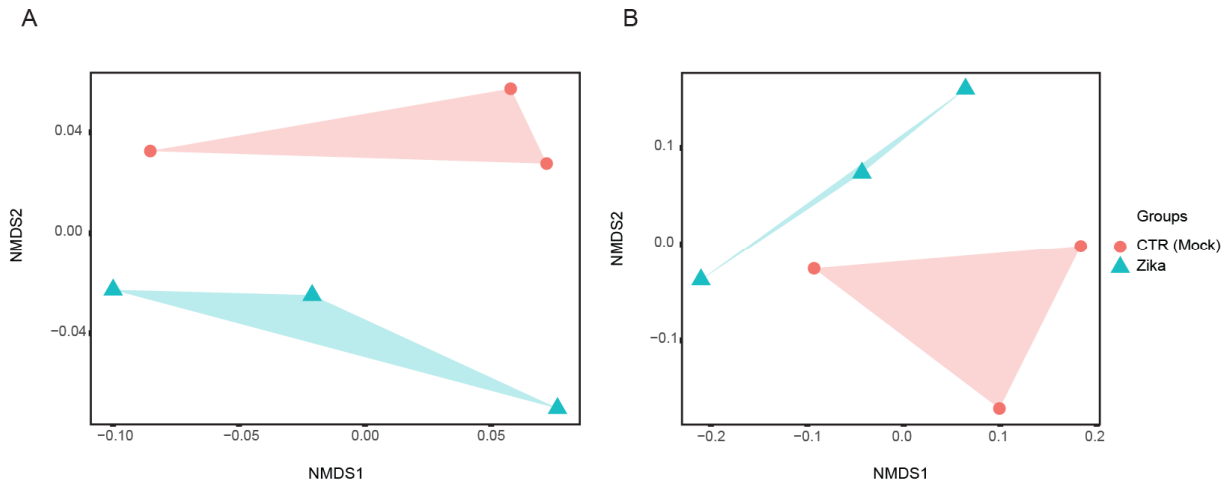

Figure 1

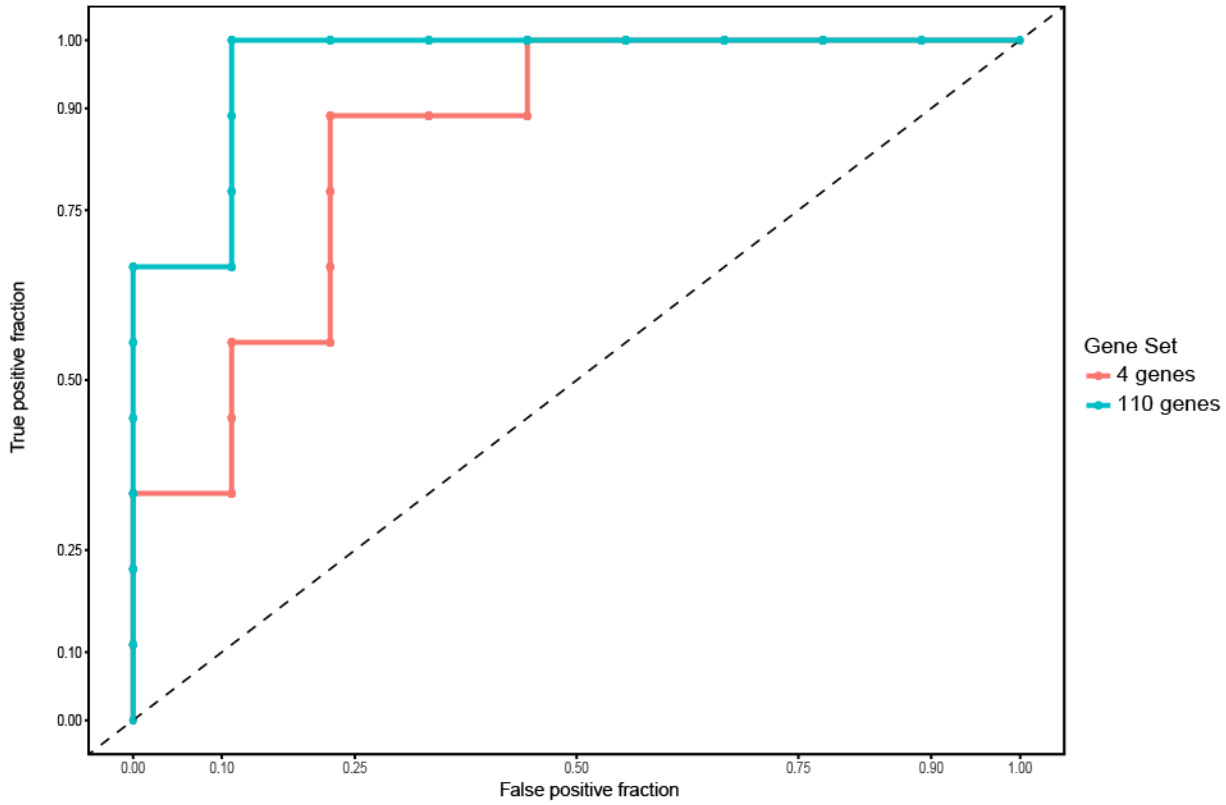

Figure 2