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

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[5,6]/+, Artur Trancoso Lopo de Queiroz[1,2]

[1]Fundação Oswaldo Cruz-Fiocruz, Instituto Gonçalo Moniz, Salvador, BA, Brasil
[2]Fundação Oswaldo Cruz-Fiocruz, Programa de Pós-Graduação em Biotecnologia em Saúde Investigativa, Salvador, BA, Brasil
[3]Universidade Federal da Bahia, Serviço de Imunologia, Salvador, BA, Brasil
[4]Universidade Tecnológica Federal do Paraná, Cornélio Procópio, PR, Brasil
[5]Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, SP, Brasil
[6]Universidade Salvador, Salvador, BA, Brasil

The mosquito *Aedes aegypti* is the main vector of several arthropod-borne diseases that have global impacts. In a previous meta-analysis, our group identified a vector gene set containing 110 genes strongly associated with infections of dengue, West Nile and yellow fever viruses. Of these 110 genes, four genes allowed a highly accurate classification of infected status. More recently, a new study of *Ae. aegypti* infected with Zika virus (ZIKV) was published, providing new data to investigate whether this "infection" gene set is also altered during a ZIKV infection. Our hypothesis is that the infection-associated signature may also serve as a proxy to classify the ZIKV infection in the vector. Raw data associated with the NCBI/BioProject were downloaded and re-analysed. A total of 18 paired-end replicates corresponding to three ZIKV-infected samples and three 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 classify the Zika-infected mosquitoes.

Key words: 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) that affects more than 2.5 billion people (WHO 2015). Moreover, other illnesses such as yellow fever (YFV) are endemic to tropical regions (Bae et al. 2005) and have recently re-emerged in close proximity to major urban centres in Brazil (Paules and Fauci 2017), while West Nile fever (WNF), which usually is associated with small outbreaks, presents high mortality rates (Pradier et al. 2012). Chikungunya virus (CHIKV), once localised to parts of Africa, has now spread globally (Cauchemez et al. 2014). Zika virus (ZIKV) 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, it was not considered dangerous as it is a self-limiting disease, but later, it was associated with a 'microcephaly outbreak' (Tang et al. 2016), leading Brazilian authorities to declare a national health state of 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 of *Ae. aegypti* in response to viral infections, and our group recently identified a viral infection meta-signature for DENV, WNV and YFV by investigating the relationship between feeding and infection (Fukutani et al. 2018). We identified a set of 110 genes highly correlated with viral infection, of which four genes (AAEL012128, AAEL014210, AAEL002477, and AAEL005350) were highly informative in identifying the infection-status classification in the vector. The role of the genes AAEL014210, AAEL002477 and AAEL005350 is regulatory based on the prediction of a zinc finger, DNA-binding domain (InterPro accession no. IPR013087), and basic-leucine zipper domain (InterPro accession no. IPR004827), and these genes harbour retinaldehyde binding and alpha-tocopherol transport domains (InterPro accession nos. IPR001071 and IPR001251, respectively). The gene AAEL012128 is described as having a 12-pass transmembrane protein with a cationic amino acid transporter and was previously reported as a retrovirus receptor (Wang et al. 1991, Fukutani et al. 2018).

The datasets used in the previous studies lacked ZIKV-infected samples, which were unavailable at the time. Recently, a next-generation sequencing study with *Ae. aegypti* infected with ZIKV was published (Etebari

doi: 10.1590/0074-02760180053
Financial support: FAPESB (process no. JCB0004/2013), FAPESP (no. 2017/03491-6), CNPq Universal MCTI/CNPQ/Universal14/2014 (no. 454505/2014-0).
+ Corresponding author: ferreirafk@gmail.com
EF, MR, KFF, PIPR and ATLQ contributed equally to this work.
Received 29 January 2018
Accepted 17 April 2018

CC BY

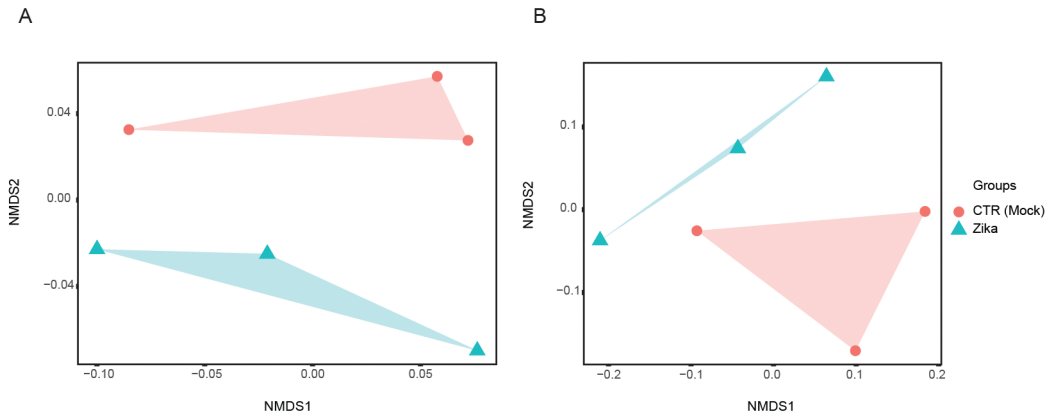

Fig. 1: nonmetric multidimensional scaling analysis (NMDS) based on the Bray dissimilarity index from the 110 gene set (A) and four gene set (B) to discriminate infected (blue triangles) and uninfected mosquitoes (red circles).

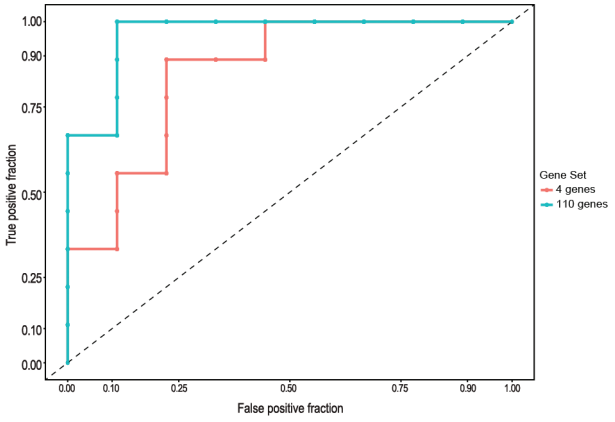

Fig. 2: receiver operating characteristic (ROC) curve for the gene sets. The area under the curve (AUC) for predicting Zika virus infection was 0.94 for the 110 gene (blue) set and 0.83 for the four gene (red) set.

et al. 2017) with a new data, allowing a test of our signature in this infection. Our hypothesis is that the expression of signature genes, as we have demonstrated for other viruses, also plays an important role in the classification of the 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 the raw data. In total, there are 18 paired-end replicates corresponding to six samples (three ZIKV-infected and three 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 version 0.32 GNU (Bolger et al. 2014), and transcripts were quantified using the *Ae. aegypti* reference transcriptome AaegL3.5 as a reference within Salmon v0.9.1 (Patro et al. 2017). Transcripts were summarised 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 value was 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 the 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 resulting from metaMDS were used in a logistic regression model to access the predicted probability of a sample belonging to a given class (i.e., ZIKV- or mock-infected). To compare both models (the complete set of 110 genes or the restricted set of four 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 previously identified signature composed of 110 genes. This set consists of the previously correlated genes with infection without the blood-feeding influence (Fukutani et al. 2018). The nMDS model was able to discriminate the groups (infected mosquitoes with ZIKV and uninfected samples) (Fig. 1A). The same approach was used with the smaller gene set (four genes) (Fig. 1B). Both the 110 and four gene sets were able to discriminate the ZIKV-infected from the uninfected samples with 7 x 10^{-2} and 2 x 10^{-4} stress values, respectively (lower stress values indicate a more reliable ordination of the dataset). To measure the accuracy of each gene set, we calculated the area under curve (AUC); using the 110 gene dataset, an AUC of 0.94 was found, whereas using a four gene dataset, the value was 0.83 (Fig. 2). However, there is no significant difference between both the AUC of both gene sets (D = -1.48, boot.n = 2000, p-value = 0.13). Despite this result, our original meta-signature was identified for the mosquito classification infected with YFV, WNF and DENV, and when applied in ZIKV-infected mosquitoes, these genes were able to discriminate the infected and healthy samples. In addition, irrespective of the set of genes selected (the 110 or four gene set), the signature remained robust with good predictive power for the classification of the Zika infection.

The limitation of the current approach is the relatively low number of samples re-analysed (n = 6). However, there are few datasets available of mosquitoes infected with the ZIKV. Another available dataset of *Ae. aegypti* (GSE96605) does not have an adequate number of samples (three samples: one DENV, one ZIKV and one control with three 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 infections. Applications of this meta-signature are promising and suggest a process that could be similar in other vector infections such as CHIKV and Oropouche virus. However, to date, there are no publicly available datasets of vectors infected with these viruses. These results improve the predictive 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 result suggests common processes are involved in all infections. Moreover, these processes impact vector pathways related to the maintenance of virus replication, such as host protein machinery and amino acid transportation processes (Mosso et al. 2010).

ACKNOWLEDGEMENTS

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

AUTHORS' CONTRIBUTION

EF, MR, JIK, CFA and ARP downloaded the data and performed the analysis; PIR, KFF and ATLQ wrote the manuscript.

REFERENCES

Angleró-Rodríguez YI, MacLeod HJ, Kang S, Carlson JS, Jupatanakul N, 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.

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(4): 274-80.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014; 30(15): 2014-20.

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(28): 20854.

Etebari K, Hegde S, Saldaña MA. Global transcriptome analysis of *Aedes aegypti* mosquitoes in response to Zika virus infection. mSphere. 2017; 2(6): pii:e00456-17.

Faria NR, Azevedo RDSDS, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: early epidemiological and genetic findings. Science. 2016; 352(6283): 345-9.

Fukutani KF, Kasprzykowski JI, Paschoal AR, Gomes MS, Barral A, de Oliveira CI, et al. 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.

Lorenzo MG, Vidal DM, Zarbin PHG. Control of neglected disease insect vectors: future prospects for the use of tools based on behavior manipulation-interference. J Braz Chem Soc. 2014; 25: 1799-809.

Mosso C, Galván-Mendoza IJ, Ludert JE, del Angel RM. Endocytic pathway followed by dengue virus to infect the mosquito cell line C6/36 HT. Virology. 2010; 378: 193-9.

NCBI - National Center for Biotechnology Information [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/.

Oksanen J, Guillaume FB, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: community ecology package. 2017; 1(2): 1-12.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017. 14(4): 417-19.

Paules CI, Fauci AS. Yellow fever - Once again on the Radar Screen in the Americas. N Engl J Med. 2017; 376:1397-9.

Pradier S, Lecollinet S, Leblond A. West Nile virus epidemiology and factors triggering change in its distribution in Europe. Rev Sci Tech. 2012; 31: 829-44.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12: 77.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1): 139-40.

Slavov SN, Otaguiri KK, Kashima S, Covas DT. Overview of Zika virus (ZIKV) infection in regards to the Brazilian epidemic. Braz J Med Biol Res. 2016; 49(5): e5420.

Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000 Res. 2015; 4: 1521.

Tang H, Hammack C, Ogden SC, Wen Z, Quian X, Li Y, et al. Zika virus infects human cortical neural progenitors and attenuates their growth. Cell Stem Cell. 2016; 18(5): 587-90.

Wang H, Kavanaugh MP, North RA, Kabat D. Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. Nature. 1991; 352: 729-31.

WHO - World Health Organization). WHO director-general summarizes the outcome of the emergency committee regarding clusters of microcephaly and Guillain-Barré syndrome. 2016. Available from: http://www.who.int/mediacentre/news/statements/2016/emergency-committee-zika-microcephaly/en/.

WHO - World Health Organization. Global strategy for dengue prevention and control. [updated 2015; cited 2018]. Available from: http://www.who.int/denguecontrol/9789241504034/en/.

Zanluca C, de Melo VCA, Mosimann ALP, dos Santos GIV, dos Santos CND, Luz K. First report of autochthonous transmission of Zika virus in Brazil. Mem Inst Oswaldo Cruz. 2015; 110(4): 569-72.