Transcriptome profiles and novel lncRNA identification of *Aedes aegypti* cells in response to dengue virus serotype 1.

Azali Azlan[1], Sattam M. Obeidat[1], Muhammad Amir Yunus[2] and Ghows Azzam[1]*

[1]*School of Biological Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia*

[2]*Infectomics Cluster, Advanced Medical & Dental Institute, Universiti Sains Malaysia, Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia.*

*Corresponding author

Email: ghows@usm.my (G.A)

Key words*: Aedes aegypti*, long non-coding RNA, dengue virus, transcriptome

Abstract

Dengue virus (DENV) is a single-stranded, positive-strand RNA virus that belongs to the family of *Flaviviridae,* and it is mainly transmitted by the mosquito *Aedes aegypti* (*Ae. aegypti*). Understanding the interaction of the virus with mosquito vector is vital for devising new strategies for preventing virus transmission. Although protein-coding genes have been the central focus, many reports indicated that long non-coding RNAs (lncRNAs) were also involved in virus-host interaction. Recently, the latest version of *Ae. aegypti* genome (AaegL5) was released, and the assembly was up to chromosome level. This prompted us to perform lncRNA identification and characterization using the latest genome release as reference. In this study, we investigated the transcriptome profiles of both protein-coding and lncRNA genes in *Aedes aegypti* cells upon DENV infection. By combining RNA-seq libraries generated in this study with publicly available datasets, we identified a total of 7,221 novel lncRNA transcripts, of which 3,052 and 3,620 were intronic and intergenic respectively, while 549 were antisense to the reference genes. A total of 2,435 differentially expressed transcripts, of which 956 of them were lncRNAs. Overall, the distribution of lncRNA expression and fold change upon virus infection were lower than that of protein-coding genes. We found that the expression of immune-related genes involved in IMD and MAPK signaling pathways were altered. In addition, the expression of major genes involved in RNA-interference (RNAi) pathway that confers antiviral resistance in mosquitoes were found to be unchanged upon DENV infection. Gene ontology analysis suggests that differentially expressed transcripts, either upregulated or downregulated, generally belong to the same functional categories or working in similar signaling pathways. Taken together, besides providing a new set of lncRNA repertoire, the outcomes of our study offer better understanding of *Ae. aegypti* responses to DENV infection at gene level.

Author Summary

Dengue virus (DENV), a single-stranded and positive-strand RNA virus of the family *Flaviviridae*, is primarily transmitted by *Aedes aegypti* (*Ae. aegypti*) mosquitoes. There are four closely related but antigenically different serotypes of dengue virus namely DENV1-4. Our understanding on the interaction of each serotype of DENV with its mosquito vector is still very limited. Since vector-borne viruses pose significant burden to public health, knowledge on the virus-host interaction at the molecular level is essential, especially in developing effective strategies to control virus transmission. In this study, we embarked on investigating the transcriptional response of long non-coding RNAs (lncRNAs) and protein-coding genes upon dengue virus serotype 1 (DENV1) infection. Besides, we also generate a comprehensive list of novel lncRNAs identified from the latest and improved genome version of *Ae. aegypti*. Similar to protein-coding genes, we discovered that the overall expression of lncRNA was significantly altered, suggesting that lncRNAs were involved in virus-host interaction. The results of this study provide basic understanding on the interaction between DENV1 and *Ae. aegypti* vector at the transcriptional level.

Introduction

Dengue virus (DENV) is a single-stranded, positive-strand RNA virus that belongs to the family of *Flaviviridae* [1]. There are four serotypes of dengue virus (DENV1-4), which are categorized based on the variation of antigens present on the viral particles [1]. Symptoms of the dengue infection include fever, headache, muscle, joint pains and rashes. In some cases, the disease develops into a life-threatening dengue hemorrhagic fever (DHF), which results in bleeding, low levels of blood platelets, and extremely low blood pressure [2] All four serotypes of DENV are transmitted by *Aedes aegypti* (*Ae. aegypti)* mosquito, which is the primary vector.

Ae. aegypti, a blood-sucking mosquito, is the principal vector responsible for replication and transmission of dengue virus. Due its pathological effects and mortality worldwide, this mosquito-borne virus is a serious threat to human health. Although virus infections of the vertebrate host are acute and often associated with disease, the maintenance of these pathogens in nature generally requires the establishment of a persistent non-lethal infection in the insect host. This is also true in the context of dengue virus infection. While dengue-infected humans show serious symptoms of illnesses, interestingly, infected *Ae. aegypti* mosquitoes are generally asymptomatic, suggesting that these vectors possess intrinsic antiviral immunity to resist or tolerate the virus infection

A number of studies have shown that, upon flavivirus infection, major transcriptomic changes occurred within *Ae. aegypti* [3–5]. Such changes involved a wide range of host genes including genes related to immunity, metabolic pathways, and trafficking. Beside protein-coding genes, non-coding RNA (ncRNA) including small and long ncRNA (lncRNA) of *Aedes* mosquitoes have been shown to undergo massive changes in transcription following virus infection [6,4].

Here, we aimed to investigate the transcriptome profiles in *Ae. aegypti* cell (Aag2) in response to DENV serotype 1 (DENV1) infection. Detailed characterization of transcriptomic changes in adult mosquitoes is convoluted by the complexity of different tissue types with unique niche and functions.

Therefore, we decided to use an immune-competent Aag2 cells in our study as it provided a more homogenous system composed of clonal cells of the same niche. To provide a global overview of transriptomic changes occuring within mosquito cells, we analyzed the expression of both protein-coding genes and lncRNAs. The latest version of *Ae. aegypti* (AaegL5) genome has been released [7], and the assembly is composed of chromosome-length scaffolds, which is more contiguous than the previous AaegL3 and AaegL4 assemblies [8]. We realized that previous lncRNA identification done in *Ae. aegypti* was based on AaegL3 assembly [6]; hence, this motivated us to perform re-annotation and identification of lncRNA using AaegL5 genome as reference. The results generated in this study provide better understanding of the mosquito-virus interaction at the molecular level, offering insights into devising strategies to inhibit viral replication within mosquitoes.

Identification of lncRNA

To perform lncRNA prediction, we used a total of 117 datasets of Illumina high-throughput sequencing derived from *Ae. aegypti* mosquito and Aag2 cell, a widely used *Ae. aegypti* derived cell line [9]. An overview of lncRNA identification pipeline can be found in Fig 1A. The pipeline developed in this study was adapted with few modifications from previous reports [6,10,11]. Briefly, each dataset (both paired-end and single-end) was individually mapped using HISAT2 [12]. The resulting alignment files were used for transcriptome assembly, and the assemblies were merged into a single unified transcriptome. Both transcriptome assembly and merging were performed using Stringtie [13]. Then, we used Gffcompare to annotate and compare the unified transcriptome assembly with reference annotation (AaegL5.1, VectorBase). We classified lncRNA transcripts based on their position relative to annotated genes derived from AaegL5.1 assembly (VectorBase). We only selected transcripts with class code “i”, “u”, and “x” that denote intronic, intergenic, and antisense to reference genes for downstream analysis.

To get confident lncRNA transcripts, we performed multiple steps of filtering transcripts having coding potential or open-reading frame (Fig 1A). The steps involved TransDecoder [14], CPAT [15] and finally BLASTX. Detailed description on the prediction analysis and parameter used can be found in Material and Methods section. Using this pipeline, we identified a set of 7,221 lncRNA transcript isoforms derived from 6,577 loci. Of these 7,221 lncRNA transcripts, 3,052 and 3,620 were intronic and intergenic respectively, while the remaining 549 transcripts were antisense to reference genes. Currently, AaegL5.1 annotation catalogs 4,155 lncRNA transcripts. Here, we provided another set of 7,221 lncRNA transcripts, making a total of 11,376 lncRNA in *Ae. aegypti*, which made up of 27.28% of the *Ae. aegypti* transcriptome.

Characterization of lncRNA

To examine whether lncRNAs identified in this study exhibit typical characteristics observed in other species [11,16–18], we analyzed features such as coding potential, sequence length, GC content and sequence conservation with closely related species. Since lncRNAs were strictly defined by their inability to code for protein, we determined coding probability of our newly identified lncRNAs and compared them with known lncRNA, 3'UTR, 5'UTR, and protein-coding mRNA. We discovered that, similar to other non-coding sequence such as known lncRNA, 3'UTR, and 5'UTR, our novel lncRNA transcripts have extremely low coding probability when compared to protein-coding mRNA (Fig 2A). Besides that, we found that both novel and known lncRNAs (provided by AaegL5.1 annotation) were shorter than protein-coding transcripts. Mean length of novel and known lncRNAs was 825.4 bp and 745.6 bp respectively, while protein-coding mRNA has an average length of 3330 bp.

Similar to previous reports [6,11], we observed that lncRNAs identified in this study had slightly lower GC content compared to protein-coding mRNAs (Fig 2C). For instances, mean GC content of novel lncRNA and mRNA was 40.1% and 46.4% respectively. Known lncRNA, on the other hand, had

143 relatively similar mean GC content to novel lncRNA (40.8%), while average GC of 5'UTR and 3'UTR
144 sequence was 43.1% and 34.6% respectively. Overall, GC content of non-coding sequence was relatively
145 lower than coding sequence.

146 To determine the *Ae. aegypti* lncRNAs that were highly conserved to other related insect species,
147 we perform BLASTN against other insect genomes including *Ae. albopictus, C, quinquifasciatus*, *An,*
148 *gambiae* and *D. melanogaster*. Similar to previous study [6], we applied the same BLASTN e-value cut-
149 off (10-50) to identify highly conserved lncRNAs. By grouping these highly conserved lncRNAs, we
150 found that most of them only shared high sequence similarity with *Ae. albopictus* (Fig 2D), suggesting
151 that they were presumably genus specific. Although *Ae. aegypti* lncRNAs shared high sequence
152 similarity with *Ae. albopictus*, compared to total lncRNAs, the fraction of conserved lncRNAs was
153 relatively lower. For example, only 11% (1,258 out of 11,376 lncRNAs) showed high sequence similarity
154 with *Ae. albopictus*.

155

156 **Identification of differentially expressed lncRNAs following DENV1 infection**

157 We next assessed whether lncRNA transcripts in Aag2 cell undergo differential expression upon
158 DENV1 infection. To answer this question, we generated paired-end RNA-seq libraries derived from
159 triplicates of Aag2 cells infected with DENV1. To avoid bias in library size which affects fold change
160 calculation, we performed differential expression analysis of protein-coding genes and lncRNAs together
161 in a single transcriptome reference using Salmon v0.9 [19] followed by edgeR [20]. A total of 2,435
162 transcripts were differentially expressed (P-value < 0.05), of which 956 of them were lncRNAs, while
163 the remaining 1,479 were protein-coding (Supplemental Table 3).

164 In general, we found that across all six samples (3 uninfected and 3 DENV1-infected), expression
165 of lncRNA was lower than protein-coding mRNAs (Fig 3A). Besides, distribution of fold change (FC)
166 of differentially expressed lncRNAs (P-value < 0.05) were much lower than that of mRNAs (Fig 3),

suggesting that unlike mRNA, lncRNA expression was less responsive to viral infection. For example, log2 FC of mRNA can extend to more than 10 fold in either direction (-10 and 10 fold), but in lncRNA, log2 FC ranged from -3 to 5. Beside, most of the differentially expressed lncRNAs (952 out of 956) transcripts had log2 FC of less than 2 in either direction.

Differential expression of protein-coding genes

We next analyzed the protein-coding transcriptome of DENV1-infected Aag2 cells. A total of 1,479 transcripts that derived from 1,272 genes were found to be differentially expressed ($|\log2\ FC| > 0$, P-value < 0.05). In subsequent analysis of protein-coding transcriptome, we only focused on transcripts having log2 FC of more or less than 1. The number of upregulated transcripts of more than one fold log2 FC was 141. Meanwhile, the number of downregulated transcripts having less than -1 log2 FC was 167.

Previous studies highlighted the importance of immune-related genes upon infection with bacteria and viruses. Thus, we examined our differentially expressed transcripts for the presence of immune-related genes. We found that key players of immunodeficiency (IMD) pathway such as NF-kappa B Relish-like transcription factor (AAEL007624-RB) and I kappa B kinase gamma-subunit (AAEL005079-RB) were downregulated [21]. Furthermore, transcript encoding defensin anti-microbial peptide (AAEL003857-RA) was also downregulated. Meanwhile, Clip-domain serine protease (AAEL002601-RA) that functions in insect innate immunity including Toll receptor activation [22] was found to be upregulated. We also found that several components of MAPK signaling pathway [23], an important pathway in immunity, were both upregulated and downregulated. For example, MAPKKK5 (AAEL008306-RG), GTP-binding protein alpha (O) subunit, gnao, (AAEL008641-RE, AAEL008641-RK), rho GTPase activator (AAEL009157-RE), eph receptor tyrosine kinase (AAEL010711-RE), and regulator of G protein signaling (AAEL010676-RB) were discovered to be upregulated. On the other

hand, GTP-binding protein alpha subunit, gna, (AAEL008630-RD) and MAPKKK (AAEL010835-RC) were downregulated.

Beside immune-related genes, we also looked at the expression of key genes in RNA-interference (RNAi) pathways. It has been reported that one of the most important immune responses against viruses known in insects is the RNAi system that make use of three major conserved small RNAs – microRNA (miRNA), small-interfering RNA (siRNA) and PIWI-interacting RNA (piRNA) [24–26]. These small RNA pathways require the binding of Argonaute/Piwi (AGO/PIWI) proteins to execute their functions [24]. We discovered that transcripts encoding AGO/PIWI proteins were not represented in our differentially expressed transcript list. To validate this observation, we performed qRT-PCR on *AGO/PIWI* genes using primers from previous studies [25]. In agreement with our RNA-seq data, qRT-PCR result showed that the expression of *AGO/PIWI* genes were largely unaffected upon DENV1 infection (Fig 4).

We next asked if there was any correlation between differentially expressed lncRNAs and protein coding genes. Co-expression analyses of lncRNAs and their adjacent protein-coding genes have been reported to provide insights into the potential roles of lncRNAs in various biological processes [27,28]. Therefore, in this study, we determined the relative distance of differentially expressed lncRNAs (956 transcripts) with the nearest protein-coding genes *in cis*, resulting in a total of 1409 lncRNA-mRNA pairs formed by a total of 1156 and 956 transcripts of mRNAs and lncRNAs respectively. We found that 68% of these pairs were located within 10kb upstream and downstream from each other. However, only 39 mRNA transcripts within this 10kb range underwent differential expression upon differential expression.

Gene ontology analysis

To investigate the gene ontology (GO) associated with the differentially expressed transcripts, we submitted both the downregulated and upregulated transcripts (|log2FC|>1) to Blast2GO for GO

214 analysis [29]. In GO analysis of upregulated transcripts, Blast2GO recognized 38, 31, 29 GO terms in
215 biological process, molecular functions and cellular components respectively. Meanwhile, in
216 downregulated transcripts, Blast2GO identified GO terms of 61 in biological process, 33 in molecular
217 function, and 24 in cellular components. We found that in biological process category, similar GO terms
218 such as signal transduction (GO:0007165), cellular process (GO:0009987), and cellular protein
219 modification process (GO:0006464) were represented in both upregulated and downregulated GO
220 analysis (Fig 5). In fact, transcripts that could be assigned to those GO terms were among the largest in
221 number.

222 Similar observation could be found in other GO categories namely molecular function and
223 cellular component. For example, in molecular function category, similar GO terms such as binding
224 (GO:0005488), protein-binding (GO:0005515), metal ion binding (GO:0046872) were represented in
225 both upregulated and downregulated analysis (Fig 5). Meanwhile in cellular component category, in both
226 cases, GO terms such as membrane (GO:0016020) and cytoplasm (GO:0005737) were among the largest
227 (Fig 5).

228 This observation led us to speculate that, upon DENV1 infection, the differentially expressed
229 transcripts, either upregulated or downregulated, generally perform relatively similar function inside the
230 cells. Therefore, we then searched GO IDs that overlapped in both upregulated and downregulated
231 transcripts based on GO categories namely biological process, molecular function and cellular
232 component. We found that most GO IDs in each category were overlapped in both upregulated and
233 downregulated data (Fig 6), suggesting that upon DENV1, genes with relatively similar functional
234 categories or working in similar signaling pathways experienced changes in expression.

235

236

237

238

239 **Discussion**

240 The field of lncRNA has become increasingly important in many areas of biology particularly
241 infectious disease, immunity, and pathogenesis [6,11,17,30]. High-throughput sequencing combined
242 with bioinformatics enable scientists to uncover comprehensive repertoire of lncRNA in many species.
243 Here, we present a comprehensive lncRNA annotation using the latest genome reference of *Ae. aegypti*
244 (AaegL5). Previous study [6] reported that a total of 3,482 intergenic lncRNA (lincRNA) was identified
245 in *Ae. aegypti*. However, the identification was performed using previous version of *Ae. aegypti* genome
246 (AaegL3), and only lncRNAs residing in intergenic region were retained.

247 Due to the recent release of *Ae. aegypti* genome (AaegL5) equipped with dramatically improved
248 gene set annotation [7], we decided to perform lncRNA identification using this latest genome reference.
249 Unlike previous study on *Ae. aegypti* lincRNA, we also annotated lncRNAs residing within the introns
250 (intronic lncRNA) and antisense to reference genes (antisense lncRNA), resulting in the identification of
251 3,052 and 549 intronic and antisense lncRNA respectively. A total of 7,221 lncRNA transcripts were
252 identified in this study, which was more than previously reported lncRNAs in *Ae. aegypti.* This
253 observation was presumably due to the high-quality assembly of AaegL5 genome and larger RNA-seq
254 datasets with high sequencing depth being used in lncRNA prediction pipeline.

255 Similar to previous reports, we discovered that lncRNAs identified in our study exhibited typical
256 characteristics of lncRNAs found in other species including vertebrates [10]. Such characteristics are
257 lower GC content, shorter in length, and low sequence conservation even among closely related species.
258 Upon DENV1 infection, we observed that both protein-coding and lncRNAs experienced overall changes
259 in expression level. However, the overall level of fold change displayed by lncRNAs was not as high as
260 protein-coding genes. This observation however, did not necessarily negate the possibility of lncRNAs
261 having potential roles in immunity. For example, RNAi-mediated knockdown of one lncRNA candidate

(lincRNA_1317) in *Ae. aegypti* resulted in higher viral replication [6]. Therefore, intensive investigation using loss or gain of function approach on lncRNAs is crucial in dissecting their functions in mosquito-virus interaction.

Transcriptome analysis of protein-coding genes showed that cellular response to DENV infection involved numerous pathways and functional gene classes, suggesting that virus infection resulted in substantial changes on cell physiology. Many viruses including DENV suppress immune signaling pathways to ensure the success of their replication. For instance, infection of *Ae. aegypti* mosquitoes with Sindbis virus (SINV) resulted in the inhibition of Toll signaling pathway [31]. Likewise, infection of mosquito cells with Semliki Forest virus (SFV) led to the reduced activation of Toll, JAK-STAT, and IMD pathways [32]. In agreement with previous studies done in *Ae. aegypti* [33], we observed that the expression level of several genes implicated in immunity was simultaneously up- and down-regulated upon DENV1 infection. For example, core components of IMD signaling pathways such as NF-kB and I kappa B kinase gamma-subunit were transcriptionally repressed.

The IMD pathway plays important role in immunity by promoting the production of anti-microbial peptides (AMP) such as defensin and cecropin [21]. Due to reduced expression of NF-kB, we discovered that the expression of defensin in DENV1-infected cell was also reduced. On the other hand, genes involved in Toll and MAPK signaling, both of which are intimately connected to JAK-STAT pathways [23], experienced up- and down-regulation upon infection with DENV1 in Aag2 cells. This simultaneous up- and down-regulation of immune-related genes that function in similar signaling pathways suggest that there is a physiological arm race in survival between host cell and DENV1.

In summary, we provided a comprehensive list of lncRNA in *Ae. aegypti* using the latest genome version. We then explored the transcriptome of both lncRNA and protein-coding genes upon infection with DENV1. Results generated in this study provide a global picture of transcriptional response in *Ae. aegypti* cell during virus infection; thereby offering a fundamental understanding of virus-host interaction.

Materials and Methods

Cell culture and virus

Ae. aegypti Aag2 cell and *Ae. albopictus* C6/36 cell (ATCC: CRL-1660) were cultured in Leibovitz's L-15 medium (Gibco, 41300039), supplemented with 10% Fetal Bovine Serum (FBS, Gibco, 10270) and 10% Tryptose Phosphate Broth (TPB) solution (Sigma, T9157). Both Aag2 and C6/36 cells were incubated at 25ºC without CO_2. Vero E6 cells (ATCC:CRL-1586) were cultured at 37ºC in Dulbecco's modified Eagles Medium (DMEM, Gibco, 11995065) supplemented with 10% FBS (Gibco, 10270), and 5% CO_2. DENV1, Hawaiian strain was propagated in C6/36 cells and titered using Vero E6 cells. Determination of DENV1 titer was done using 50% tissue culture infectious dose – cytopathic effect (TCID50-CPE) as previously described [34]. DENV1 used in this study was a gift from Dr. David Perera, University Malaysia Sarawak. Aag2 cell line was a kind gift from Dr. Ronald P. Van Rij from Radboud Institute for Molecular Life Sciences, Netherlands.

Virus infection, RNA isolation and sequencing

Aag2 cells were infected with DENV1 at multiplicity of infection (MOI) of 0.5. At 72 hours post infection (hpi), the culture medium was removed, and RNA isolation was carried out using miRNeasy Mini Kit 50 (Qiagen, 217004) according to the manufacturer's protocol. Total RNA was then subjected to next-generation sequencing. The RNA-sequencing libraries were prepared using standard Illumina protocols and sequenced using HiSeq2500 platform generating paired-end reads of 150 in size.

Verification of DENV1 infection

Total RNA of both uninfected and DENV1-infected samples were subjected to cDNA synthesis using Tetro cDNA synthesis kit (Bioline, BIO-65042). Reverse primer for DENV1 was included in the reaction

set-up. This followed with PCR and gel electrophoresis using forward and reverse primers of DENV1. Primers used for DENV1 verification were taken from previous study [35].

RNA-seq data preparation

Publicly available datasets were downloaded from NCBI Sequence Reads Archive (SRA) and ArrayExpress with accession numbers SRP041845, SRP047470, SRP046160, SRP115939, E-MTAB-1635, SRP035216, SRP065731, SRP065119, SRA048559, PRJEB13078 [4,36–41]. Adapters were removed using Trimmomatic version 0.38 [], and reads with average quality score (Phred Score) above 20 were retained for downstream analysis.

lncRNA identification

Each library (both paired-end and single-end) was individually mapped against *Ae. aegypti* genome (AaegL5) using HISAT2 version 2.1.0 [12]. The resulting alignment files were used as input for transcript assembly using Stringtie version 1.3.2 [13]. We used reference annotation file of AaegL5 (VectorBase) to guide the assembly only without discarding the assembly of novel transcripts. Here, we set the minimum assembled transcript length to be 200 bp. Then, the output gtf files were merged into a single unified transcriptome using Stringtie merge [13]. Only input transcripts of more than 1 FPKM and TPM were included in the merging. Then, we compared the assembled unified transcript to a reference annotation of AaegL5 (VectorBase) using Gffcompare (https://github.com/gpertea/gffcompare). For the purpose of lncRNA prediction, we only retained transcripts with class code “i”, “u”, and “x”. The transcripts were then subjected to coding potential prediction. We used TransDecoder [14] to identify transcripts having open-reading frame (ORF), and those having ORF were discarded. The remaining transcripts were then subjected to a coding potential assessment toll (CPAT) [15]. We set the same cut-off as previous study in *Ae. aegypti* which is less than 0.3 [6]. Transcripts having coding potential more

than 0.3 were discarded. To exclude false positive prediction, we used BLASTX against Swissprot database, and transcripts having E-value of less than 10-5 were removed.

Identification of differentially expressed transcripts

We used Salmon version 0.10.1 [19] to quantify the expression of transcripts. Read count for each transcript was subjected for differential analysis using edgeR [20] in R/Bioconductor environment.

Quantitative real-time PCR (qRT-PCR)

cDNA synthesis was done using QuantiNova Reverse Transcription Kit (Qiagen) followed by qRT-PCR with QuantiNova SYBR Green PCR kit (Qiagen) according to manufacturer's protocol. *RPS17* was used as housekeeping gene [42], and the experiment was performed using Applied Biosystems Step One PlusTM Real-Time PCR System.

Acknowledgements

We would like to thank all our collaborators and colleagues for the discussion and the work conducted in this lab. This study was funded by the USM Research University Grant (1001/PBIOLOGI/811320) and Sciencefund (305/PBIOLOGI/613238).

Figure Legend

Fig 1 lncRNA identification

(A) Overview of lncRNA identification pipeline. (B) Overview number of transcripts in *Ae. aegypti*

Fig 2 Characterization of *Ae aegypti* lncRNA

(A) Coding probability of lncRNA, 3'UTR, 5'UTR and mRNA. (B) Sequence length distribution of lncRNA and mRNA. (C) GC content. (D) Venn Diagram showing lncRNA transcripts that share high sequence similarity with other insect species (BLASTN cut-off < 10-50).

Fig 3 Differentially expressed lncRNA

(A) Overall expression of lncRNA and protein-coding mRNA. (B) Volcano plot of lncRNA. (C) Violin plot of log2 FC distribution. (D) Distribution of distance of differentially expressed lncRNA to neighboring protein-coding genes.

Fig 4 qRT-PCR of *AGO/PIWI* genes

qRT-PCR analysis of *AGO/PIWI* genes in uninfected and DENV1-infected in Aag2 cells. Error bars represent standard error of the mean across three replicates of uninfected and DENV1-infected samples.

Fig 5 GO of upregulated and downregulated transcripts

Fig 6 Shared GO IDs found in both upregulated and downregulated genes

List of Supporting Informations

S1 Table. Genomic coordinates of lncRNA candidates identified in *Ae. aegypti*

S2 Table. List of differentially expressed lncRNA

382 **S3 Table. List of differentially expressed protein-coding transcripts**

383 **S4 Table. List of GO IDs**

384 **S5 Table. List of primers used**

385

386 **References**

1. Kuhn RJ, Zhang W, Rossmann MG, Pletnev SV, Corver J, Lenches E, et al. Structure of Dengue Virus: Implications for Flavivirus Organization, Maturation, and Fusion. Cell. 2002 Mar 8;108(5):717–25.
2. Lee K-S, Ng L-C, Thein T-L, Leo Y-S, Yung C-F, Wong JGX, et al. Dengue Serotype-Specific Differences in Clinical Manifestation, Laboratory Parameters and Risk of Severe Disease in Adults, Singapore. Am J Trop Med Hyg. 2015 May 6;92(5):999–1005.
3. Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Marinotti O, James AA. Complex Modulation of the Aedes aegypti Transcriptome in Response to Dengue Virus Infection. PLOS ONE. 2012 Nov 27;7(11):e50512.
4. Etebari K, Hegde S, Saldaña MA, Widen SG, Wood TG, Asgari S, et al. Global Transcriptome Analysis of *Aedes aegypti* Mosquitoes in Response to Zika Virus Infection. Fernandez-Sesma A, editor. mSphere. 2017 Nov 22;2(6):e00456-17.
5. Shrinet J, Srivastava P, Sunil S. Transcriptome analysis of Aedes aegypti in response to mono-infections and co-infections of dengue virus-2 and chikungunya virus. Biochem Biophys Res Commun. 2017 Oct 28;492(4):617–23.
6. Etebari K, Asad S, Zhang G, Asgari S. Identification of Aedes aegypti Long Intergenic Non-coding RNAs and Their Association with Wolbachia and Dengue Virus Infection. PLoS Negl Trop Dis. 2016 Oct 19;10(10):e0005069.

7. Matthews BJ, Dudchenko O, Kingan S, Koren S, Antoshechkin I, Crawford JE, et al. Improved Aedes aegypti mosquito reference genome assembly enables biological discovery and vector control. 2017 Dec 29 [cited 2018 Sep 9]; Available from: http://biorxiv.org/lookup/doi/10.1101/240747
8. Sinkins S. Genome sequence of Aedes aegypti, a major arbovirus vector. Science. 2007 Jun 22;316(5832):1718–23.
9. Walker T, Jeffries CL, Mansfield KL, Johnson N. Mosquito cell lines: history, isolation, availability and application to assess the threat of arboviral transmission in the United Kingdom. Parasit Vectors. 2014;7(1):382.
10. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. Cell Rep. 2015 May;11(7):1110–22.
11. Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R, et al. Systematic Identification and Characterization of Long Non-Coding RNAs in the Silkworm, Bombyx mori. PLOS ONE. 2016 Jan 15;11(1):e0147147.
12. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr;12(4):357–60.
13. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015 Mar;33(3):290–5.
14. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat Protoc [Internet]. 2013 Aug [cited 2018 Jul 16];8(8). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3875132/

15. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013 Apr;41(6):e74.
16. Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, et al. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol. 2010 Jul 12;11:R72.
17. Clark MB, Mattick JS. Long noncoding RNAs in cell biology. Semin Cell Dev Biol. 2011 Jun;22(4):366–76.
18. Jenkins AM, Waterhouse RM, Muskavitch MA. Long non-coding RNA discovery across the genus anopheles reveals conserved secondary structures within and beyond the Gambiae complex. BMC Genomics. 2015 Apr 23;16:337.
19. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat Methods. 2017 Apr;14(4):417–9.
20. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.
21. Myllymaki H, Valanne S, Ramet M. The Drosophila Imd Signaling Pathway. J Immunol. 2014 Apr 15;192(8):3455–62.
22. Kanost MR, Jiang H. Clip-domain serine proteases as immune factors in insect hemolymph. Curr Opin Insect Sci. 2015 Oct 1;11:47–55.
23. Chen RE, Thorner J. Function and Regulation in MAPK Signaling Pathways. Biochim Biophys Acta. 2007 Aug;1773(8):1311–40.
24. Azlan A, Dzaki N, Azzam G. Argonaute: The executor of small RNA function. J Genet Genomics. 2016 Aug 20;43(8):481–94.
25. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015 Jul 27;43(13):6545–56.

26. Miesen P, Ivens A, Buck AH, Rij RP van. Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. PLoS Negl Trop Dis. 2016 Feb 25;10(2):e0004452.

27. Li L, Cong Y, Gao X, Wang Y, Lin P. Differential expression profiles of long non-coding RNAs as potential biomarkers for the early diagnosis of acute myocardial infarction. Oncotarget [Internet]. 2017 Oct 24 [cited 2018 Sep 9];8(51). Available from: http://www.oncotarget.com/fulltext/20101

28. Silva JP, Booven D van. Analysis of diet-induced differential methylation, expression, and interactions of lncRNA and protein-coding genes in mouse liver. Sci Rep. 2018 Aug 1;8(1):11537.

29. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005 Sep 15;21(18):3674–6.

30. Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu J-L, et al. Identification and Properties of 1,119 Candidate LincRNA Loci in the Drosophila melanogaster Genome. Genome Biol Evol. 2012;4(4):427–42.

31. Sanders HR, Foy BD, Evans AM, Ross LS, Beaty BJ, Olson KE, et al. Sindbis virus induces transport processes and alters expression of innate immunity pathway genes in the midgut of the disease vector, Aedes aegypti. Insect Biochem Mol Biol. 2005 Nov;35(11):1293–307.

32. Fragkoudis R, Chi Y, Siu RWC, Barry G, Attarzadeh-Yazdi G, Merits A, et al. Semliki Forest virus strongly reduces mosquito host defence signaling. Insect Mol Biol. 2008 Dec;17(6):647–56.

33. Sim S, Dimopoulos G. Dengue Virus Inhibits Immune Responses in Aedes aegypti Cells. PLOS ONE. 2010 May 18;5(5):e10678.

34. Li J, Hu D, Ding X, Chen Y, Pan Y, Qiu L, et al. Enzyme-Linked Immunosorbent Assay-Format Tissue Culture Infectious Dose-50 Test for Titrating Dengue Virus. PLOS ONE. 2011 Jul 25;6(7):e22553.

35. Johnson BW, Russell BJ, Lanciotti RS. Serotype-Specific Detection of Dengue Viruses in a Fourplex Real-Time Reverse Transcriptase PCR Assay. J Clin Microbiol. 2005 Oct;43(10):4977–83.

36. Li Y, Piermarini PM, Esquivel CJ, Drumm HE, Schilkey FD, Hansen IA. RNA-Seq Comparison of Larval and Adult Malpighian Tubules of the Yellow Fever Mosquito Aedes aegypti Reveals Life Stage-Specific Changes in Renal Function. Front Physiol. 2017;8:283.

37. Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, et al. SEX DETERMINATION. A male-determining factor in the mosquito Aedes aegypti. Science. 2015 Jun 12;348(6240):1268–70.

38. McBride CS, Baier F, Omondi AB, Spitzer SA, Lutomiah J, Sang R, et al. Evolution of mosquito preference for humans linked to an odorant receptor. Nature. 2014 Nov 13;515(7526):222–7.

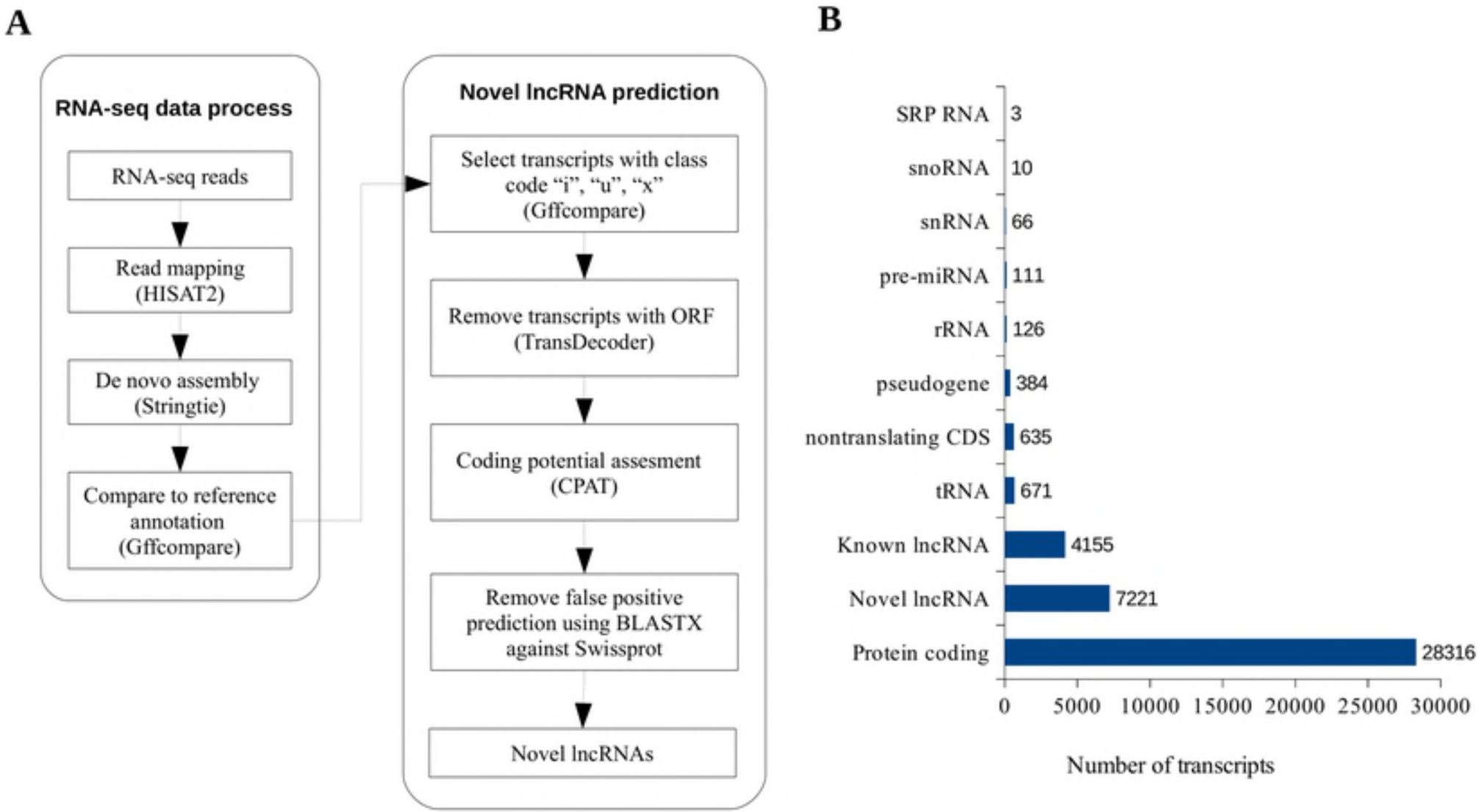
39. Canton PE, Cancino-Rodezno A, Gill SS, Soberón M, Bravo A. Transcriptional cellular responses in midgut tissue of Aedes aegypti larvae following intoxication with Cry11Aa toxin from Bacillus thuringiensis. BMC Genomics. 2015 Dec 9;16:1042.

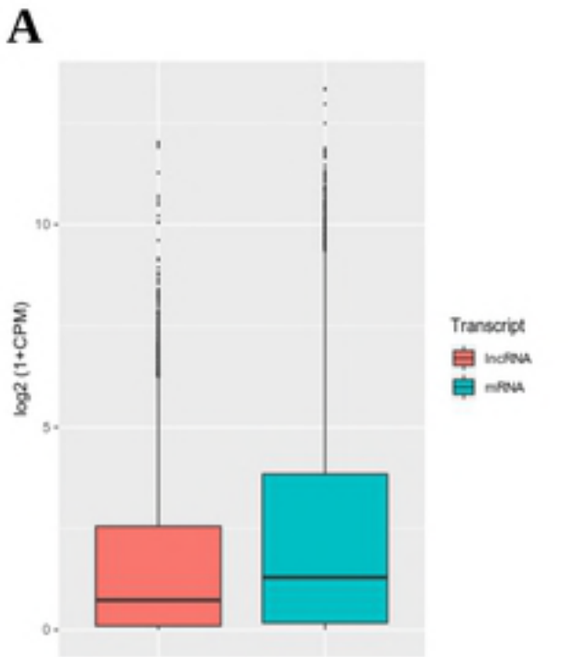
40. David J-P, Faucon F, Chandor-Proust A, Poupardin R, Riaz MA, Bonin A, et al. Comparative analysis of response to selection with three insecticides in the dengue mosquito Aedes aegyptiusing mRNA sequencing. BMC Genomics. 2014 Mar 5;15(1):174.

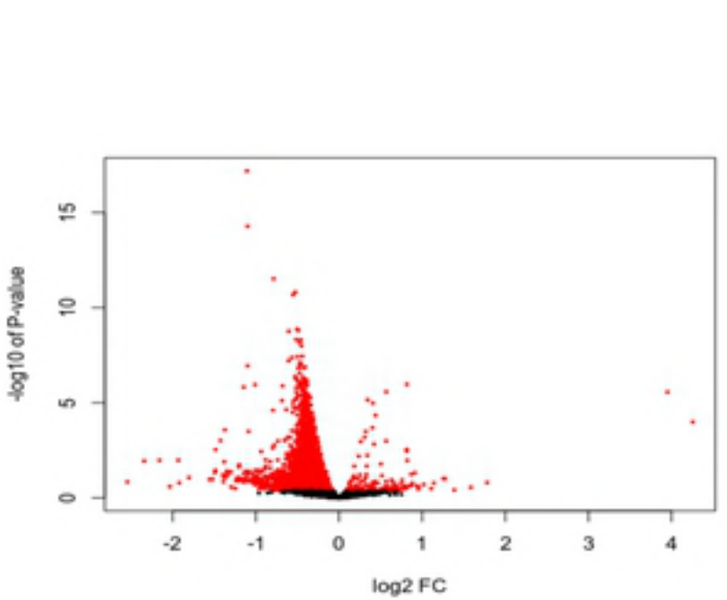
41. Maringer K, Yousuf A, Heesom KJ, Fan J, Lee D, Fernandez-Sesma A, et al. Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in Aedes aegypti. BMC Genomics [Internet]. 2017 Jan 19 [cited 2018 Sep 9];18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5248466/

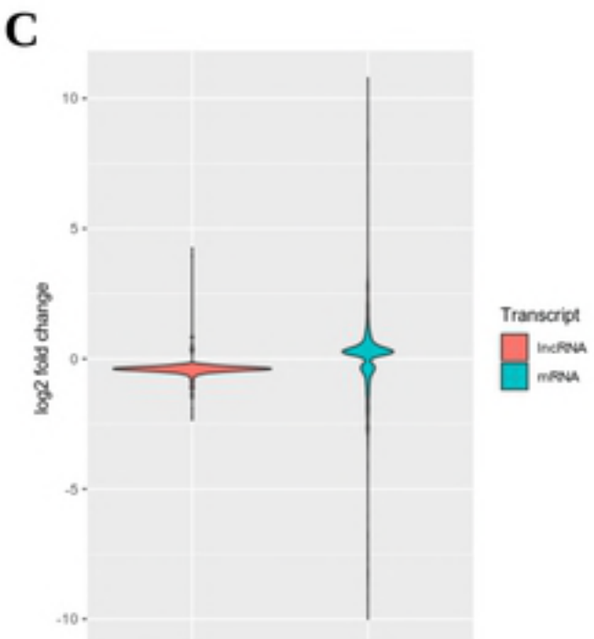
42. Dzaki N, Ramli KN, Azlan A, Ishak IH, Azzam G. Evaluation of reference genes at different developmental stages for quantitative real-time PCR in Aedes aegypti. Sci Rep. 2017 16;7:43618.

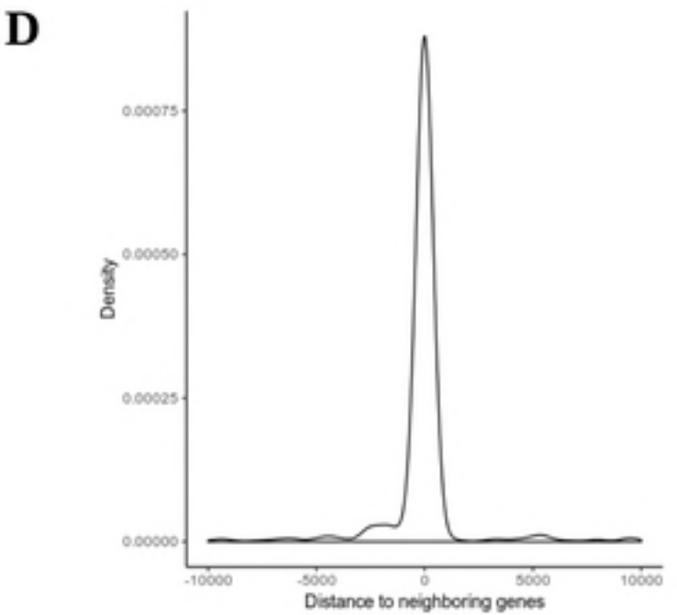
387
388


A
RNA-seq data process
RNA-seq reads
Read mapping (HISAT2)
De novo assembly (Stringtie)
Compare to reference annotation (Gffcompare)
Novel lncRNA prediction
Select transcripts with class code “i”, “u”, “x” (Gffcompare)
Remove transcripts with ORF (TransDecoder)
Coding potential assesment (CPAT)
Remove false positive prediction using BLASTX against Swissprot
Novel lncRNAs
B
SRP RNA 3
snoRNA 10
snRNA 66
pre-miRNA 111
rRNA 126
pseudogene 384
nontranslating CDS 635
tRNA 671
Known lncRNA 4155
Novel lncRNA 7221
Protein coding 28316
0 5000 10000 15000 20000 25000 30000
Number of transcripts


A
log2 (1+CPM)
10
5
0
Transcript
lncRNA
mRNA


B
-log10 of P-value
15
10
5
0
-2
-1
0
1
2
3
4
log2 FC


C
log2 fold change
10
5
0
-5
-10
Transcript
lncRNA
mRNA


D
Density
0.00075
0.00050
0.00025
0.00000
-10000
-5000
0
5000
10000
Distance to neighboring genes