Epigenetics & Chromatin

RESEARCH

Open Access

Functional characterization of transcriptional enhancers in an *Anopheles* genetic locus controlling natural resistance to the malaria parasite, *Plasmodium falciparum*

Natalia Marta Zmarlak-Feher[1†], Kathryn S. Taquet[2†], Renée Zakhia[1,3], Adrien Pain[1,4], Emma Brito-Fravallo[1,3], Cameron E. Anderson[2], Kenneth D. Vernick[1], Christian Mitri[1,3*] and Michelle M. Riehle[2*]

Abstract

Background *Anopheles* mosquitoes and the malaria parasites they transmit remain a significant global health problem. Most genomic and functional genomic studies of mosquitoes have focused on the protein-coding genome, and comparatively little is known about the importance of noncoding transcriptional enhancers in controlling their gene expression and phenotypic variation. Here we evaluate nine enhancers previously identified in a STARR-seq screen and present in a genetic locus that was identified as a major influence on susceptibility to malaria infection in wild *Anopheles coluzzii* mosquitoes.

Result We developed an analytical pipeline to filter nine enhancers in the malaria susceptibility locus on chromosome 2L. First, ATAC-seq revealed that only three of the nine enhancers were located in open chromatin and thus likely to be active in somatic cells. Next, we cloned these three enhancers from malaria-susceptible and resistant mosquitoes and measured their enhancer activity by luciferase reporter assays. Only two of the three open-chromatin enhancers displayed significantly different enhancer activity between resistant and susceptible alleles. Finally, alleles of just one of these enhancers, ENH_2L-03, contained nucleotide variants which also segregated in wild mosquitoes, and ENH_2L-03 was prioritized for further study. A noncoding RNA was detected within ENH_2L-03, consistent with an enhancer RNA (eRNA), which we depleted in mosquitoes using RNAi in order to silence the enhancer activity. Transcriptional profiling of ENH_2L-03-silenced mosquitoes revealed 15 differentially expressed genes, which share a transcription factor binding motif suggestive of coordinate regulation. However, silencing ENH_2L-03 did not influence infection levels of either human or rodent malaria parasites.

Conclusion Despite the absence of an ENH_2L-03 effect on infection outcome, multiple enhancers can cooperate to influence a phenotype, and further examination of this enhancer is warranted. Overall, we provide a pipeline for the in vivo functional study of transcriptional enhancers in *Anopheles,* towards understanding how enhancer function may control important vector phenotypes.

Keywords *Anopheles*, Enhancer, Epigenetics, Non-coding DNA

†Natalia Marta Zmarlak-Feher and Kathryn S. Taquet are co-first authors.

*Correspondence:
Christian Mitri
christian.mitri@pasteur.fr
Michelle M. Riehle
mriehle@mcw.edu
Full list of author information is available at the end of the article

Introduction

Mosquitoes are the vectors of multiple and widespread emerging and re-emerging infectious diseases, particularly malaria and a range of arboviruses. In nature, individual mosquitoes vary in their vector capacity due to both extrinsic and intrinsic factors, including genetic variation [1, 2]. Previous work, using linkage mapping approaches, mapped malaria susceptibility in wild *A. coluzzii* and *A. gambiae* mosquito pedigrees to an interval on chromosome 2L [2, 3]. Functional dissection of this region has identified factors important in mosquito innate immunity and the immune response to *Plasmodium* infection, such as the APL1 gene family [2, 4, 5], but identifying the specific genetic basis underlying natural variation in susceptibility to *Plasmodium falciparum* infection remains elusive. Extensive community effort has been dedicated to the annotation and functional characterization of *Anopheles gambiae/coluzzii* genes with hypothesized and experimentally verified roles in innate immunity, pathogen recognition, and insecticide resistance [6–8]. However, comparatively little is known about the regulatory elements that occupy the non-coding portion of the *Anopheles* genome, despite the preponderance of evidence across many systems pointing to the importance of genetic variation in the non-coding genome as underlying most phenotypic variation, genome-wide association signals and complex phenotypes [9–13]. The most important class of non-coding DNA elements, transcriptional enhancers, are responsible for the regulation of gene transcription above basal levels in eukaryotes. Genetic variation in enhancer sequences can impact transcription factor binding, enhancer activity, the regulation of downstream target genes and ultimately phenotype [9–13] and may underlie phenotypes relevant for pathogen transmission and vector control, such as insecticide resistance, parasite susceptibility, mosquito behavior, or adaptation to ecological conditions.

Current vector control methods rely heavily on chemical and mechanical tools through the use of insecticide-treated bed nets, indoor residual spraying, and larval source management (habitat modification and/or larviciding) [14–18]. However, strong selective pressure has resulted in increased insecticide resistance and diminished control efficacy. Research investment into genetically modified mosquitoes as a population replacement or population suppression strategy has expanded and would be useful to add to the malaria control toolbox [19–21]. However, many of these strategies are not informed by natural modes of malaria resistance as we do not yet understand the mechanisms of natural resistance. Our work aims to identify and characterize genetic variation underlying natural resistance, with current focus on the non-coding regulatory genome, and in doing so, provide means to evolutionarily inform future control measures. Identifying an enhancer candidate that functionally controls *Plasmodium* infection outcomes, particularly a candidate with an agonistic function for infection, could be targeted by the CRISPR gene drive system to spread a genetically modified population with an altered specific-enhancer function [20–22].

Despite once being labeled as junk DNA, the functional importance of the non-coding genome is now widely acknowledged [23] and has taken center stage in the dissection of intricate regulatory networks, which are at the root of complex phenotypes. However, there remain inherent challenges in querying the noncoding genome for function including the lack of an amino acid-like code and low levels of sequence conservation across species [24, 25]. Attempts to identify regulatory elements in mosquitoes have employed -omics scale technologies which either use massively parallel reporter assays (MPRAs) [26] or capitalize on properties of open chromatin to infer presence of regulatory elements, reviewed in Farley et al. [23]. In addition to patterns of open chromatin, epigenetic signatures written through histone modifications and detected by ChIP-seq have been compared in blood-fed and *Plasmodium* challenged *Anopheles* [27].

Even when enhancers and epigenetic signatures have been cataloged and identified, it is not possible to predict without empirical studies if and how genetic variation within an enhancer sequence will affect the underlying activity and ultimately the regulation of downstream target genes [11, 28]. While work in *Drosophila melanogaster*, a related model arthropod, is more advanced [29, 30], knowledge and functional characterization of the non-coding genome in mosquitoes remains in its infancy. Identification and characterization of genome regulation in this insect of public health importance could foster the development of vector control tools rooted in the natural response of the mosquito to pathogen infection.

Here we explore the enhancer regulatory neighborhood flanking the genetic marker most significantly linked to malaria susceptibility in wild *A. gambiae* and *A. coluzzii* mosquito pedigrees from East and West Africa [2, 3]. We previously generated a genome-wide catalog of *A. coluzzii* enhancers using a STARR-seq MPRA functional screen [26]. From this genome-wide catalog, we now find that nine enhancers are located in this malaria susceptibility locus of interest. Using complementary omics-based approaches alongside computational predictions and mosquito whole genome sequence data, we prioritize one of the enhancers in this locus, ENH_2L-03, for functional testing. First, to develop a tool to experimentally perturb enhancer function, we identified and

molecularly characterized a candidate enhancer RNA (eRNA) colinear with the enhancer, and depleted it using RNAi. Transcriptome profiling then identified genes whose expression was modulated by silencing the candidate eRNA. Finally, mosquitoes were challenged with both rodent malaria, *Plasmodium berghei* and human malaria, *Plasmodium falciparum* after candidate eRNA silencing to functionally assess the role of this non-coding transcript in the outcome of infection.

Methods

Identify regions of open chromatin using ATAC-seq

Preparation of nuclei

4a-3A hemocyte like cells [31] at 80% confluence were washed in 1X PBS at room temperature. Cells were then harvested and counted. For each of the 2 replicate reactions 500,000 cells were spun down at 500 g for 10 min at 4 °C. The resulting cell pellet was washed with 50 µl of cold 1X PBS and spun at 500 g for 10 min at 4 °C. The resulting cell pellet was resuspended in 50 µl of cold lysis buffer (10 mM Tris–HCL pH7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1% IGEPAL) and again spun at 500 g for 10 min at 4 °C.

Transposition

The resulting pellet was subjected to transposition using TDE1 transposase. Nuclei were resuspended in the transposition reaction mix containing 1X TDE buffer and 2.5 µl of TDE1 transposase and incubated at 37 °C for 30 min with gentle mixing. Reactions were purified using a Qiagen MinElute PCR Purification Kit and eluted in 11.5 µl Elution Buffer. Purified DNA was stored at −20 °C and sent to the University of Minnesota Genomics Center (UMGC) for sequencing. The UMGC performed the PCR amplification including indexing, library preparation and sequencing. Libraries were sequenced to 8 million 75bp paired-end reads on an Illumina MiSeq.

ATAC-seq analysis

Raw read quality was assessed using FastQC version 0.11.9 [32] and MultiQC version 1.9 [33]. Cutadapt version 2.10 [34] was used to clean the reads and remove adapter sequences with the following parameters: -m 30, -q 30, 30. BWA mem version 0.7.17-r1188 was used with the default parameters to align sequence reads against the reference genome of *Anopheles gambiae* str. PEST version AgamP4 (VectorBase release 53). Genrich version 0.6.1 [35] was used for peak calling with the parameters -j for ATAC-seq mode and -r for remove PCR duplicates.

Methods to assess ATAC-seq data quality

Peak annotation

Annotation was performed in R using the CHIPseeker (version 1.5.1) and GenomicFeatures (version 1.38.2) packages. The GFF annotation was loaded using makeTxDbFromGFF and peak annotation was done with AnnotatePeak, setting tssRegion set to −2000, 2000). The plot was generated with ggplot2 (3.3.3) and cowplot (version 1.1.1). TSS positions were extracted using the promoters function (from GenomicFeatures).

TSS enrichment plot

Deeptools, a suite of python tools, was used to produce the TSS enrichment plot. First, BigWig files were generated for the two libraries using bamCoverage (option normalizeUsing CPM) and the signal matrix around the TSS was generated using computeMatrix reference-point using the following options: –referencePoint TSS -b 2000 -a 2000 and -R with the bed file containing the TSS position previously extracted.

The PEST reference genome [36] was used as it represents the most mature genome assembly and annotation of *Anopheles gambiae* or *coluzzii* and allows direct comparison with other published work mapped to the same default genome reference. While the PEST genome has been historically referred to as pure *A. gambiae*, it is a hybrid genome of both *A. gambiae* and *A. coluzzii,* and there is no current pure A. coluzzii reference of similar assembly quality [37]. To assess the quality of ATAC-seq data prior to focusing on the genetic locus of interest, several metrics were assessed including correlation across replicate samples, mapping rates, insert sizes, enrichment at ATG/TSS and the genomic location of detected peaks. The two biological replicate samples were correlated at 0.989. Mapping rates were 99.9% for each experimental replicate and duplicate rates were 1.7% and 3.2%, respectively. The total number of ATAC-seq peaks detected was 15,948 and the majority of these were located in promoter regions or introns with patterns consistent across replicates (Additional File 1_Figure S1). The majority of ATAC-seq peaks were within 3kb of the TSS, with a vast majority of these within 1kb and this pattern was also significant across replicates. A list of genome-wide ATAC-seq peaks is available as a bed compatible file (Additional File 2_Table S1). ATAC-seq raw data are available at PRJEB79891 https://www.ebi.ac.uk/ena/browser/view/PRJEB79891.

Measurement of enhancer activity using luciferase assays

Cloning of candidate enhancer fragments

Candidate enhancer regions were PCR amplified from genomic DNA isolated from malaria susceptible and

resistant homozygote individuals from an original mapping pedigree, generated as described [2, 3]. PCR reactions consisted of 10ng template DNA, 1X Phusion Ultramix, and 0.5 µM primer. Cycling conditions were an initial denaturation at 98 °C for 30 s, 30 cycles of 98 °C for 10 s, 55–62 °C annealing temp for 30 s and a 45 s extension at 72 °C followed by a final 10 min extension at 72 °C. Resulting PCR fragments were either restriction enzyme cloned into a Firefly luciferase reporter vector pGL-Gateway-DSCP (AddGene, vector ID# 71506) or gateway cloned into pGL-Gateway-DSCP. Ligation products were transformed into OneShot OmniMax 2 T1 Phage-Resistant Cells (Invitrogen), grown up overnight, plasmid purified and sequenced [26]. Amplification primers for candidate enhancers can be found in Additional File 1_TableS2. These primers are designed to be inclusive of published enhancers [26], and as such the amplicons are slightly larger than the STARR-seq published enhancer. Sequences for cloned alleles can be found in Additional File 1.

Lipid based transfection and luciferase assays

A. coluzzii SUA 4.0 hemocyte-like cells [31] were seeded at 2.5×10^4cells/well in 65 µL in a 96 well plate. Cells were agitated on a MixMate (Eppendorf) for 30 s at 300 rpm for even distribution. Following a 24 h incubation period, lipid-based transfections were performed using Lipofectamine 3000 (Invitrogen). Cells were transfected with two plasmids (1) the pGL-Gateway-DSCP (described above) carrying a single amplified fragment of the candidate enhancer upstream of a firefly luciferase gene and (2) a renilla control vector pRL-ubi-63E (AddGene, #74280). Plasmids were transfected at a ratio of 1:5 (renilla:firefly). Plates were again agitated for 30 s at 300 rpm on a MixMate (Eppendorf) and incubated for 24 h at 27 °C. The Dual-Glo Luciferase Assay System (Promega) was used for luciferase assays, according to supplier instructions. Measurements were recorded on the GloMax Discover (Promega) at 25 °C. All test plates contained cells transfected with control plasmid constructs; a negative control fragment, which was a size-matched fragment within intron 1 of AGAP007058, and a highly active positive control enhancer peak nearby AGAP008980 [38]. All samples were run in sixfold technical replication within a single plate and across at least two independent plates for at least two biological replicates. Firefly luciferase measurements (relative light units, RLUs) were normalized to the renilla measurements from the same well. These measurements were then expressed relative to the firefly/renilla mean for the negative control on the same plate.

Computational analysis of wild mosquito genetic variation

Ag1000G phase 1 AR3 variant data was downloaded from MalariaGEN [39]. The variable positions within each enhancer of interest were selected. The following groups of individuals were analyzed: 2La/2La mosquitoes from all countries combined, consisting of 246 mosquitoes (492 alleles) that were a mix of *A. gambiae* (from Guinea, Burkina Faso, Cameroon, Uganda, Kenya, and Gabon), and *A. coluzzii* (from Burkina Faso and Angola), and admixed (from Guinea-Bissau); 2La/2La *A. gambiae* mosquitoes from Cameroon only, consisting of 48 mosquitoes (96 alleles); and 2La/2La *A. gambiae* mosquitoes from Burkina Faso only, consisting of 66 mosquitoes (132 alleles). For each group, each position was confirmed to have a missing genotype frequency < 0.05, and for all positions meeting this criteria, maximum allele frequency was calculated using vcftools [40]. Because the positions from the variant data include bi-, tri-, and quadr- allelic positions, commonly variable positions were defined as positions in which the maximum allele frequency was < 0.95 for at least one group of mosquitoes (all $n = 246$; Cameroon only $n = 48$ or Burkina Faso only $n = 66$). This cut-off means the minor allele (or alleles in the case of tri and quadr- allelic positions) frequency was > 0.05. The analysis of genetic variation was performed on enhancers as defined by STARR-seq analysis. Fragment lengths used in luciferase assays are necessarily slightly longer than STARR-seq defined enhancers as primers are designed to flank the entire enhancer.

Comparison of STARR-seq and ATAC-seq peaks in the genomic locus with other published databases

Published open chromatin and epigenetic marker information was extracted from previously published work, and peaks were compared. Peaks within ± 1500bp of a given STARR-seq peak were considered overlapping with that STARR-seq peak. Each of the nine STARR-seq peaks within the genomic locus were compared to a FAIRE-seq study conducted in the 4a-3B hemocyte cell line [41], a ChIP-seq experiment examining blood-fed and *Plasmodium*-infected mosquitoes [27], as well as a combined ATAC-seq data from midgut and salivary gland tissues of *Plasmodium*-infected mosquitoes [42]. While the focus here is on the genomic region controlling malaria susceptibility, the genome wide set of ATAC-seq peaks is available in Additional File 2_Table S1.

In vivo detection of EN2L-03 candidate eRNA and functional assays

Mosquitoes

The Ngousso strain (*An. coluzzii*) was colonized in 2006 from mosquitoes collected in Yaoundé, Cameroon. The

mosquito colony was reared under standard insectary conditions (26 °C, 12:12 L:D, and 70% relative humidity) by the Centre for the Production and Infection of *Anopheles* (CEPIA) of the Institute Pasteur in Paris, France.

Candidate eRNA detection by conventional PCR

Total RNA was extracted from a pool of 10 female mosquitoes with TRI Reagent (Molecular Research Center, Inc.) and 1 µg of total RNA was reverse transcribed to cDNA using M-MLV reverse transcriptase (Invitrogen) and random hexamers (Invitrogen) as previously described [7]. The primers used for detection of eRNA expression from enhancer 3 (ENH_2L-03) span the genomic region targeted by ds1 and ds2 fragments (Fig S2) and consist of a forward primer in the ds1 region and a reverse primer in the ds2 region (Primer sequences in Additional File 1_Table S2). For the PCR amplification, Ex Taq (Takara) was used with cycling condition: [95 °C for 3 min, 35 cycles of (95 °C for 30 s, 59 °C for 20 s, 68 °C for 30 s), 68 °C for 5 min].

Synthesis of double-stranded RNA and microinjection

ENH_2LH-03 eRNA silencing

To maximize the efficiency of eRNA silencing, three double-stranded (ds) RNAs (ds1, ds2 and ds3) spanning a total of 248 bp were designed and synthesized for targeting ENH_2L-03 candidate eRNA. A dsRNA directed against the irrelevant GFP target (dsGFP) was synthesized as a control. The dsRNAs were prepared using the MEGAscript™ RNAi Kit (Invitrogen, Thermo Fisher Scientific) as described [4]. Primers bearing the T7 RNA polymerase recognition sequence were designed to amplify 87, 85 and 76 base pairs for ds1, ds2 and ds3, respectively. All primer sequences are listed in Additional File 1_Table S2. For silencing of eRNA expression, three-day-old cold-anaesthetized female mosquitoes were injected into the thorax using a NanoJect II microinjector (Drummond) either with 1.5 µg of an equal quantity of ds1, ds2 and ds3 (dsENH_2L-03$_{Mix}$; 500 ng of each fragment per mosquito) or with 1.5 µg of dsGFP for the control group. Following injections, mosquitoes were maintained on 10% sucrose solution at 26 °C for 2 days before challenge with *P. falciparum*; or at 21 °C for 2 days before infection with *P. berghei*.

AGAP007051RNA silencing

Double-stranded RNA was designed and synthesized using T7-primers to target 534 bp of the AGAP007051 transcript. dsGFP was used as a control. Primer sequences are listed in Additional File 1_Table S2. For silencing 7051 mRNA, mosquitoes were treated as above with 850 ng of either ds7051 or the dsGFP control. Following injection, mosquitoes were maintained on 10% sucrose solution at 21 °C for 4 days before being processed for RNA silencing verification.

Test for silencing efficiency and eRNA/gene expression

For each condition, two days post dsRNA injection, 1 µg of total RNA extracted with TRI Reagent (Molecular Research Center, Inc.) was reverse transcribed to cDNA using M-MLV reverse transcriptase and random hexamers (Invitrogen) as previously described [7]. The efficiency of gene silencing in mosquitoes was monitored by quantitative real-time PCR (RT-qPCR) using SYBR Green Supermix (KAPA SYBR FAST ABI, Kapa Biosystems) and the CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories). Primers are listed in Additional File 1_Table S2. The gene for ribosomal protein S7 (rpS7) was used as an internal control. The quantification of candidate eRNA and gene transcript levels was done relative to rpS7. Analysis of transcript abundance relative to rpS7 was determined according to the $2^{-\Delta\Delta Ct}$ method [43]. PCR cycling conditions were: 95 °C for 10 min, 40 cycles of (95 °C for 15 s, 60 °C for 1 min).

Confirmation of a candidate eRNA molecule antisense to AGAP007051

Specifically primed reverse transcription was carried out to detect the candidate eRNA antisense to the direction of the AGAP007051 intron. The MMLV-RNase H point mutant (Promega), which functions between 40 and 55 °C, was used to enhance the binding specificity of the primer used for reverse transcription. 1 ug of total RNA isolated from dsGFP and dsENH_2L-03, dsGFP and ds7051 treated mosquitoes with 2 pmol of specific primers and 20 pmol of dNTP in a total volume of 25 µL were reverse transcribed at 54 °C, for 1 h, followed by 15 min at 70 °C for enzyme inactivation. The primer used was EN03_Int_F (Additional File 1_Table S2) for specifically detecting ENH_2L-03 candidate eRNA.

For PCR detection following the transcript-specific cDNA synthesis, Taq polymerase (Takara) was used with the following program: 3 min at 94 °C, 40 cycles of (20 s at 94 °C, 20 s at 60 °C and 25 s at 68 °C), 68 °C 10 min. For the primers, EN03_Int_F and ds3_R (Additional File 1_Table S2) were used to detect ENH_2L-03 transcript.

Plasmodium falciparum gametocyte culture and mosquito infection

Plasmodium falciparum isolate NF54 was cultured using an automated tipper-table implemented in the CEPIA mosquito infection core facility of the Institute Pasteur, as previously described [4, 6, 44]. Briefly, 14 days after initiating the subculture, gametocyte maturity was tested by exflagellation of microgametes, and parasitemia and numbers of mature male and female gametocytes were

counted on Giemsa-stained slides. For experimental infection of mosquitoes, 10 ml of medium containing mature gametocytes were centrifuged at 2000 rpm, and the cell pellet was resuspended in an equal volume of normal type AB human serum. The infected erythrocytes were added to fresh erythrocytes in AB human serum, mixed gently, and transferred to a membrane feeder warmed to 37 °C. At 2 days post-dsRNA injection (dsGFP or dsENH_2L-03$_{Mix}$), Ngousso colony mosquitoes were allowed to feed for 20 min, unfed females were discarded, and only fully engorged females were used for further analysis. Blood-fed mosquitoes were maintained at 26 °C and 70% relative humidity on 10% sucrose solution.

Rodent malaria infection

One to three-day old Ngousso colony mosquitoes were injected with dsGFP or dsENH_2L-03$_{Mix}$ and on the second day post-injection were fed on mice infected with *P. berghei* strain PbGFPCON [45], which constitutively expresses green fluorescent protein (GFP). The mosquitoes were allowed to feed for 15 min. Unfed females were discarded, and only fully engorged females were maintained at 21 °C (necessary for *P. berghei* development) and 70% relative humidity on 10% sucrose.

Analysis of phenotypes

Effects on infection prevalence and intensity were calculated from biological replicates of ≥ 30 dissected mosquitoes each with three independent replicates performed for each species of *Plasmodium*. Mosquito midguts were dissected 8 days post-infection. For *P. falciparum*, midguts were stained with 0.4% mercury dibromofluorescein (Sigma) and the number of oocysts was counted by light microscopy. For *P. berghei*, oocysts were counted by fluorescence microscopy. Measured phenotypes included oocyst infection prevalence (the proportion of mosquitoes carrying ≥ 1 oocyst among the total number of dissected mosquitoes), and oocyst intensity (the number of oocysts counted in mosquitoes with ≥ 1 oocyst). Differences in infection prevalence were statistically analyzed using the χ^2 test, and analysis of oocyst intensity used the Wilcoxon signed rank non-parametric test. Statistical differences in prevalence and intensity were first tested separately for each independent replicate as described above, and p-values were empirically determined using 10^5 Monte-Carlo permutations.

Transcriptomic analysis to identify genes regulated by EN2L-03 candidate eRNA

Sample preparation

RNAseq was carried out to compare transcriptome profiles in mosquitoes depleted for the ENH2L-03 candidate eRNA as compared to dsGFP controls. Three-day-old cold-anaesthetized mosquito Ngousso females were injected using the NanoJect II (Drummond) either with 1.5 µg of an equal quantity of ds1, ds2 and ds3 (500 ng each per mosquito) or with 1.5 µg of dsGFP for the control group as described above. Two days post-injection, for each condition, total RNA was extracted from a pool of 10 mosquitoes with TRI Reagent (Molecular Research Center, Inc.). Three independent experiments were performed, and the extracted total RNA from each condition and each experiment were submitted to transcriptome analysis.

RNA-seq analysis

RNA-seq was performed by Fasteris (Genesupport Lifescience, Plan-les-Ouates, Switzerland) using the PolyA mRNA-seq protocol. Each library was sequenced to a depth of ~ 30 million reads with 2*150 bp paired end reads on a NovaSeq 6000. Raw read quality was assessed using FastQC version 0.11.9 [32] and MultiQC version 1.12 [33]. Cutadapt version 2.10 [34] was used to clean the reads and remove adapters sequences with the following parameters: -O 6, -m 30 (discards reads shorter than 30bp), -q 30 –-trim-n (removes N bases from the read ends) and –-max-n 1 (removes reads with too many Ns). STAR version 2.7.9a [46] was used for alignment against the reference genome of *Anopheles gambiae* str. PEST version AgamP4 (VectorBase release 66). First the genome was indexed with the GFF annotation file from VectorBase release 66 and the following parameters: –sjdbGTFtagExonParentTranscript Parent and –genomeSAindexNBases 8 and the mapping was done with the following parameters: –seedSearchStartLmax 20 and –alignIntronMax 25000. Genes were counted using featureCounts version 2.0.0 [47] with the annotation from VectorBase release 66 and the parameters –t exon –g gene_id -p -B. Counts data were analyzed using R version 4.3.2 [48], the SARTools package version 1.8.1 [49] and template script for DESeq2 [50] with default settings, including a 5% false discovery rate, except for the blocking factor which reflects replicate number. The median of ratios method was used for normalization. The template script for analysis is available https://github.com/PF2-pasteur-fr/SARTools/blob/master/template_script_DESeq2.r. Genes with adjusted p-values below 0.05 were considered differentially expressed. All RNA-seq data is available at E-MTAB-14478 at Biostudies https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-14478.

MEME for discovery of predicted transcription factor binding sites

MEME version 5.5.5 [51] was used to perform motif discovery to find significantly enriched sequence motifs

near the 15 differentially expressed genes (DEG) identified by RNAseq. In particular, shared predicted binding sites among the 15 DEGs for transcription factors (TF) would suggest a common regulatory pathway associated with the activity of ENH2L-03 candidate eRNA upon gene expression. To do so, 500 nucleotides upstream of the TSS for all *A. gambiae* genes from the Agam4 genome version 68 assembly were extracted and upstream sequences from differentially expressed genes were used as input, while the upstream sequences from all *A. gambiae* annotated genes were input to MEME with the parameter -neg as background sequences. The other parameters used were -nmotifs 5, -minw 5 and -maxw 10. All other parameters had default values. Enriched motifs were then compared to characterized motifs using the online version of TOMTOM version 5.5.7 [52] to find for similar motifs in the FLY database with the option Combined Drosophila Databases.

Results

We previously identified a genetic locus linked to malaria susceptibility in wild mosquito pedigrees [2, 3] and recently cataloged non-coding regulatory elements using STARR-seq [26]. Here, we further explored STARR-seq identified non-coding regulatory elements within the previously mapped locus and genetically and functionally characterized an enhancer of interest.

Examination of enhancers near the mapped locus

From a genome wide STARR-seq enhancer map [26], we identified nine enhancers located ~0.4MB on either side of the previously described genetic marker (microsatellite H603) most significantly linked to differential malaria infection outcome in wild mosquito pedigrees challenged with sympatric wild parasite genotypes [2, 3]. To filter these nine enhancers for further characterization, multiple lines of evidence were integrated (i) ATAC-sequencing to identify regions of open chromatin overlapping with STARR-seq identified enhancers [26], (Fig. 1, Additional File 1_Figure S1), (ii) cloning of enhancer alleles from malaria susceptible and resistant mosquitoes and measurement of differential enhancer activity between alleles and (iii) using sequence data from the *Anopheles* 1000 Genomes project [39] to examine genetic variation shared in wild mosquitoes sampled from across Africa. Of the nine candidate enhancers detected by STARR-seq only three were also in regions of open chromatin, ENH_2L-02, ENH_2L-03 and ENH_2L-07 (solid black rectangles in Fig. 1). These candidate enhancers and open chromatin regions are shown relative to the position of the most significant genetic marker from previous mapping work (Fig. 1) and the PEST genome annotation (Fig. 1). Of the three enhancers in the STARR-seq catalog also located in regions of open chromatin, only two had genetic alleles that displayed significantly different enhancer activity between alleles cloned from malaria resistant and susceptible mosquitoes in the mapping pedigrees (ENH_2L-02 and ENH_2L-03, Fig. 1), and of these two enhancers, for one of them (ENH_2L-03) all nucleotide variants observed between alleles cloned from malaria susceptible and resistant mosquitoes also segregated in a large collection of *A. gambiae* and *A. coluzzii* sampled across Africa, indicating that they are widespread in nature. Based on the application of these three filters, ENH_2L-03 was prioritized for further functional characterization. In addition to being the only enhancer to pass all three filters, ENH_2L-03 is also the most proximal of the nine enhancers to the most significant genetic marker linked to malaria infection outcome (Fig. 1, [2, 3]). Characterization of ENH_2L-03 focused on, first, developing general methods to study the phenotypic function of enhancers in mosquitoes, and second, testing the functional significance of this enhancer in mosquito biology.

Enhancer genetic variation in wild mosquitoes

Using publicly available sequence generated as part of the *Anopheles gambiae* 1000 genomes project [39], genetic variation in wild mosquitoes from across Africa were examined for all nine enhancers in the genomic region controlling *Plasmodium* infection outcome. The variable sites across malaria susceptible and resistant individuals in the mapped locus [2] were then compared to natural variants from the Ag1000 genomes project and of the nine enhancers examined. Only in enhancers ENH_2L-03 and ENH_2L-09 were all variable sites captured in mapping pedigrees also variable in wild mosquitoes sampled across Africa. We prioritized enhancers for further characterization where variation segregating between malaria susceptible and resistant mosquitoes was also variable in nature, because the results would be more directly relevant to the natural population and be potentially more translatable into vector control tools. For each of the other seven enhancers, some genetically variable sites across malaria-resistant and susceptible individuals were either non-variable in nature or were variable at very low frequency (< 5%). For ENH_2L-03, each of the 5 nucleotide sites that were variable in malaria-resistant and susceptible pedigree mosquitoes was also variable in at least one natural population examined (see positions marked by stars in Fig. 2). Linkage disequilibrium analysis of the wild mosquito sequences by pairwise comparison of these 5 nucleotides in ENH_2L-03 indicate that none of the 5 positions are genetically linked, with all pairwise comparisons having an r^2 of < 0.12. Therefore, these nucleotides, despite that they distinguish susceptible and

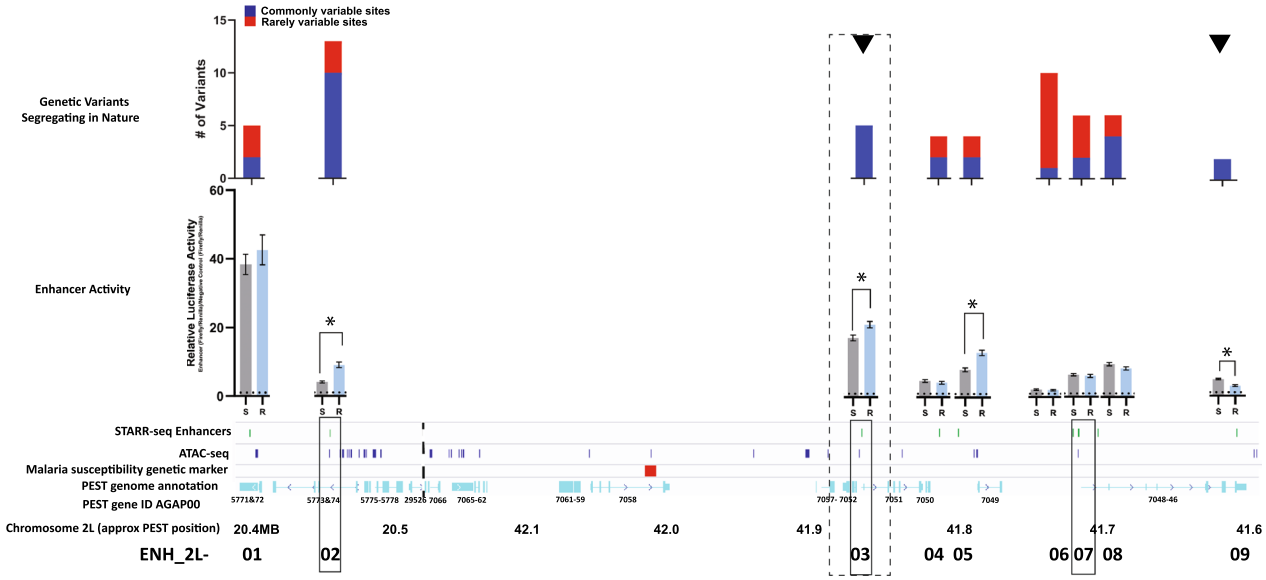

Fig. 1 Evidence filters for enhancer screening in malaria infection outcome locus. The enhancer screening strategy is summarized on a map of the PEST AgamP4 genome assembly. Reading from bottom to top, the line (ENH_2L-) displays positions and names of the nine STARR-seq identified enhancers within the locus. The next three lines indicate the PEST genome coordinates (Chromosome 2L (approx PEST position), gene ID (PEST gene ID AGAP00) and annotation if any (PEST genome annotation). The vertical dashed line at ~ 20.53 MB marks the proximal breakpoint of the 2La inversion; the genomic interval to the right of this breakpoint have been rotated from the PEST assembly inverted (2La+) orientation to the uninverted (2La) orientation of the mosquitoes used in the genetic linkage mapping study. Next, the line (Malaria susceptibility genetic marker) indicates the position of the most significant genetic marker for the infection outcome locus (red square). The line (ATAC-seq) indicates positions of ATAC-seq peaks from the current study, and (STARR-seq Enhancers) indicates the positions of the enhancers in the locus. Black outline rectangles highlight the 3 enhancers that overlap with ATAC-seq peaks. For all nine enhancers present in the locus, alleles were cloned from mapping pedigree mosquitoes susceptible and resistant to malaria infection and assayed for enhancer activity, displayed in line (Enhancer Activity). Enhancers with significantly ($p < 0.05$) different activity between susceptible (S, grey bars) and resistant alleles (R, light blue bars) are indicated by black asterisks. Finally, the top line (Genetic Variants Segregating in Nature) indicates the number of nucleotide variants distinct between alleles of S and R pedigree mosquitoes that also segregate in wild *Anopheles* sequences sampled across Africa, to detect the frequency of variants (dark blue, common variant, maximum allele frequency < 0.95; red, rare variants and fixed positions, maximum allele frequency > 0.95). Two enhancers harbored R versus S variable sites for which 100% also segregate in nature (black arrowhead). ENH_2L-03 (large, dashed rectangle) was the only enhancer that i) overlapped an ATAC-seq peak, and ii) displayed significantly different enhancer activity between alleles cloned from R and S individuals from the original mapping study, and iii) displayed nucleotide variants distinguishing R and S alleles that are common in nature

resistant alleles of ENH_2L-03, do not segregate in nature as a haplotypic linkage block.

Considering wild mosquitoes from the Ag1000 genomes project, there are 22 sites in the 555 bp ENH_2L-03 enhancer (including the 5 mentioned above) that are variable in wild mosquito samples sequenced as part of the Phase I Ag1000 genomes project. Analysis of wild variation was repeated for the other eight enhancers in the genomic region of interest and is available in Additional File 1 _Figure S2). Similar variation trends were noted across all enhancers examined, with enhancers having 25–33 commonly variable sites in nature for enhancers ranging in size from 503 to 702 bp.

Overlap of STARR-seq enhancers in the genomic locus with other chromatin surveys

The nine STARR-seq enhancers were compared to three published studies [27, 41, 42] that characterized open chromatin or epigenetic signatures within mosquito cell lines, whole mosquitoes, or mosquito tissues. Only ENH_2L-03 was also detected as a region of differentially accessible or epigenetically marked chromatin across all other studies (Additional File 1_Table S3).

Detection and depletion of ENH_2L-03 eRNA

Enhancer RNA (eRNA) is frequently associated with enhancer function. Detection of candidate eRNA colinear with ENH_2L-03 is required to develop RNAi assays to experimentally perturb enhancer function. To detect ENH_2L-03 candidate eRNA, primers were designed to amplify across two independent fragments of the dsENH_2L-03 eRNA. Amplification of the ENH_2L-03 eRNA and rpS7 control fragments was performed on cDNA synthesized from total RNA isolated from dsENH_2L-03 and dsGFP treated mosquitoes. Amplified products were visualized on a 2% agarose gel, which allows for both detection of the eRNA fragment in wild type (dsGFP) samples and provides evidence for

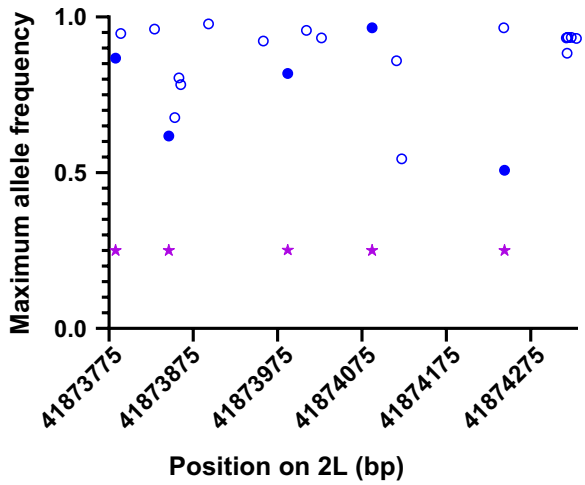

Fig. 2 Nucleotide variants that distinguish susceptible and resistant alleles of ENH_2L-03 are also variable in nature. Wild *Anopheles* sequences sampled across Africa were analyzed (n = 246 mosquitoes, 492 alleles) to calculate maximum allele frequency for each variant position in the 554bp STARR-seq interval of ENH_2L-03. All common variants (maximum allele frequency < 0.95) are indicated as blue circles (n = 22). Filled circles indicate the five natural population variant positions that distinguish S and R alleles in mapping pedigree mosquitoes (highlighted by violet stars); open circles indicate natural population variants that are not variant between S and R alleles of mapping pedigree mosquitoes. X-axis, PEST genome nucleotide coordinates for the 554bp STARR-seq enhancer interval. Genetic variation data for all nine enhancers are in Additional File 1_Supplementary Figure S2

depletion of the ENH_2L-03 eRNA in dsENH_2L-03 treated samples (Fig. 3A). Quantification of ENH_2L-03 candidate eRNA expression was determined by qRT-PCR, assessing the expression of ENH_2L-03 eRNA alongside the expression of AGAP007051, the gene in which ENH_2L-03 lies (Fig. 3B). This demonstrates that not only can an ENH_2L-03 candidate eRNA be detected, but it also can be depleted by dsRNA treatment. This also demonstrates that the silencing of the ENH_2L-03 candidate eRNA located in the intron of gene AGAP007051 does not influence AGAP007051 transcript abundance, indicating independence of the candidate eRNA and gene transcript. To reinforce this observation, treatment of mosquitoes with dsRNA directed against the AGAP007051 transcript depleted AGAP007051 transcript, but did not influence the abundance of the ENH_2L-03 candidate eRNA (Fig. 3C, D). Finally, to confirm candidate eRNA transcription in the direction antisense to the gene AGAP007051, specifically primed cDNA synthesis was performed on total RNA samples generated from mosquitoes treated with dsGFP, ds7051 or dsENH_2L-03, using a specific primer (EN03_Int_F in Additional File 1 Table S2), which detected an RNA molecule transcribed antisense to the AGAP007051 gene, and which is depleted only by treatment with the dsRNA directed against the ENH_2L-03 candidate eRNA, and not by dsRNA targeting AGAP007051 (Fig. 3E, Additional File1_Figure S3).

The effect of silencing the EN2L-03 candidate eRNA on gene expression

ENH_2L-03 eRNAs were silenced through the injection of cognate dsRNAs and total RNA was collected from 3–4-day old adult female mosquitoes. RNA-seq was performed to identify direct and indirect target genes whose expression was influenced, at least in part, by modulation of the level of candidate eRNA expressed from ENH_2L-03. Of the 13,845 annotated transcripts, 1970 had a sum of normalized read counts across samples below 10. A total of 15 genes whose expression was significantly ($p < 0.05$) altered following depletion of the ENH_2L-03 candidate eRNA were identified. With an FDR of 5%, less than one (precisely 0.75) differentially expressed gene would be expected due to false discovery. A volcano plot of RNA-seq results is provided in Additional File 1_Fig S4. Thirteen of these fifteen (87%) differentially expressed genes (DEG) were significantly upregulated, suggesting that the candidate eRNA expressed from ENH_2L-03 normally plays an inhibitory role in regulating their level of expression, and 2 genes were downregulated following dsRNA treatment, indicating that the candidate eRNA transcribed from ENH_2L-03 normally plays a positive regulatory role in their expression (Table 1). For these 15 DEGs, the sum of normalized read counts across samples averaged 21,922 reads with a range of 1457–104,734 reads.

An examination of the upstream regulatory region for these 15 genes (the 500 bp immediately upstream of the transcription start sites) using MEME suite, identified a 6-base-pair motif, CAGTYG, in 11 of the 15 upstream regions. The significance of enrichment of this 6-base-pair motif in the upstream region of DEGs was evaluated by comparing to the sequence 500 bp upstream of transcription start sites for all 13,780 predicted genes in the PEST genome. Across the genome, the same 6 bp motif occurs in 34% of upstream regulatory regions, while in the set of 15 DEGs after ENH_2L-03 eRNA depletion, the same motif is found in 73% of upstream sequences, representing a significant increase in the frequency of the motif presence ($p = 0.0012$). This motif enriched in the DEGs is similar to the characterized binding site for the Adf1 transcription factor [53]. Presence of an enriched motif similar to a known transcription factor binding site in a majority of genes with altered expression levels after ENH_2L-03 candidate eRNA depletion suggests a mechanism of potential co-regulation by the enhancer.

Enhancer silencing and infection with human and rodent malaria parasites

To examine the potential role of the ENH_2L-03 candidate eRNA in controlling infection with *Plasmodium* parasites, the ENH_2L-03 candidate eRNA was depleted by treatment with dsRNA (Fig. 3). Nevertheless, despite efficient silencing of ENH_2L-03 candidate eRNA expression, there was no significant difference in *Plasmodium falciparum* infection prevalence or intensity (Fig. 4A, B). Similarly, there was also no significant effect on either infection prevalence of infection or intensity after feeding on a *Plasmodium berghei*-infected mouse (Fig. 4C, D).

Discussion

A mosquito genetic locus discovered by a population-based study in Africa displays a large effect on mosquito susceptibility to malaria in nature [2, 3], but the mechanism of its action is not known. The most significant genetic marker linked to infection outcome is located in a large noncoding region, which raised the hypothesis of causative genetic variation in noncoding regulators such as enhancers, and led to the generation of an enhancer map using STARR-seq [26]. Here, we evaluated the nine enhancers located within the malaria susceptibility locus using computational and experimental approaches, which prioritized one of them as the enhancer of major interest, ENH_2L-03, for methods development and functional characterization. Such efforts to characterize the function of noncoding regulatory elements have the potential to inform future vector control strategies rooted in a region of the genome known to control resistance and susceptibility to malaria infection in wild mosquitoes [2, 3].

The current examination of ENH_2L-03 is grounded in natural genetic variation, both from the standpoint that examined differences in malaria-resistant and susceptible alleles were captured as wild variation from nature at the time of experimental infections [2], and also because the genetically variable sites in ENH_2L-03 underlying observed differences in luciferase activity are also variable across the African continent from population resequencing data [39]. Furthermore, in addition to an ATAC-seq peak from our new survey that overlaps with ENH_2L-03, multiple public chromatin surveys also identified peaks overlapping with ENH_2L-03. Thus, the focus on ENH_2L-03 is well-supported by genetic and chromatin evidence.

(See figure on next page.)

Fig. 3 Detection of ENH_2L-03 enhancer RNA and depletion using RNAi. **A** The image is an agarose gel showing detection of the candidate enhancer RNA (eRNA), contiguous with ENH_2L-03, which is depleted by treatment of mosquitoes with specific double-stranded RNA (dsRNA) targeting the eRNA molecules from three independent biological replicate (REP) templates. Ribosomal protein rpS7 RNA is amplified as an RNA input control. For each biological replicate (REP1, REP2 and REP3), labels above lanes indicate dsRNA treatment of mosquitoes with control directed against GFP (dsGFP) or against the eRNA (dsENH_2L-03); lanes labelled as "no cDNA control" are PCR control reactions without DNA template. Labels below the lanes indicate primers used for RT-PCR detection of either rpS7 or eRNA. **B** Quantitative assessment of expression using qRT-PCR approaches measuring the expression of ENH_2L-03 eRNA and AGAP007051. As ENH_2L-03 lies in the first intron of AGAP007051, this verifies that depletion of ENH_2L-03 eRNA by dsRNA approaches does not disrupt levels of AGAP007051 expression but does significantly deplete the expression of ENH_2L-03 eRNA. The fold change in the expression of AGAP007051 and ENH_2L-03 eRNA was measured by qPCR in samples collected from dsGFP (control) and dsENH_2L-03 groups. Results are normalized to the fold change in the dsGFP (control) group (dotted line). dsENH_2L-03 does not influence the expression of 7051 mRNA ($W = -9.00$; $p = 0.5$), whereas it strongly reduces the expression of ENH_2L-03 enhancer RNA in *A. coluzzii* ($W = 28.00$; $p = 0.01$). The ratio of the normalized gene expression in dsENH_2L-03/dsGFP was calculated using triplicates. Statistical analyses were performed using a Wilcoxon test between dsGFP and AGAP007051 or dsENH_2L-03. $N = 7$ biological replicates. **C** Agarose gel image shows specific depletion of AGAP007051 transcript when mosquitoes are injected with dsRNA targeting AGAP007051 and no depletion of the ENH_2L-03 candidate eRNA when mosquitoes are treated with dsRNA targeting AGAP007051. Amplification of rpS7 serves as a control for cDNA input into amplification reactions. Genomic DNA (gDNA) template is used as a positive control to confirm primer efficacy via successful gDNA amplification and to observe size difference with mRNA after intron splicing for AGAP007051 and rpS7, which demonstrates an absence of genomic contamination in Total RNA samples. Labels above lanes indicate templates and dsRNA treatments. Labels below indicate detected PCR amplicons. **D** Quantitative assessment of expression using qRT-PCR approaches measuring the expression of ENH_2L-03 eRNA and AGAP007051 in samples treated with dsRNA targeting AGAP007051. This verifies that depletion of AGAP007051 transcript by dsRNA approaches does not disrupt levels of ENH_2L-03 candidate eRNA expression ($p = 0.64$) but does significantly deplete the expression of AGAP007051 transcript ($p = 0.01$). The fold change in the expression of AGAP007051 and ENH_2L-03 eRNA was measured by qPCR in samples collected from dsGFP (control) and dsENH_2L-03 groups. Results are normalized to the fold change in the dsGFP (control) group (dotted line). **E**. The image is an agarose gel showing detection of the candidate ENH_2L-03 transcript primed by specific primer, which confirms transcript of a lncRNA over the ENH_2L-03 region, antisense to AGAP007051 transcript, that is only depleted in cDNA templates from mosquitoes treated with dsRNA targeting the ENH_2L-03 eRNA. Mosquitoes treated with other experimental (AGAP0070751) or control (dsGFP) dsRNAs and whose cDNA was specifically primed with the same primer designed against the ENH_2L-03 transcript did not show decreased levels of the targeted transcript. Labels above lanes indicate templates and dsRNA treatments. Labels below indicate detected PCR amplicons. Genomic DNA (gDNA) template is used as a positive control of primer efficacy. Uncropped gel images are available in Additional File 1_Fig S6. Gels in A and C use GeneRuler 100bp Plus DNA Ladder and in E GeneRuler 100bp DNA ladder

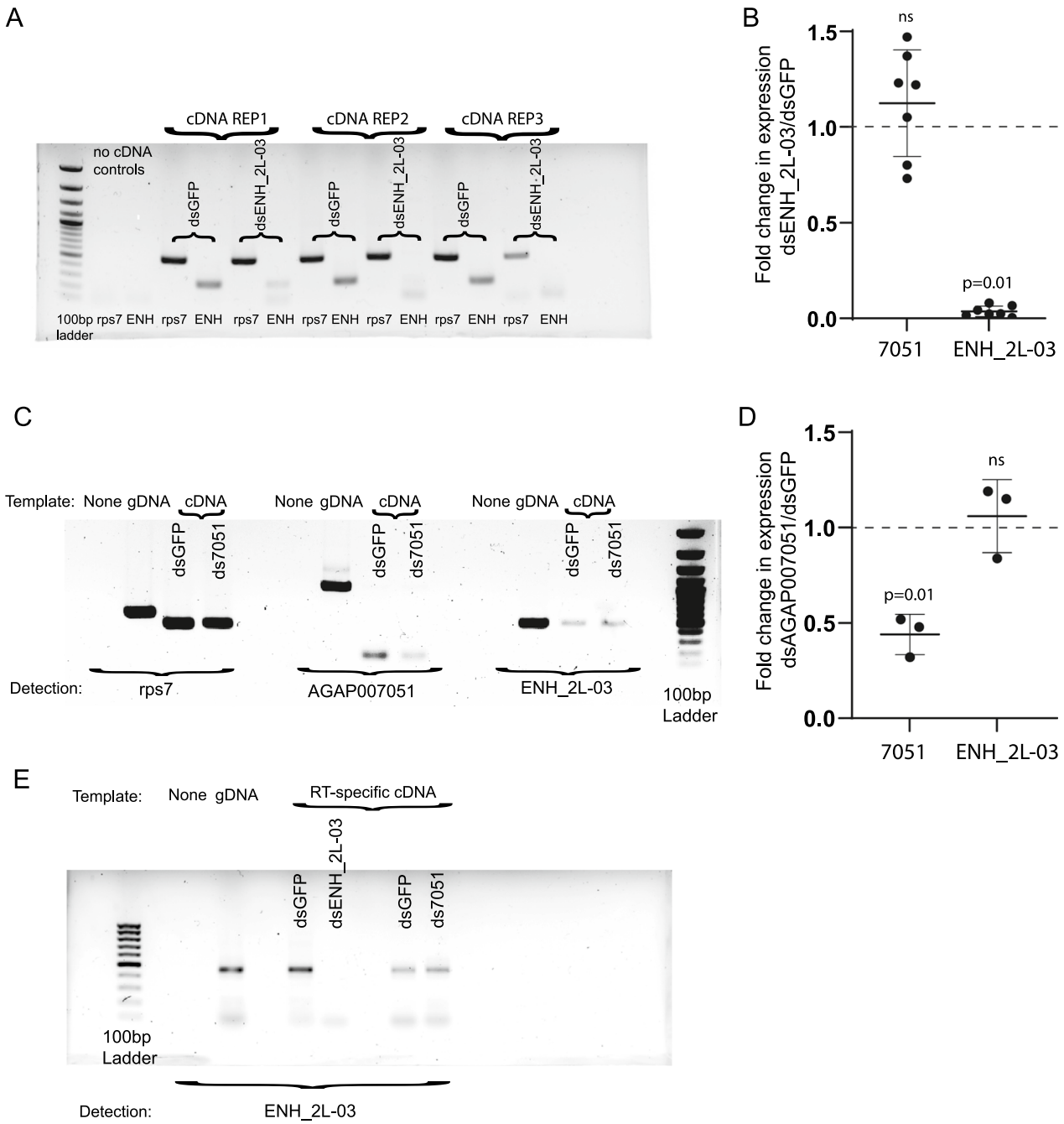

Fig. 3 (See legend on previous page.)

The current work demonstrates a lncRNA transcript contiguous with the ENH_2L-03 interval, and expressed antisense to the AGAP007051 gene, which harbors ENH_2L-03 within an intron. The lncRNA is consistent with the features of an eRNA transcript generated from the enhancer, although additional evidence would be required to definitively categorize the molecule as an eRNA, and we currently consider this molecule to be a candidate or putative eRNA. For example, global run-on sequencing (GRO-seq) would provide important evidence on the structure of the RNA molecule, although GRO-seq has not previously been performed in mosquitoes, the method would require process development, and the less mature non-model organism genome annotation could pose a limitation for mapping and interpretation. GRO-seq is thus beyond the scope of the current

Table 1 Fifteen genes displayed significantly altered gene expression upon silencing of candidate ENH_2L-03 eRNA expression

Id	Gene position	Gene name	FoldChange	padj	Contained Enriched Sequence motif similar to Adf1
AGAP011765	3L:33005140..33009840	Spondin like;glycoprotein; signaling ligand	1.89	0.00019	Yes
AGAP010363	3L:2242724..2243715		1.60	0.00115	No
AGAP001190	2R:1174559..1175564	Female reproductive tract protease GLEANR_896	1.76	0.00136	Yes
AGAP029723	3L:10281980..10283611		1.81	0.00136	Yes
AGAP000605	X:10822501..10823703		1.51	0.00148	Yes
AGAP004501	2R:57097933..57099532	Allantoicase	1.62	0.01005	Yes
AGAP000693	X:12440533..12441132	CEC1; cecropin anti-microbial peptide	−1.58	0.01305	Yes
AGAP006327	2L:29733363..29734698	LRIM6; leucine-rich immue protein (short)	1.44	0.01527	Yes
AGAP010550	3L:6251373..6252489	Venom allergen	1.42	0.02607	Yes
AGAP008632	3R:13774427..13779616	Alpha-aminoadipic semialdehyde synthase	1.37	0.04180	No
AGAP008892	3R:20451456..20453496		1.40	0.04180	No
AGAP029365	2L:47490581..47492360		1.36	0.04571	Yes
AGAP004115	2R:50171785..50177084	Cytinosin	1.38	0.04631	Yes
AGAP010814	3L:11184704..11190462	TEP6; thioester-containing protein 6	−1.35	0.04687	No
AGAP008209	3R:6915902..6917581	Cytochrome P450	1.29	0.04940	Yes

Thirteen genes were upregulated suggesting an inhibitory role of ENH_2L-03 eRNAs in their expression while 2 genes were downregulated, suggesting an activating role of ENH_2L-03 eRNA in their expression

manuscript, but it should represent future work aimed at cataloguing the eRNA catalog for *Anopheles* by us or other groups.

Experimental perturbation of enhancer activity can be achieved through different approaches. Here, we used RNAi silencing to modulate the abundance of the ENH_2L-03 candidate eRNA. CRISPR modification would also be possible, but would require sufficiently strong incrimination of a specific candidate to warrant that approach, which is more appropriate for confirmation rather than discovery. Upon depletion of ENH_2L-03 eRNA levels, bulk RNA sequencing identified 15 genes whose transcript abundance was significantly altered. Of these 15 differentially expressed genes (DEGs), three have predicted immune function, as a leucine rich repeat containing protein 6 (LRIM6) (AGAP006327), CEC1 antimicrobial peptide (AGAP000693), and TEP6 (AGAP010814). Of the 15 DEGs, only CEC1 and TEP6, were significantly downregulated when ENH_2L-03 eRNA was depleted, suggesting that ENH_2L-03 normally has a positive effect on expression of CEC1 (i.e., CECA) and TEP6. CEC1 expression has been previously shown to be induced by infection with bacteria and *Plasmodium* [54]. Single-cell transcriptomic analysis revealed co-regulation of CEC1 and TEP6 in the same hemocyte cluster cell upon blood feeding [55]. The same study also demonstrated expression of LRIM6 in a granulocyte (phagocytic cell) population cell cluster [55]. None of the 15 DEGs is likely a direct target gene of ENH_2L-03 as most (13/15) are in trans to ENH_2L-03 on other chromosome arms, and all are outside of topologically associating domains (TAD) boundaries using available published information [56]. Thus, the DEGs likely represent indirect target genes whose expression is influenced by depletion of ENH_2L-03 as part of a regulatory network.

(See figure on next page.)

Fig. 4 Depletion of ENH_2L-03 candidate eRNA does not influence *Plasmodium* infection. **A** Mosquitoes were injected with dsRNA targeting the ENH_2L-03 candidate eRNA or GFP (control) and were subsequently challenged with the human malaria *P. falciparum*. Mosquito midguts were dissected 7–8 days post feeding and parasite oocysts counted. Infection prevalence (the proportion of mosquitoes with at least one oocyst) was not significantly different between dsENH_2L-03 and dsGFP treated mosquitos across 3 biological replicates ($p > 0.256$). n indicates the total number of dissected mosquitoes over the three biological replicates. **B** *P. falciparum* infection intensity (the number of midgut oocysts in mosquitoes that had at least one), also was not significantly different between dsENH_2L-03 and dsGFP samples ($p > 0.14$). **C.** Similar experiments were conducted for rodent malaria *P. berghei* by feeding mosquitoes on a *P. berghei*-infected mouse and measuring infection prevalence. There were no consistent differences in infection prevalence upon treatment with dsENH_2L-03. n indicates the total number of dissected mosquitoes over the three biological replicates. **D** *P. berghei* infection intensity was also not statistically significant between dsENH_2L-03 and dsGFP treated mosquitoes ($p > 0.342$)

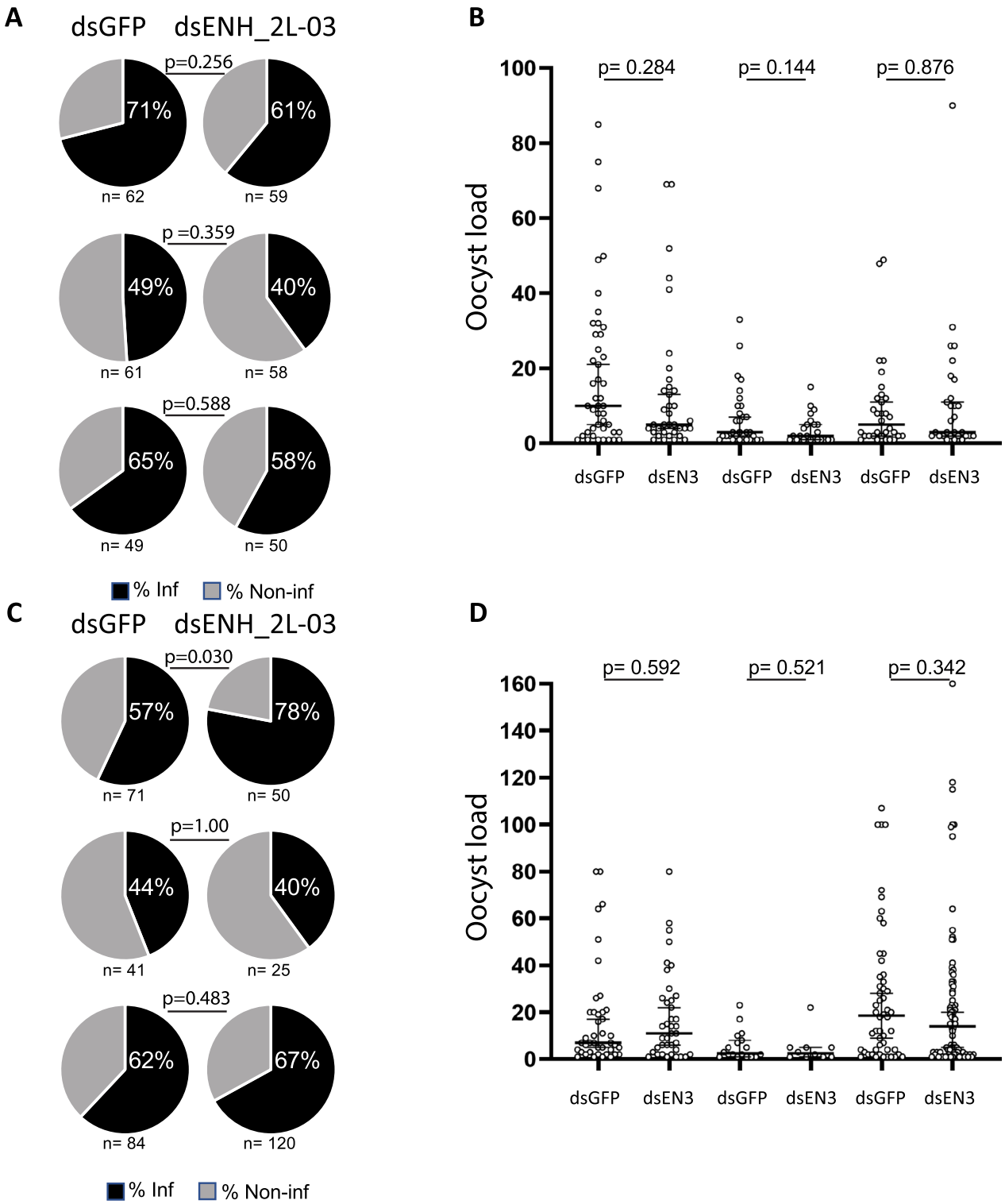

Fig. 4 (See legend on previous page.)

To examine this possibility, promoter sequences from the 15 DEGs were examined for conserved sequence and in 11/15 upstream sequences, the presence of a 6bp motif similar to the transcription factor binding motif, Adf1, has been identified, suggesting that recruitment of the Adf1 transcription factor might be required for

ENH_2L-03 activity to regulate the expression of at least these 11 of the 15 DEGs. In the dipteran ortholog *D. melanogaster,* Adf1 is required for regulating Alcohol dehydrogenase expression [57] and for long-term memory formation [58]. It also plays a role in behavioral immune response of *D. melanogaster* based on anticipatory medication of offspring, as a non-associative memory paradigm based on innate parasite recognition by the host [59]. The role of Adf1 in the coordinate expression of the ENH_2L-03 DEGs could be elucidated by silencing of Adf1 expression, which will require further work.

Silencing of the ENH_2L-03 candidate eRNA did not generate a significant phenotypic effect for either human or rodent *Plasmodium* infection. However, the function of an individual enhancer can be necessary but not sufficient to produce a phenotype, given the control of target genes by multiple enhancers, and the potential for redundant or compensatory effects of different enhancers acting upon a phenotype [60–62]. The enhancer ENH2L-03 has strong evidence pointing to its potential involvement in the infection outcome locus, and further work may be warranted, for example simultaneous silencing of ENH2L-03 with other enhancers in the locus that display lower individual priority, but that could act combinatorially to influence infection outcome.

Here we provide a proof of concept for the in vivo functional study of candidate transcriptional enhancers controlling natural resistance to malaria parasites in *Anopheles* mosquitoes. Functional characterization of active enhancers in relevant biological contexts has the potential to affect the design and implementation of strategies incorporating endogenous gene regulation in the generation of genetically modified and transgenic mosquitoes for vector control. As an example, CRISPR-based gene drive systems targeting an enhancer candidate that functionally controls *Plasmodium* infection outcomes could lead to population replacement strategies aimed at ultimately eliminating *Plasmodium*-susceptible mosquitoes. To this end, identification, and functional characterization of transcriptional enhancers in *Anopheles* represents progress and an increased tool set of regulatory factors to use in transgenic approaches.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13072-025-00597-3.

Additional file 1

Additional file 2

Acknowledgements
We thank the Institut Pasteur core facility, the Center for the Production and Infection of Anopheles (CEPIA) for rearing and infection of mosquitoes.

Author contributions
NZF, KST, KDV, CM, MMR designed the research. NZF, KST, CEA, RZ, EBF performed the research. NZF, KST, CEA, AP, RZ analyzed the data. KDV, CM and MMR wrote the manuscript with input from all authors. All authors have seen and approved the manuscript, and it has not been accepted for publication elsewhere.

Funding
This study received financial support to KV from the European Commission Horizon 2020 Infrastructures #731060 Infravec2; European Research Council Support for Frontier Research, Advanced Grant #323173 AnoPath; Agence Nationale de la Recherche #ANR-19-CE35-0004 ArboVec; to MMR from National Institutes of Health, NIAID #AI145999.

Availability of data and materials
The datasets supporting the conclusions of this article are available at the European Nucleotide Archive (ENA) (ATAC-seq data- PRJEB79891 https://www.ebi.ac.uk/ena/browser/view/PRJEB79891.) and the RNA-seq data at E-MTAB-14478 Biostudies https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-14478.

Declarations

Competing Interests
The authors declare no competing interests.

Author details
[1]Genetics and Genomics of Insect Vectors Unit, Department of Parasites and Insect Vectors, Institut Pasteur, Université Paris Cité, Paris, France. [2]Department of Microbiology and Immunology, Medical College of Wisconsin, Milwaukee, WI 53226, USA. [3]Arboviruses and Insect Vectors Unit, Department of Virology, Institut Pasteur, Université Paris Cité, Paris, France. [4]Hub de Bioinformatique Et Biostatistique, Institut Pasteur, Université Paris Cité, Paris, France.

Received: 24 January 2025 Accepted: 23 May 2025
Published online: 24 June 2025

References

1. Niare O, Markianos K, Volz J, Oduol F, Toure A, Bagayoko M, Sangare D, Traore SF, Wang R, Blass C, Dolo G, Bouare M, Kafatos FC, Kruglyak L, Toure YT, Vernick KD. Genetic loci affecting resistance to human malaria parasites in a West African mosquito vector population. Science. 2002;298(5591):213–6.
2. Riehle MM, Markianos K, Niare O, Xu J, Li J, Toure AM, Podiougou B, Oduol F, Diawara S, Diallo M, Coulibaly B, Ouatara A, Kruglyak L, Traore SF, Vernick KD. Natural malaria infection in Anopheles gambiae is regulated by a single genomic control region. Science. 2006;312(5773):577–9.
3. Riehle MM, Markianos K, Lambrechts L, Xia A, Sharakhov I, Koella JC, Vernick KD. A major genetic locus controlling natural Plasmodium falciparum infection is shared by East and West African Anopheles gambiae. Malar J. 2007;6:87.
4. Mitri C, Jacques JC, Thiery I, Riehle MM, Xu J, Bischoff E, Morlais I, Nsango SE, Vernick KD, Bourgouin C. Fine pathogen discrimination within the APL1 gene family protects Anopheles gambiae against human and rodent malaria species. PLoS Pathog. 2009;5(9): e1000576.
5. di Iulio J, Bartha I, Wong EHM, Yu HC, Lavrenko V, Yang D, Jung I, Hicks MA, Shah N, Kirkness EF, Fabani MM, Biggs WH, Ren B, Venter JC, Telenti A. The human noncoding genome defined by genetic diversity. Nat Genet. 2018;50(3):333–7.
6. Mitri C, Bischoff E, Takashima E, Williams M, Eiglmeier K, Pain A, Guelbeogo WM, Gneme A, Brito-Fravallo E, Holm I, Lavazec C, Sagnon N, Baxter RH, Riehle MM, Vernick KD. An evolution-based screen for genetic differentiation between anopheles sister taxa enriches for detection of functional immune factors. PLoS Pathog. 2015;11(12): e1005306.

7. Mitri C, Markianos K, Guelbeogo WM, Bischoff E, Gneme A, Eiglmeier K, Holm I, Sagnon N, Vernick KD, Riehle MM. The kdr-bearing haplotype and susceptibility to Plasmodium falciparum in Anopheles gambiae: genetic correlation and functional testing. Malar J. 2015;14(1):391.
8. Redmond SN, Eiglmeier K, Mitri C, Markianos K, Guelbeogo WM, Gneme A, Isaacs AT, Coulibaly B, Brito-Fravallo E, Maslen G, Mead D, Niare O, Traore SF, Sagnon N, Kwiatkowski D, Riehle MM, Vernick KD. Association mapping by pooled sequencing identifies TOLL 11 as a protective factor against Plasmodium falciparum in Anopheles gambiae. BMC Genomics. 2015;16:779.
9. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA, Hatan M, Carrasco-Alfonso MJ, Mayer D, Luckey CJ, Patsopoulos NA, De Jager PL, Kuchroo VK, Epstein CB, Daly MJ, Hafler DA, Bernstein BE. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518(7539):337–43.
10. Ivancevic A, Simpson DM, Joyner OM, Bagby SM, Nguyen LL, Bitler BG, Pitts TM, Chuong EB. Endogenous retroviruses mediate transcriptional rewiring in response to oncogenic signaling in colorectal cancer. Sci Adv. 2024;10(29): eado1218.
11. Kaplow IM, Lawler AJ, Schaffer DE, Srinivasan C, Sestili HH, Wirthlin ME, Phan BN, Prasad K, Brown AR, Zhang X, Foley K, Genereux DP, Zoonomia C, Karlsson EK, Lindblad-Toh K, Meyer WK, Pfenning AR. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. Science. 2023;380(6643): eabm7993.
12. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, Linder-Basso D, Plachetka A, Shanower G, Tolstorukov MY, Luquette LJ, Xi R, Jung YL, Park RW, Bishop EP, Canfield TK, Sandstrom R, Thurman RE, MacAlpine DM, Stamatoyannopoulos JA, Kellis M, Elgin SC, Kuroda MI, Pirrotta V, Karpen GH, Park PJ. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature. 2011;471(7339):480–5.
13. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Res. 2012;22(9):1748–59.
14. Okumu F, Moore SJ, Selvaraj P, Yafin AH, Juma EO, Shirima GG, Majambere S, Hardy A, Knols BGJ, Msugupakulya BJ, Finda M, Kahamba N, Thomsen E, Ahmed A, Zohdy S, Chaki P, DeChant P, Fornace K, Govella N, Gowelo S, Hakizimana E, Hamainza B, Ijumba JN, Jany W, Kafy HT, Kaindoa EW, Kariuki L, Kiware S, Kweka EJ, Lobo NF, Marrenjo D, Matoke-Muhia D, Mbogo C, McCann RS, Monroe A, Ndenga BA, Ngowo HS, Ochomo E, Opiyo M, Reithinger R, Sikaala CH, Tatarsky A, Takudzwa D, Trujillano F, Sherrard-Smith E. Elevating larval source management as a key strategy for controlling malaria and other vector-borne diseases in Africa. Parasit Vectors. 2025;18(1):45.
15. Pryce J, Medley N, Choi L. Indoor residual spraying for preventing malaria in communities using insecticide-treated nets. Cochrane Database Syst Rev. 2022;1(1): CD012688.
16. Tusting LS, Thwing J, Sinclair D, Fillinger U, Gimnig J, Bonner KE, Bottomley C, Lindsay SW. Mosquito larval source management for controlling malaria. Cochrane Database Syst Rev. 2013;2013(8): CD008923.
17. Wilson AL, Courtenay O, Kelly-Hope LA, Scott TW, Takken W, Torr SJ, Lindsay SW. The importance of vector control for the control and elimination of vector-borne diseases. PLoS Negl Trop Dis. 2020;14(1): e0007831.
18. Munyakanage D, Niyituma E, Mutabazi A, Misago X, Musanabaganwa C, Remera E, Rutayisire E, Ingabire MM, Majambere S, Mbituyumuremyi A, Ngugi MP, Kokwaro E, Hakizimana E, Muvunyi CM. The impact of Bacillus thuringiensis var. israelensis (Vectobac((R)) WDG) larvicide sprayed with drones on the bio-control of malaria vectors in rice fields of sub-urban Kigali, Rwanda. Malar J. 2024;23(1):281.
19. Haber DA, Arien Y, Lamdan LB, Alcalay Y, Zecharia C, Krsticevic F, Yonah ES, Avraham RD, Krzywinska E, Krzywinski J, Marois E, Windbichler N, Papathanos PA. Targeting mosquito X-chromosomes reveals complex transmission dynamics of sex ratio distorting gene drives. Nat Commun. 2024;15(1):4983.
20. Hancock PA, North A, Leach AW, Winskill P, Ghani AC, Godfray HCJ, Burt A, Mumford JD. The potential of gene drives in malaria vector species to control malaria in African environments. Nat Commun. 2024;15(1):8976.
21. Wang GH, Hoffmann A, Champer J. Gene drive and symbiont technologies for control of mosquito-borne diseases. Annu Rev Entomol. 2025;70(1):229–49.
22. Wudarski J, Aliabadi S, Gulia-Nuss M. Arthropod promoters for genetic control of disease vectors. Trends Parasitol. 2024;40(7):619–32.
23. Farley EJ, Eggleston H, Riehle MM. Filtering the junk: assigning function to the mosquito non-coding genome. Insects. 2021;12(2):186.
24. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, Turner JM, Bertelsen MF, Murchison EP, Flicek P, Odom DT. Enhancer evolution across 20 mammalian species. Cell. 2015;160(3):554–66.
25. Yang S, Oksenberg N, Takayama S, Heo SJ, Poliakov A, Ahituv N, Dubchak I, Boffelli D. Functionally conserved enhancers with divergent sequences in distant vertebrates. BMC Genomics. 2015;16:882.
26. Holm I, Nardini L, Pain A, Bischoff E, Anderson CE, Zongo S, Guelbeogo WM, Sagnon N, Gohl DM, Nowling RJ, Vernick KD, Riehle MM. Comprehensive genomic discovery of non-coding transcriptional enhancers in the african malaria vector Anopheles coluzzii. Front Genet. 2021;12: 785934.
27. Ruiz JL, Yerbanga RS, Lefevre T, Ouedraogo JB, Corces VG, Gomez-Diaz E. Chromatin changes in Anopheles gambiae induced by Plasmodium falciparum infection. Epigenetics Chromatin. 2019;12(1):5.
28. Yang MG, Ling E, Cowley CJ, Greenberg ME, Vierbuchen T. Characterization of sequence determinants of enhancer function using natural genetic variation. Elife. 2022;11: e76500.
29. Small S, Arnosti DN. Transcriptional Enhancers in Drosophila. Genetics. 2020;216(1):1–26.
30. Jacobs J, Pagani M, Wenzl C, Stark A. Widespread regulatory specificities between transcriptional co-repressors and enhancers in Drosophila. Science. 2023;381(6654):198–204.
31. Muller HM, Dimopoulos G, Blass C, Kafatos FC. A hemocyte-like cell line established from the malaria vector Anopheles gambiae expresses six prophenoloxidase genes. J Biol Chem. 1999;274(17):11727–35.
32. Bioinformatics, B. *FastQC*. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
33. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–8.
34. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10–2.
35. Stovicek J. *Genrich version 0.6.1*.
36. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL. The genome sequence of the malaria mosquito Anopheles gambiae. Science. 2002;298(5591):129–49.
37. Eggleston H, Njoya K, Anderson CE, Holm I, Eiglmeier K, Liang J, Sharakhov IV, Vernick KD, Riehle MM. Molecular characterization and genetic authentication assay for Anopheles "hemocyte-like" cell lines 4a–3A and 4a–3B. Parasit Vectors. 2022;15(1):465.
38. Nardini L, Holm I, Pain A, Bischoff E, Gohl DM, Zongo S, Guelbeogo WM, Sagnon N, Vernick KD, Riehle MM. Influence of genetic polymorphism on transcriptional enhancer activity in the malaria vector Anopheles coluzzii. Sci Rep. 2019;9(1):15275.
39. Anopheles gambiae Genomes, C., G. Data analysis, g. Partner working, c.-A. Sample, F. Burkina, Cameroon, Gabon, Guinea, B. Guinea, Kenya, Uganda, Crosses, Sequencing, p. data, d. Web application and c. Project. *Genetic diversity of the African malaria vector Anopheles gambiae.* Nature 2017;552(7683):96–100.

40. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, G. Genomes Project Analysis. The variant call format and VCFtools. Bioinformatics 2011;27(15): 2156–8.
41. Perez-Zamorano B, Rosas-Madrigal S, Lozano OAM, Castillo Mendez M, Valverde-Garduno V. Identification of cis-regulatory sequences reveals potential participation of lola and Deaf1 transcription factors in Anopheles gambiae innate immune response. PLoS One. 2017;12(10): e0186435.
42. Ruiz JL, Ranford-Cartwright LC, Gomez-Diaz E. The regulatory genome of the malaria vector Anopheles gambiae: integrating chromatin accessibility and gene expression. NAR Genom Bioinform. 2021;3(1): lqaa113.
43. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001;25(4):402–8.
44. Mitri C, Markianos K, Guelbeogo WM, Bischoff E, Gneme A, Eiglmeier K, Holm I, Sagnon N, Vernick KD, Riehle MM. The kdr-bearing haplotype and susceptibility to Plasmodium falciparum in Anopheles gambiae: genetic correlation and functional testing. Malar J. 2015;14(1):391.
45. Franke-Fayard B, Trueman H, Ramesar J, Mendoza J, van der Keur M, van der Linden R, Sinden RE, Waters AP, Janse CJ. A Plasmodium berghei reference line that constitutively expresses GFP at a high level throughout the complete life cycle. Mol Biochem Parasitol. 2004;137(1):23–33.
46. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
47. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
48. R Core Team. R: A language and environment for statistical computing. 2016; Available from: http://www.R-project.org/
49. Varet H, Brillet-Gueguen L, Coppee JY, Dillies MA. SARTools: A DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. PLoS ONE. 2016;11(6): e0157022.
50. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
51. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43(W1):W39-49.
52. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol. 2007;8(2):R24.
53. Cutler G, Perry KM, Tjian R. Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. Mol Cell Biol. 1998;18(4):2252–61.
54. Vizioli J, Bulet P, Charlet M, Lowenberger C, Blass C, Muller HM, Dimopoulos G, Hoffmann J, Kafatos FC, Richman A. Cloning and analysis of a cecropin gene from the malaria vector mosquito, Anopheles gambiae. Insect Mol Biol. 2000;9(1):75–84.
55. Kwon H, Mohammed M, Franzen O, Ankarklev J, Smith RC. Single-cell analysis of mosquito hemocytes identifies signatures of immune cell subtypes and cell differentiation. Elife. 2021;10: e66192.
56. Lukyanchikova V, Nuriddinov M, Belokopytova P, Taskina A, Liang J, Reijnders M, Ruzzante L, Feron R, Waterhouse RM, Wu Y, Mao C, Tu Z, Sharakhov IV, Fishman V. Anopheles mosquitoes reveal new principles of 3D genome organization in insects. Nat Commun. 2022;13(1):1960.
57. Heberlein U, England B, Tjian R. Characterization of Drosophila transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell. 1985;41(3):965–77.
58. DeZazzo J, Sandstrom D, de Belle S, Velinzon K, Smith P, Grady L, DelVecchio M, Ramaswami M, Tully T. nalyot, a mutation of the Drosophila myb-related Adf1 transcription factor, disrupts synapse formation and olfactory memory. Neuron. 2000;27(1):145–58.
59. Kacsoh BZ, Lynch ZR, Mortimer NT, Schlenke TA. Fruit flies medicate offspring after seeing parasites. Science. 2013;339(6122):947–50.
60. Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell. 2016;167(5):1170–87.
61. Panigrahi A, O'Malley BW. Mechanisms of enhancer action: the known and the unknown. Genome Biol. 2021;22(1):108.
62. Ray-Jones H, Spivakov M. Transcriptional enhancers and their communication with gene promoters. Cell Mol Life Sci. 2021;78(19–20):6453–85.

Publisher's Note