Ultra-deep exploration of transcription in *Anopheles gambiae*

Jaroslaw Krzywinski (PI) and Martin Donnelly
Liverpool School of Tropical Medicine, UK; jarek@liv.ac.uk

Executive summary

Catastrophic numbers of mosquito-borne disease cases and deaths remain one of the unsolved major health problems in developing countries. Malaria, arboviral diseases and filariasis kill more than one million people and/or debilitate hundreds of millions every year. Management of these diseases relies heavily on the use of insecticides. However, the evolution and spread of insecticide resistance can compromise the effectiveness of control campaigns and disease incidence continues to be extremely high. Approximately 90% of deaths caused by mosquito-borne disease are attributable to malaria in sub-Saharan Africa, where children under the age of five are the primary victims and the mosquito *Anopheles gambiae* is the major vector. Novel control strategies are urgently needed to combat mosquito-borne diseases, and to meet this challenge we must grasp the biological complexities of the vectors ranging from molecular to population level interactions.

We propose to develop a comprehensive picture of the *Anopheles gambiae* transcriptome using high-coverage RNA-seq through two discrete, complementary study aims.

Aim 1 will be to characterise the transcripts expressed during development and to utilise this ontogenetic series to drastically improve genome annotation through examining correlated patterns of gene expression. We propose next-generation sequencing of sexed whole-insect transcriptomes at 16 time points spanning from 10-hour old embryos to 20-day old adults. We will identify transcribed features and comprehensively characterize their expression dynamics throughout development.

Aim 2 is to discover additional genes that are induced following physiological stress. To document global transcriptional response to a range of physiological stressors we propose to conduct RNA-seq experiments on females subject to (1) sub-lethal insecticide exposure and (2) sub-optimal relative humidity (RH), and (3) females post-blood feeding and during vitellogenesis/oogenesis.

The RNA-seq reads from both sets of experiments will be mapped to the PEST and the S molecular form genome sequences to provide an insight into dynamic changes in transcription levels and transcript models will be created to provide a nearly exhaustive catalogue of transcribed features in each sample. Finally, the RNA-seq data will be integrated into Vectorbase as new tracks.

Our study will have a profound impact on the vector research community. Discovery of a complete gene set is a pre-requisite for an understanding of the molecular basis of the development of *An. gambiae* and its response to physiological and environmental stressors. We expect that deep insight into these processes will allow identification of novel anti-vector targets and will result in effective interventions in disease transmission. Sexed transcriptomes are expected to shed light on mosquito sex determination and sexual differentiation processes that may be targeted in genetic control approaches. A highly improved *An. gambiae* transcript dataset will also allow efficient implementation of comparative methods to the annotation of additional *Anopheles* genomes that are currently in a sequencing pipeline. Further, it will be highly informative for improvement of annotations of the already sequenced culicine, sandfly, tsetse fly and other vectors genomes.