Review

Experimental and computational applications of microarray technology for malaria eradication in Africa

Victor Chukwudi Osamor

Department of Computer and Information Science(Bioinformatics Unit)College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria. E-mail: vcosamor@gmail.com.

Various mutation assisted drug resistance evolved in *Plasmodium falciparum* strains and insecticide resistance to female Anopheles mosquito account for major biomedical catastrophes standing against all efforts to eradicate malaria in Sub-Saharan Africa. Malaria is endemic in more than 100 countries and by far the most costly disease in terms of human health causing major losses among many African nations including Nigeria. The fight against malaria is failing and DNA microarray analysis need to keep up the pace in order to unravel the evolving parasite's gene expression profile which is a pointer to monitoring the genes involved in malaria's infective metabolic pathway. Huge data is generated and biologists have the challenge of extracting useful information from volumes of microarray data. Expression levels for tens of thousands of genes can be simultaneously measured in a single hybridization experiment and are collectively called a "gene expression profile". Gene expression profiles can also be used in studying various state of malaria development in which expression profiles of different disease states at different time points are collected and compared to each other to establish a classifying scheme for purposes such as diagnosis and treatments with adequate drugs. This paper examines microarray technology and its application as supported by appropriate software tools from experimental set-up to the level of data analysis. An assessment of the level of microarray technology in Africa, its availability and techniques required for malaria eradication and effective healthcare in Nigeria and Africa in general were also underscored.

Key words: Malaria, microarray, Gene expression profile, *Plasmodium falciparum*, RNA

INTRODUCTION

A DNA microarray (also commonly known as gene or genome chip, DNA chip, or gene array) is defined by Wikipedia encyclopedia as a collection of microscopic DNA spots, commonly representing single genes, arrayed on a solid surface by covalent attachment to chemically suitable matrices. In the paper of Chen (2006), DNA microarray is defined as an array of tens of thousands of molecular sequences (that is, probes) mobilized in the form of DNA on a solid and planar platform on a microscopic and high-density scale and can be used in hybridization experiments for the parallel measurement of the quantity of bound homologous sequences in biological samples.

However, we define DNA Microarray as a technology with a grid of nucleic acid molecules of known composition called probe, placed/immobilized on a solid substrate or slide and used to hybridize messenger RNA (mRNA)

from a target cell or tissue of unknown composition to reveal changes in gene expression relative to a control sample. By hybridization, we mean that the four nitrogenous bases of the probe pair up with their complimentary nitrogenous bases in the unknown or tested sample, such that Adenine(A) pairs with Thymine (T)/Uracil(U) and Cytosine(C) pair Guanine(G). Microarray technology, which is also known as "DNA chip" technology, allows the expression of many thousands of genes to be assessed in a single experiment.

The use of microarrays for gene expression profiling was first published in Schena et al. (1995) and the first complete eukaryotic genome (*Saccharomyces cerevisiae*) on a microarray was published in 1997 (Tatusov et al., 1997). A microarray experiment involves monitoring gene expressions as the cell undergoes some biological processes. These experiments are often used to measure
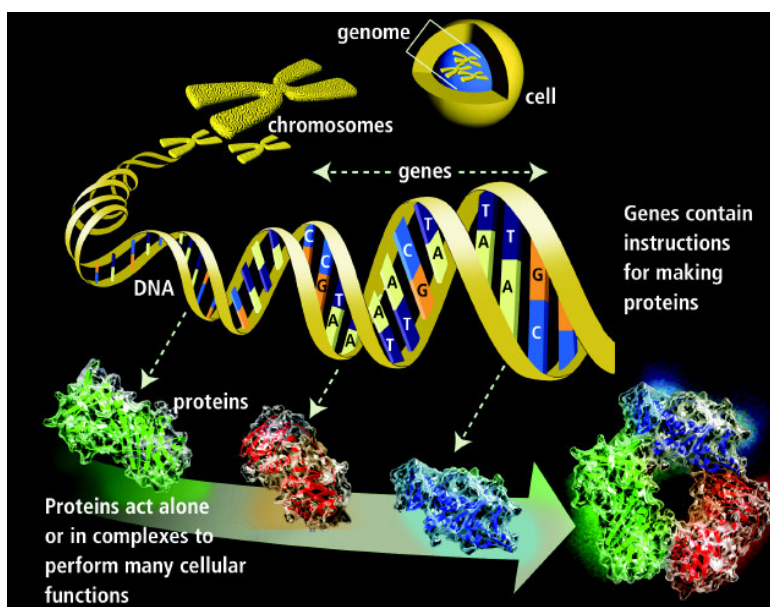
Figure 1. Process of Protein Synthesis (DNA->mRNA->Protein).

gene expression and therefore are able to detect differences in gene expression between two populations of cells; a test population (disease cell or tissue) versus a control population (normal cell or tissue). However, the experimental and control gene expression values ratio is computed and used. Huge data is generated and biologist has the challenge of extracting useful information from volumes of microarray data. Expression levels for tens of thousands of genes can be simultaneously measured in a single hybridization experiment and are collectively called a "gene expression profile".

The protozoan *Plasmodium falciparum* makes an excellent organism for this type of experiment because of its devastating economic importance coupled with the fact that its genome has been sequenced and ORFs (Open Reading Frames) have been determined. Drug resistance in evolving *P. falciparum* strains and insecticide resistance to female Anopheles mosquitoes accounts for major biomedical catastrophe standing against all efforts to eradicate malaria in Sub-Saharan Africa. Among these diseases, malaria is endemic to more than 100 countries (Le Roch et al., 2003) and by far the most costly in terms of human health causing major losses among many African nations including Nigeria. It is caused by protozoan parasites of the genus *Plasmodium* and infects approximately 500 million people annually, killing more than one million, mainly children and pregnant women in Africa (Le Roch et al., 2003).

The fight against malaria is failing (Osamor, 2009) and microarray analysis need to keep up the pace to unravel the evolving parasite's gene expression profile which is a pointer to monitoring the genes involved in malaria's infective metabolic pathway. Gene expression profiling has been commonly used to study the pathogen's or

host's responses to each other or to the external stimuli such as drug or vaccine treatments. Gene expression profiles can also be used in studying various state of malaria development in which expression profiles of different disease states at different time points are collected and compared to each other to establish a classifying scheme for purposes of diagnosis and treatments with adequate drugs.

Each particular cell or tissue in the body has a nucleus bearing a number of chromosomes with genes containing information in its DNA, about the type of needed protein to be produced by the cell. The characteristic of producing different sets of proteins passes the process called transcription by copying the DNA genetic information of the needed protein to form mRNA (an intermediate product) and finally to protein biomolecules through translation as shown in Figure 1.

This molecular biology dogma, *DNA => mRNA => Protein* is the basis of microarray technology as DNA microarray measures mRNA transcript as gene expression level. DNA code (genetic code) which is a triplet codon arising from a combination of three(3) of the four(4) nitrogenous bases: Adenine(A), Thymine(T), Guanine(G), and Cytosine(C) form the type of amino acid or protein to be produced while a stop codon terminates the process.

PATHOGENICITY OF *P. FALCIPARUM* MALARIA

When a plasmodium-infected mosquito bites a person, it transmits the sporozoites on its salivary glands into the subcutaneous layer of the skin into the bloodstream which migrates to the liver for exo-erythrocytic development
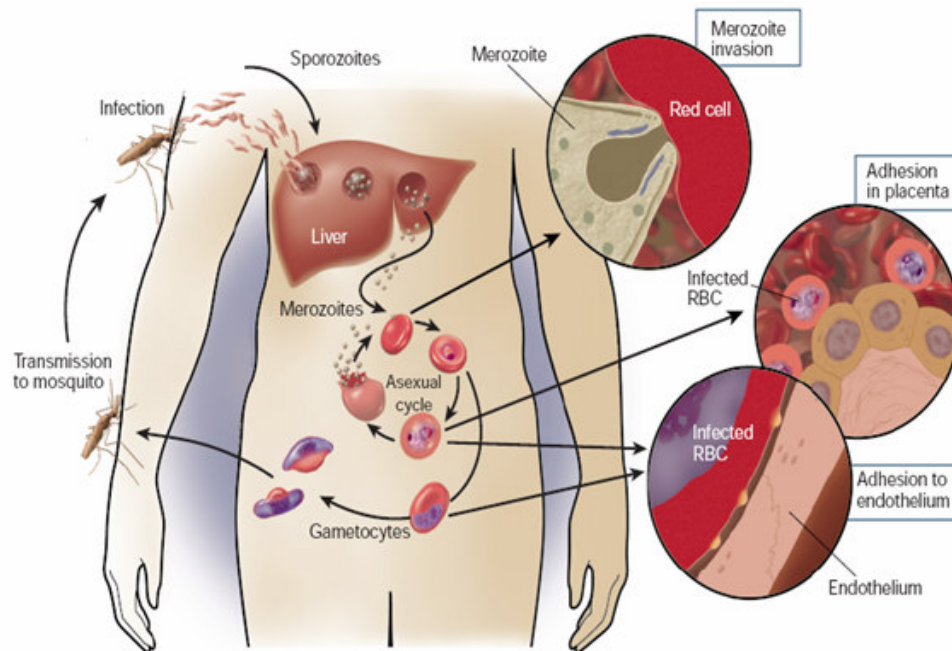
Figure 2. Life Cycle of *Plasmodiun falciparum* in Liver and Red Blood Cell (Miller et al., 2002).

by asexual means (Figure 2). The co-receptor on sporozoites that mediates invasion involves, in part, the thrombospondin domains on the circumsporozoite protein (CSP) and on thrombospondin-related adhesive protein (TRAP). These domains bind specifically to heparin on hepatocytes. Inside the hepatocyte, each sporozoite develops into tens of thousands of merozoites, which can each invade an RBC on release from the liver. Disease begins only once the asexual parasite multiplies in RBCs, producing around 20 merozoites per mature parasite within 48 hours, with each merozoite able to invade other RBCs.. This is the only gateway to disease. A small proportion of asexual parasites is converted to gameto-cytes that are essential for transmitting the infection to others through female anopheline mosquitoes (Miller et al., 2002).

Invasion of RBCs and binding of parasitized RBCs to vascular endothelium and placenta

In respect to RBC invasion, Miller et al. (2002) noted that what remains completely unknown is which merozoite surface molecules recognize the RBC surface and then signal the start of the invasion process. The parasite induces a vacuole derived from the RBC's plasma membrane and enters the vacuole. Three organelles on the invasive (apical) end of the parasite (rhoptries, micronemes and dense granules) define the phylum Apicomplexa. Receptors that mediate invasion of RBCs by merozoites and invasion of liver by sporozoites are

found in micronemes, on the cell surface, and in rhoptries. Identifying the signaling pathways that release organelle contents on contact with a host RBC is a critical issue in parasite biology (Figure 2). Malaria parasites have intracellular signaling pathways mediated by phosphoinositide, cyclic AMP and calcium- dependent mechanisms. Invasion events include releasing essential molecules from apical organelles and initiating the actin-myosin moving junction that brings the parasite inside the vacuole that forms in the RBC.

Although other parasite proteins on the merozoite surface and in apical organelles have been proposed as receptors, there is no direct evidence so far. Because invasion is such a complex series of events from RBC binding, to apical reorientation, to entry, it seems likely that several proteins are required for efficient invasion. For example, evidence has suggested that RBC invasion requires the cleavage of a surface protein on the RBC by an unknown parasite serine protease (Miller *et al.* (2002). Thus, the molecular and cellular events surrounding each step in invasion still remain to be elucidated. Understanding these pathways will give insight into parasite virulence and will facilitate rational vaccine design against merozoite invasion. A single parasite protein, *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), which is expressed at the infected erythrocyte surface mediates parasite binding to all the various receptors. PfEMP1 is encoded by the large and diverse var gene family that is involved in clonal antigenic variation and has a central role in *P. falciparum* pathogenesis. Adherence protects the parasite from des-
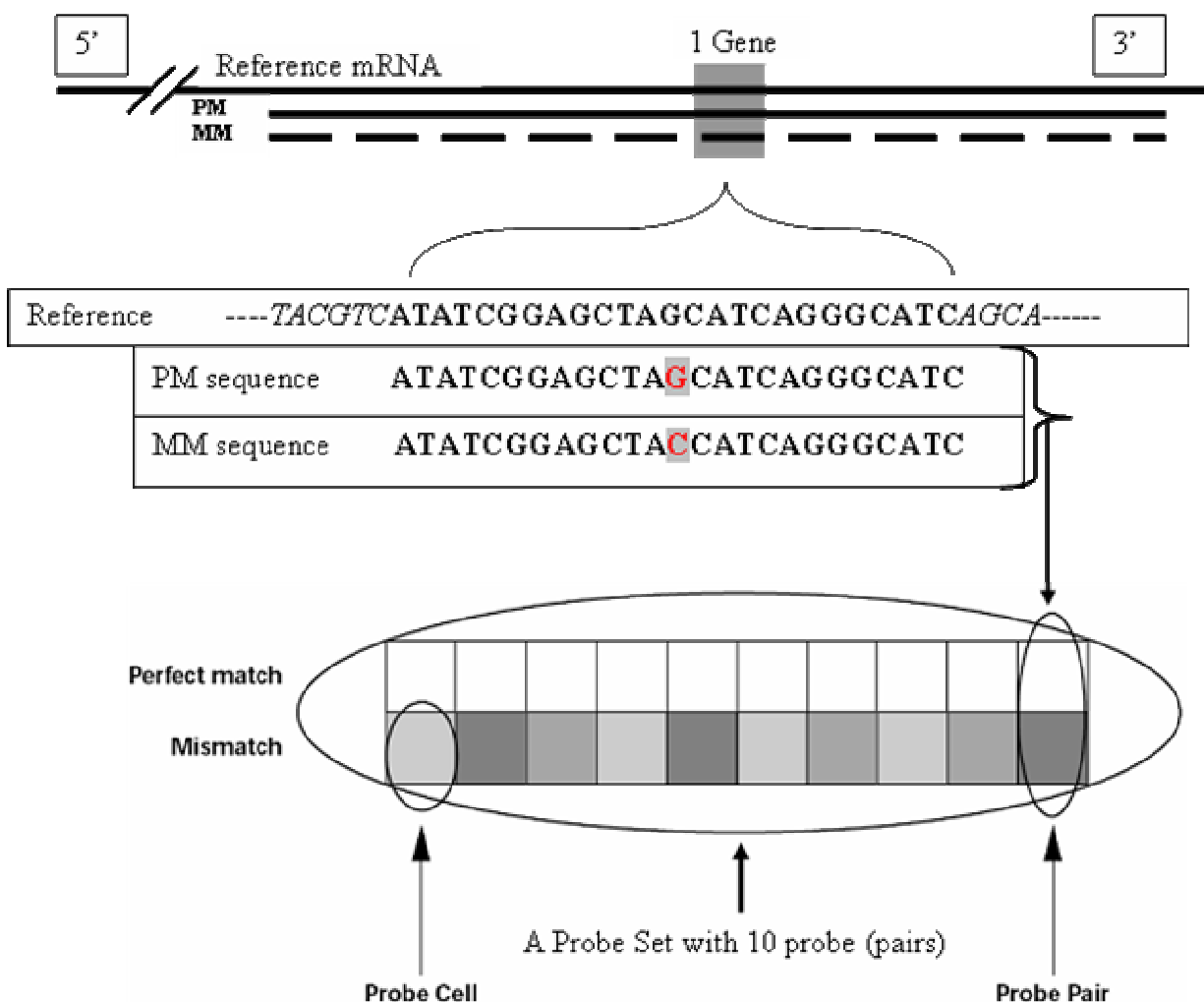
Figure 3. Design of Oligonucleotide probe (feature) showing its location in the Probe set and 1 gene represented by 1 probe.

truction, as non-adherent mature parasitized RBCs are cleared rapidly in the spleen.

DNA MICROARRAY TECHNOLOGY

Microarray fabrication and probe design

Microarrays fabrication is achieved through two techno-logies and involves either DNA *deposition* or *in situ* syn-thesis. While the deposition method allows the deposition of PCR-amplified cDNA clones and printing of already synthesized oligonucleotides with fine-pointed pins onto glass slides, *in situ* manufacturing is by photolithography using pre-made masks, ink-jet printing, or electrochemis-try on microelectrode arrays.

Let nt be nucleotides, nucleic acid microarrays primarily use short oligonucleotides (15 - 25 nt), long oligo-nucleotides (50 - 120 nt) and PCR-amplified cDNAs (100 - 3,000 base pairs) as array elements. Short probes are possible because they are selected using special algorithms to run on already sequenced genome data to find unique sequences that serves as a representative of each gene in an organism. Figure 3 shows a probe with both Perfect match (PM) and Mismatch (MM) pair aligned to a reference mRNA that is provided from information on sequenced organism.

A probe cell also called feature contains 25 nucleotides and can be a PM or MM. Usually probes are manufac-tured in pairs such that PM has same sequence of 25 nucleotides as the MM except for one nucleotide at the middle (13th) position which are complementary. PM hybridizes with the experimental sample to measure the degree of signal intensity while the MM hybridizes to give value for background subtraction which improves data accuracy. One or more probes can be used to represent a gene and a typical Affymetrix probe contains about 16 - 20 probes in a probe set.
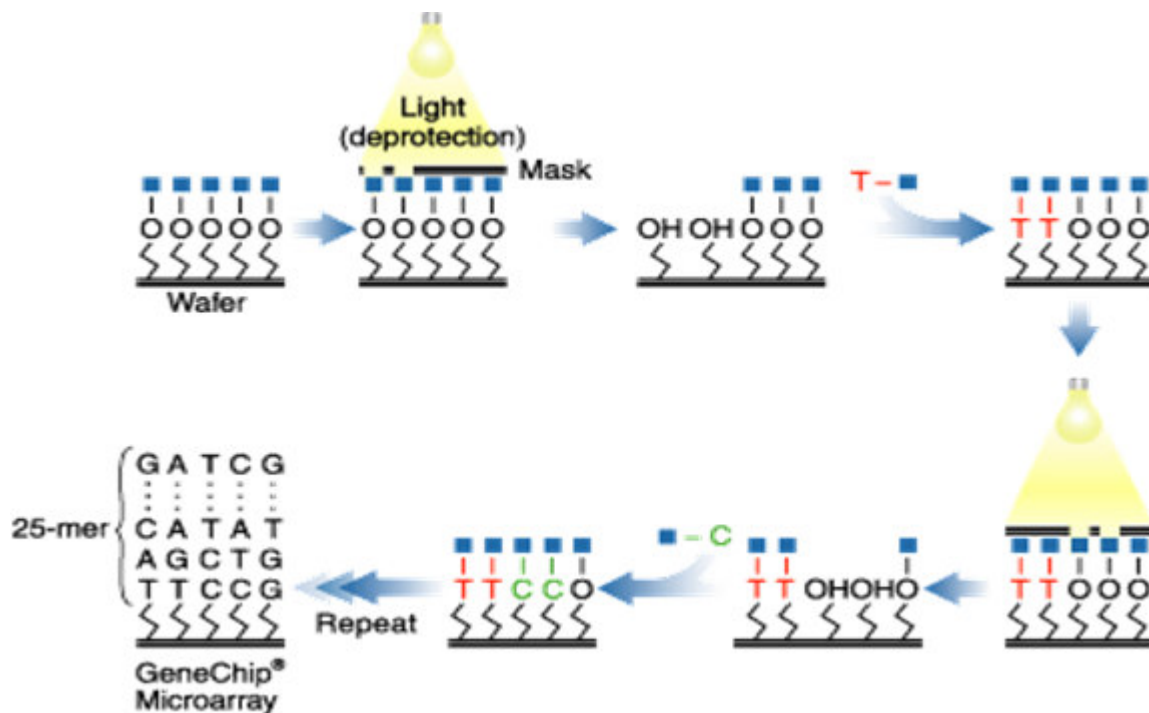
Figure 4. Photolithographic manufacture of Oligonucleotide array.

Photolithographic manufacturing process produces GeneChip arrays with millions of probes on a small glass chip or substrate called wafer or array. The photolithographic process begins by coating a 5″ x 5″ quartz wafer with a light-sensitive chemical compound that prevents coupling between the wafer and the first nucleotide of the DNA probe being created.

Lithographic masks are used to either block or transmit light onto specific locations of the wafer surface. The surface is then flooded with a solution containing either adenine, thymine, cytosine, or guanine, and coupling occurs only in those regions on the glass that have been deprotected through illumination (Affymetrix). It continues till probes reach their full length of about 25 nucleotides (Figure 4) with about 1.3 million features on one array.

Experimental design

Microarray experimental design allows researchers to test and vary the input variables that impact on the microarray experiment to get correct output. These involve three major principles that entail: Replication, Randomisation and Blocking. In replication, duplicate or repeat of the same experiment more than once but varying one factor like changing the location of same probe on the array to monitor its behaviour. This gives estimate of the experimental error if the observed differences in the data are significant. Randomisation means that experimental data can be monitored by allowing probes to be placed on the array in no particular order (random). Blocking allows the experimenter to keep all nuisance factors (factors that are not of interest but can affect the experimental outcome) while the interesting factor is varied e.g. Slide is a block (Draghici, 2003).

Some microarray specific types of experimental design include reference design and loop design. In reference design many sample conditions or timepoints $t_1, t_2, \ldots, t_n$ are pairwise compared to only one reference sample, *ref,* that samples 1 to n while *ref* is measured *n* times (Quackenbush, 2005). Variability error is introduced by dyes and can be stabilized by swapping the R and G dyes to the optimized measurement (twice) for each sample and thereby collect more information on interesting data using same 2n arrays. This is the loop design. Vinciotti et al. (2005) studied the relative efficiency of both a loop and a reference design using the same RNA preparations. Their results of these experiments show that (1) the loop design attains a much higher precision than the reference design, (2) multiplicative spot effects are a large source of variability, and if they are not accounted for in the mathematical model, for example, by taking logratios or including spot effects, then the model will perform poorly.

cDNA microarray experiment

Spotted or cDNA or two-channel microarrays consists of thousands of individual DNA sequences called probe printed in a high density array on a glass microscopic
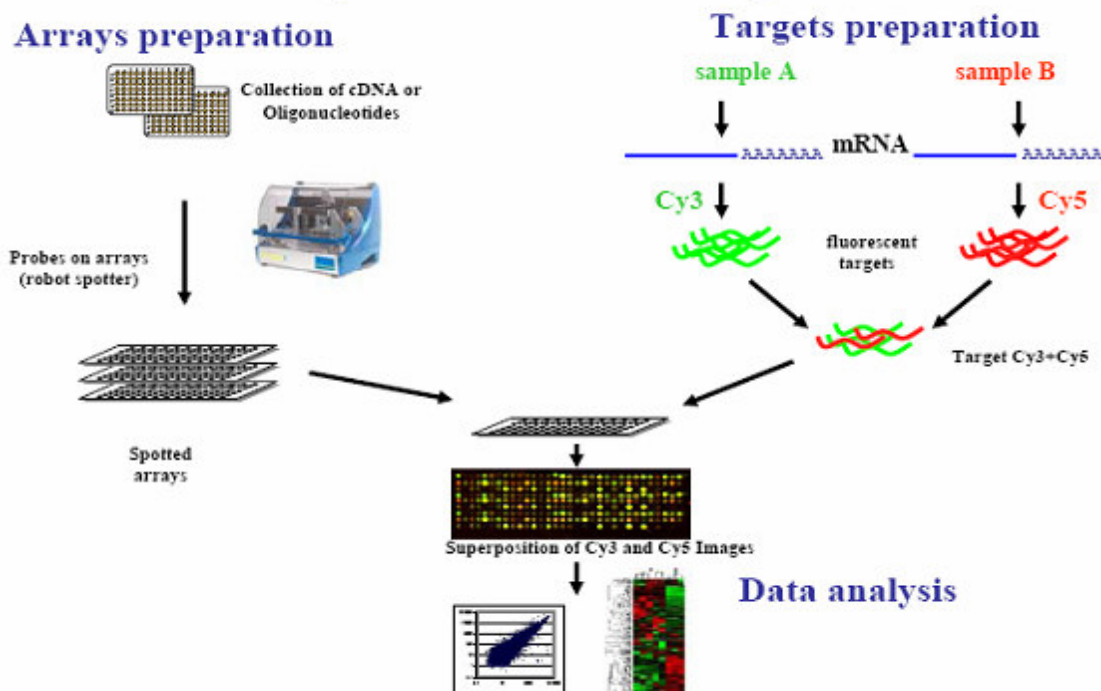
Figure 5. cDNA microarray experiment.

slide using a robotic arrayer. The process of probe printing is shown in Figure 5 during array preparation.

During target preparation mRNA is extracted from two samples called A and B to be studied. These targets mRNA samples are reverse-transcribed into cDNA and labeled using different fluorescent dyes where Cy3 represents green and Cy5 represents red Figure 5. Labeled samples are mixed together and competitively hybridized with the probe on the array to give rise to image for analysis. Relative abundance of spotted DNA sequences or probe in two samples may be assessed by monitoring the differential hybridization of the two samples to the sequences on the array.

Oligonucleotide microarray experiment

Once the microarray is constructed, oligonucleotide chip experiment requires the preparation of sample for GeneChip arrays. Messenger RNA (mRNA) is extracted from the cell and converted to cDNA as shown in Figure 6. It then undergoes amplification and labeling where the target mRNA population is labeled, typically with a fluorescent dye, so that hybridization to the probe spot can be detected when scanned with a laser (Gibson, 2003). Fragmentation and hybridization of the sample to 25-mer oligos on the surface to the chip takes place under an appropriate temperature.

The next step is the washing of unhybridized material and the chip is scanned in a confocal laser scanner and the image analyzed by computer. This approach provides a way to use directly the growing body of sequence information for experimental investigations (Wosik, 2006). However, one sample is hybridized on one array for one-colour chanelled oligonucleotide microarray unlike two-colour chanelled cDNA microarray capable of hybridizing 2 distinctly labeled(R and G) samples on one array.

Microarray images and data analysis

After competitive hybridization, images of slides are taken by microarray scanner which makes florescent measurement of each dye. The ratio of the florescence intensities for each spot is indicative of the relative abundance of the corresponding DNA sequence in the nucleic acid samples. cDNA microarray image processing steps generate two main quantities R (red) and G (green) for each spot on the array, thereby measuring transcript abundance for red and green mRNA labeled samples. The R and G values are usually combined or normalized into a single log-intensity ratio, $log_2 R/G$, measuring relative transcript abundance in the two samples. A positive log-ratio denotes gene over-expression while a negative log-ratio denotes gene under-expression. Normalisation needs to be done before clustering for further data analysis so as
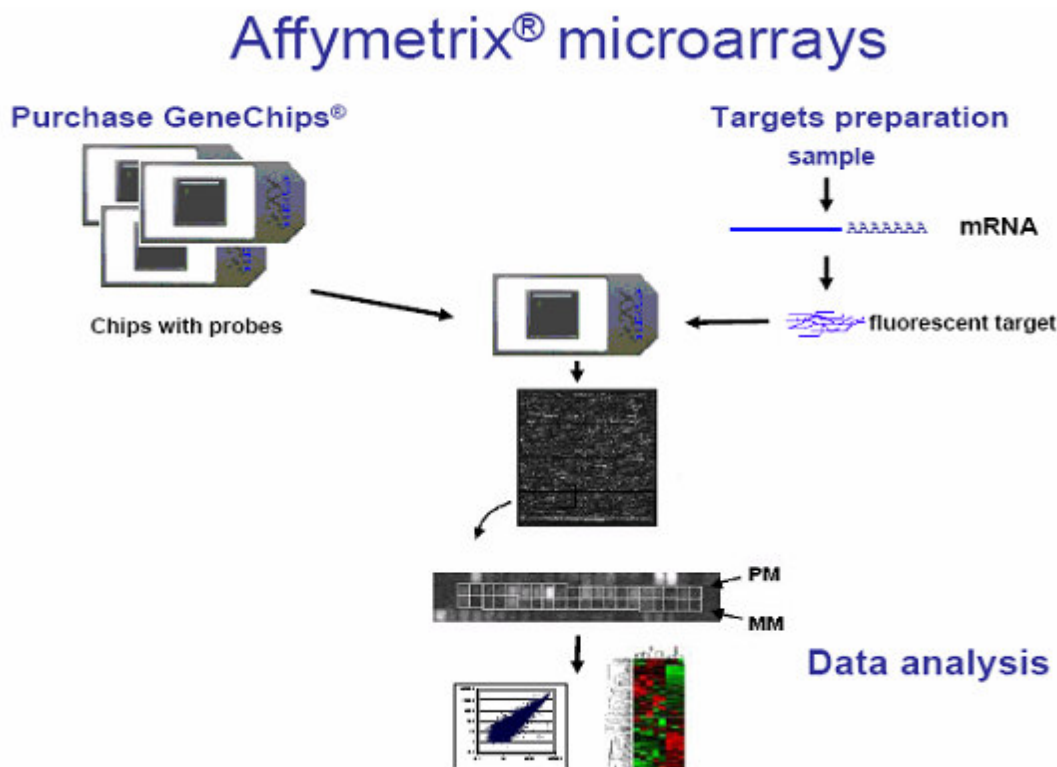
Figure 6. Oligonucleotide chip experiment.

to identify and remove systematic sources of variation such as different labeling efficiencies, scanning properties of dyes used, and print tip of robotic arrayer during probe spotting.

Microarray gene expression profiles are often subjected to cluster analysis and pathway mapping to unveil groups of co-regulated genes - a practice that is referred to as regulatory network and metabolic pathways discovery or reconstruction. A common early step in microarray data analysis is log transformation. Typically, log base 10 is used; however, log base 2 or natural log will work equally well. Log transformation has several important effects on the data. The most critical reason to log transform microarray data is that some of the error in the signal intensity measurement increases as the magnitude of signal intensity increases. That is, small numbers have less error in an absolute sense than higher numbers. Fortunately, higher numbers have roughly the same percentage error as small numbers. This roughly constant factor can be simply calculated and subtracted to normalize the data once the signals have been log transformed. There are additional effects of logging that make log transformed microarray data more closely fit statistical assumptions when applying statistical test methods. Log transformation makes data more symmetrical, one of the standard assumptions of normality. Log transformation also reduces the influence of a single measurement. Means on a log scale are more like geometric means,

which are resistant to the effects of outliers, and it follows that outliers result in better estimates of variance. So, by log transforming data, common statistical methods are made more reasonable and provide more accurate insights to the biologist (Affymetrix, 2001).

The relative difference between a Perfect Match and its Mismatch is Discrimination Score [R]: R= (PM-MM)/(PM+MM))(Affymetrix). A qualitative measurement indicating if a given transcript is detected (present), not detected (absent), or marginally detected (marginal) is a measure of value[R]. Statistical model both parametric and non-parametric such as Students t-test, ANOVA, Mann-Whitney test, f-distribution, Wilcoxon signed rank test etc form the basis of several formulae and algorithms that are used to compute differential gene expressions, image segmentations, normalizations and clustering algorithms. However, many software tools are available to serve as resource that supports most microarray analysis functions.

Microarray software support

There are various commercial and open source software products existing currently to support DNA microarray analysis. For this work, we consider a free open source software TM4 (Saeed et al., 2003) from The Institute of Genomic Research (TIGR) used for cDNA microarray and a commercial GeneChip Operating Software (GCOS)

from Affymetrix and dChip software for Oligonucleotide microarray.

TM4: TM4 suite of tools consists of four major applications, Microarray Data Manager (MADAM), TIGR_Spotfinder, Microarray Data Analysis System (MIDAS), and Multiexperiment Viewer (MeV), as well as a Minimal Information about a Microarray Experiment (MIAME)-compliant MySQL database, all of which are freely available to the scientific research community at http://www.tigr.org/software. The MADAM data entry interface provides access to data associated with a microarray study. It has a navigation panel on the left-hand side which leads users through the process of data entry during a microarray experiment. TIGR Spotfinder provides image processing with direct connections to the microarray database. MIDAS allows users to define data normalization and filtering protocol using a simple graphical scripting interface. The diagram on the left shows the steps in the analysis to be carried out; the panel on the right allows users to enter parameters for each stage in the analysis. MeV allows users to apply a number of sophisticated data mining tools to their array data and provides integrated graphical depictions of the results of the analyses conducted. Three of the TM4 applications, MADAM, MIDAS, and MeV, were developed in Java and have been tested on Microsoft® Windows, Linux, Unix, and MacOS X platforms; TIGR Spotfinder was written in C/C++ and runs only on Windows systems (Saeed et al., 2003).

GeneChip Operating Software (GCOS): GCOS is a proprietary software currently used to analyse Affymetrix Oligonucleotide microarray after hybridization of the probe with target samples since Microarray Suite (MAS) has been discontinued. Image of the slide is captured via a scanner and expression values are generated into a *.dat file (Data File). The software derives the *.cel file (Cell Intensity File) from a *.dat file and automatically creates it upon opening a *.dat file. It contains a single intensity value for each probe cell delineated by the grid (calculated by the cell analysis algorithm). Chip File *.chp, the output file generated from the analysis of a probe array contains qualitative and quantitative information for every probe set. Report File *.rpt is a text file summarizing data quality information for a single experiment and is generated from the analysis of output file (*.chp). There are also other output file involved in the use of GCOS such as *.cab (Cab File) which is a compressed file that is a backup copy of a process or publish database, project, sample, and/or experiment. *.txt and *.xls are standard format for text files and spreadsheet files and GCOS exports text in these file formats. The library files (probe information) *.cif, *.cdf, and *.psi contain information about the probe array design characteristics, probe utilization and content, and scanning and analysis parameters. These files are unique for each probe array type. Fluidics Files*.bin, *.mac. T he fluidics files contain information about the washing, staining, and/or hybridization steps for a particular array format.

DNA-Chip analyser (dChip): DNA-Chip Analyzer (dChip) is a software package for probe-level and high-level analysis of Affymetrix gene expression microarrays and SNP microarrays (Li and Wong, 2001a; Li and Wong, 2001b; Lin et al., 2004). However, gene expression or SNP data from other microarray platforms can also be analyzed by importing as external dataset. High-level analysis in dChip includes comparing samples, hierarchical clustering, Loss of Heterozygosity (LOH) and copy number analysis of SNP arrays. To use dChip, the user needs to provide Affymetrix array data files (in CEL or DAT format, or see public CEL files), and the CDF file (Chip Description File). Obtain the dChip and gene information files and CEL file if required into local directory and double click to start dChip program. Affy conversion tool will convert all CEL files in a directory from version 4 to version 3, while leaving the file name the same. You can check the CEL file change in size to confirm if conversion is done and also ensure that CEL files are not read-only. If conversion fails DAT files can also be read by dChip instead of CEL files.

To read in cDNA array data, an external data file with every two columns as the green and red channel intensities from one array (e.g. obtained from GenePix GPR file), is read into dChip by "Analysis/Get external file" before continuing with data analysis. dChip is a freeware, single executable program developed on Windows 2000 but preferring windows NT/XP computers with 512 Megabytes memory for maximal operation. dChip is written in Visual C++ 6.0 and uses Windows-specific functions for graphic tasks, and the source code is freely available for academic purposes.

Compilation of *P. falciparum* and microarray software resources: Computer brought an ease in handling microarray technology tasks at nearly all stages through the provision of abundant software resources. The inexhaustive list of *Plasmodium* and microarray software resources (Coppel et al., 2004; http://www.wellcome.ac.uk/assets /wtx033402.pdf) is depicted in Tables 1 - 5.

APPLICATION OF DNA MICROARRAYS TECHNOLOGY AND RELATED WORKS ON MALARIA

Gene expression profiling applications

Gene expression profiling applications includes: pathogenesis studies, pathogen's responses to drugs; pathogen's responses to host, host's responses to infection, host's responses to treatments and host response to vaccines. Expression levels for tens of thousands of genes can be simultaneously measured in a single hybridization experiment and are collectively called a "gene expression profile". In the gene expression profiling experiments, the

Table 1. *Plasmodium falciparum* general information sites.

Name	URL
PlasmoDB	http://www.PlasmoDB.org
P. falciparum GeneDB site	http://www.genedb.org/genedb/malaria/index.jsp
TIGR parasites database	http://www.tigr.org/tdb/parasites/
Sanger Institute parasite genomes project	http://www.sanger.ac.uk/Projects/Protozoa/
Sanger institute P. falciparum genome project	http://www.sanger.ac.uk/Projects/P_falciparum/
Malaria full length complementary DNA project	http://fullmal.ims.u-tokyo.ac.jp/
KEGG implementation of the P. falciparum metabolic pathways	http://www.genome.ad.jp/dbget- bin/www_bfind?p.falciparum
PlasmoCyc genome pathway database	http://plasmocyc.stanford.edu/
Malaria parasite metabolic pathways	http://sites.huji.ac.il/malaria/
WHO/TDR malaria database	http://www.wehi.edu.au/MalDB-www/who.html
NCBI malaria genetics and genomics	http://www.ncbi.nlm.nih.gov/projects/Malaria/
DeRisis laboratory malaria transcriptome database	http://malaria.ucsf.edu/index.php
Structural genomics of parasitic protozoa consortium	http://depts.washington.edu/sgpp

Table 2. Microarray information sites.

Name	URL
ArrayExpress repository for microarray data.	http://www.ebi.ac.uk/arrayexpress
Microarray gene expression data society homepage	http://www.mged.org,http://www.microarrays.org, http://microarrayworld.com, http://www.deathstarinc.com/science/biology/chips.html

Table 3. Spot information extraction.

Name	Description
GenePix (http://www.axon.com /gn_Genomics.html)	Spot extraction information from TIFF files
Microarray Suite (MAS 5.0)	Old and discontinued commercial software for Affymetrix oligonucleotide microarray used during image analysis and generation of expression values.
GeneChip Operating Software (GCOS)	Latest commercial software for Affymetrix oligonucleotide Microarray for image analysis and generation of expression values.

biological samples that the probes are designed to interrogate are RNA extracted from cells or tissues. These types of experiments answer the question of "what genes and how much of them are expressed in the biological sample?". The RNA molecules are first converted to cDNA by reverse transcription and labeled with a fluorescent dye. The expression level of a gene are measured as the light intensities emitted, after excitation with laser light, by the fluorescent dye attached to the cDNA that bound the homologous probes on the array (Chen, 2006).

Gene expression profiles of the host to pathogen may also be used in diagnosis for identifying possible pathogens. DNA microarray technology enables scientists to perform global survey of novel virulence factors, antimicrobial drug resistance genes, and potential vaccine targets by monitoring the transcription profiles of the pathogens in response to host environments.

Osamor (2009) reviewed recent studies that applied microarray approach to monitor gene expression of malaria pathogen, Plasmodium falciparum (put in italic) in the host cells for identification of potential vaccine candidates or drug targets. Daily et al. (2005) studied the gene expression profiles of P. falciparum that was isolated from blood samples of infected patients and compared them (Daily et al. 2007) with the in vitro profiles of a reference P. falciparum strain at the ring stage (Le Roch et al., 2003). A new family of hypothetical protein that may encode surface antigens was found to be over-expressed in the *in vivo* samples, making these potential candidates for vaccine development.

Gaur et al. (2006) identified new virulence genes by comparing gene expression profiles between two *P. falciparum* clones. The *P. falciparum* Dd2, a parasitic

Table 4. Normalisation and data analysis software.

Name	Description
TM4 software- (Saeed *et al.,* 2003; http://www.tigr.org/software).	TIGR suite of programs for data analysis and visualisation. In particular, MIDAS and MEV are useful for analyzing cDNA Microarray experiment
dChip software	A freeware that contains suite of program for that supports Affymetrix gene chip (oligonucleotide) microarray experiment.
GeneSpring – (http://www.silicongenetics.com).	commercial and comprehensive multipurpose data analysis/visualization software
Bioconductor (http://www.bioconductor.org and http://www.r-project.org).	Numerous advanced statistical analysis and data visualisation programs that run under the R-project language
SAM (Significant Analysis of Microarray) (http://www-stat.stanford.edu/~tibs/SAM)	Popular and easy to use Excel plug-in for statistical analysis of microarray data
MAANOVA (http://www.jax.org/staff/churchill/labsite/software/Jmaanova/index.html)	Advanced statistical analysis of cDNA microarray data
Resourcerer (http://pga.tigr.org/tigr-scripts/magic/r1.pl).	Identify common transcripts between different array platforms

Table 5. Functional tools applicable to microarray data.

Name	Description
EASE – (http://david.niaid.nih.gov/david/ease.htm).	Software for functional classification and analysis of gene lists.
GO (http://geneontology.org).	Biological/functional gene annotations.
EXPANDER (http://www.cs.tau.ac.il/~rshamir/expander/expander.html).	Software for functional/promoter analysis of clustered gene lists.
KEGG (http://www.genome.jp/kegg/).	Biological pathways and gene annotations.
GenMAPP- (http://www.genmapp.org).	Biological pathways and gene annotations.
Ingenuity (http://www.ingenuity.com).	Commercial, comprehensive hand-curated resource for gene function annotation and analysis of gene lists.
Biolayout from Enright and Ouzounis (2001)	A java-based general networks visualization tool with custom layout algorithm that preferentially places functionally similar nodes together so that biologically relevant functional clustering is more easily seen.
Pathway Tools	Software for studying molecular pathways from Karp et al. (2002).

clone which requires sialic acid residues on the erythrocyte surface for successful invasion, is capable of undergoing a genotypic switch to become a subclone Dd2 (NM), which can invade erythrocytes without the sialic acid residues. By comparing the expression profiles of these two parasitic clones, four novel genes were initially identified to be up-regulated in the sialic independent clone Dd2 (NM). Two of these genes, PfRH4 and PEBL, were confirmed by RT-PCR and the expression of PfRH4 at protein level was further confirmed to be only in Dd2 (NM) (Chen, 2006).

Genotyping applications

Genotyping applications includes pathogen identification, drug resistance survey, host susceptibility, pathogen cataloging and vaccine re-evaluation. In the genotyping experiments, the targets are DNA extracted from the biological samples and the probes are designed to survey the sequence variations in or among the samples. "Single nucleotide polymorphism (SNP) microarray" is an example of genotyping microarray. A variation of the SNP microarray is called "sequencing microarray" or "rese-

sequencing microarray" and can be used to re-sequence a specific region of a closely related genome, of which the sequences have to be decoded.

The most direct and perhaps also the most widely used applications of DNA microarray technology in infectious diseases fall in this category. DNA microarrays allow quick identification of the pathogens based on the unique sequence signature detectable by the large number of the probes on the array.

A good example is the identification of a new corona virus that caused the severe acute respiratory syndrome (SARS) epidemic outbreak in 2003. Before the outbreak, Wang et al. (2003) had devised a microarray intended for detecting the widest possible range of both known and unknown viruses. This viral microarray platform contained probes representing all the approximately 1,000 known virus sequences at the time from GenBank. In March 2003, DeRisi and Wang, then a postdoc in DeRisi's lab at the University of California, San Francisco, used this microarray to quickly identify the viral agent in SARS samples as a new type of corona virus.

Other microarray related works on malaria

Bozdech et al. (2003b) developed a software package, ArrayOligoSelector, to design an open reading frame (ORF)-specific DNA microarray using the publicly available P. falciparum genome sequence. Each gene was represented by one or more long 70 mer oligonucleotides selected on the basis of uniqueness within the genome, exclusion of low-complexity sequence, balanced base composition and proximity to the 3' end. A first-generation microarray representing approximately 6,000 ORFs of the P. falciparum genome was constructed. Arrays were washed and scanned with GenePix 4000B and GenePix Pro 3.2 software normalized the data. Cluster analysis was done using the CLUSTER and TREEVIEW software (Esien et al., 1998) to extensively characterize the gene-expression profile of the intra-erythrocytic trophozoite and schizont stages of P. falciparum. The results revealed extensive transcriptional regulation of genes specialized for processes specific to these two stages and may serve as the basis for future drug targets and vaccine development.

A good example of the ability of microarray analyses to monitor gene expression is provided by the study reported in Bozdech et al. (2003a). They reasoned that profiling transcript abundance throughout the erythrocyte phase of the lifecycle of the malaria parasite, P. falciparum, might identify a handful of genes that are induced at critical times and hence might be novel drug targets. The Affymetrix Custom Express program enables researchers to use any available sequence information to create a GeneChip microarray focused on their specific research needs. Using the Affymetrix CustomExpress Array program, Le Roch et al. (2003) created a customized GeneChip microarray, representing the complete genome

of P. falciparum, to better understand global gene activity during the different stages of the P. falciparum life cycle that transfers from mosquito to man and back to mosquito. Having developed a custom P. falciparum Gene-Chip microarray, the research team was able to examine the simultaneous activity of 95 percent of the 5300 genes present within the parasite. The first step was arranging genes based on their expression at different stages of the P. falciparum life cycle. They then identified groups of genes that were expressed at the same times and categorized them into 15 distinct 'clusters' using a robust k-means algorithm.

Llinas et al. (2006) have used a genome-wide approach to characterize transcriptional differences between strains of P. falciparum by characterizing the transcriptome of the 48h intraerythrocytic developmental cycle (IDC) for two strains, 3D7 and Dd2 and compared these results to HB3 strain obtained from Bozdech et al. (2003a). These three strains originate from geographically diverse locations and possess distinct drug sensitivity phenotypes. The majorities of these genes are uncharacterized and have no homology to other species. Strains 3D7 and Dd2, 60 and 58 DNA microarray hybridizations were performed, respectively. The 3D7 dataset comprised of 53 hourly time points with hours 3, 15, 27, 30, 32, 35 and 39 represented by more than one array hybridization. The Dd2 dataset were comprised of 50 time points with hours 4, 7, 12, 14, 20, 25, 28, 30 and 47 represented by more than one array hybridization. DNA microarrays were scanned using an Axon 4000B scanner and images analyzed using Axon GenePix software to obtain raw microarray data for various strains P. falciparum.

MICROARRAY TECHNOLOGY DEVELOPMENT EFFORTS IN AFRICA

Microarray technology is making an inroad into technological development of some African countries but is highly hampered by the dearth of infrastructural development and lack of technical human capital to carry out the required deliveries. It's justifiable to say that Africans are beginning to understand the strength of microarray technology and genomics in general and its ability to solve peculiar challenges of disease and food. Strong linkages between genomics research and national malarial control programs will facilitate the translation of research findings into intervention tools. This new techno-logy awareness is important for the communities in endemic countries to have total understanding of genomics research in their healthcare system (Alibu and Egwang, 2003).

A policy for developing biotechnology and genomics in a report entitled *Biotechnology platforms: strategic review and forecast* was launched in South Africa. The policy advocates the establishment of world-class genomics capacity, through the creation of a national facility and centres of excellence. The New Partnership for Africa's

Development (NEPAD) has likewise committed to developing regional capacity in science and technology, including genomics, using networks of centres of excellence. The African Biosciences Facility in South Africa is a concrete achievement in this effort to make possible Africa's active participation in the advances of genomics (WHO, 2005). Some other efforts include the formation of Nigeria Society of Bioinformatics and Computational Biology (NSBCB), African Society of Bioinformatics and Computational Biology (ASBCB). Organisation of genomics conferences has also improved microarray technology in Africa, for example International Workshop on Pattern Discovery in Biology (IWPDB 2005 and 2009) at Covenant University, Ota, Ogun State, Nigeria; Bioinformatics for Africa-Nairobi in 2007 and Mali in 2009 respectively. Also included is a WHO/TDR sponsored functional genomics workshop (African Center for Training in Functional Genomics- AFRO VECTGEN) in Mali.

It may be that such networks facilitate the development of more "networked" approaches to innovation including microarray technology and the provision of solution to the problem of malaria. To the best of our knowledge, South Africa seem to be ahead with an ACGT Microarray facility which exist at University of Pretoria and are involved in training and carrying out microarray study in organisms other than *Plasmodium falciparum*. A part WHO/TDR initiative, African Network for Drugs and Diagnostics Innovation (ANDI), was launched at Abuja, Nigeria in 2008 (http://meeting.tropika.net/andi/) to promote & sustain African-led R&D innovation through the discovery, development and delivery of affordable new tools including those based on traditional medicines. Malaria treatment discovery is one of the major challenges facing this initiative.

Conclusion

It is obvious that using microarray technology affords great ease of analyzing the expression of batteries of microbial and malaria parasite genes at different phases of infection. Our ability to define virulence determinants and to understand how *P. falciparum* adapt and excel in their environments in mosquito and human by evading host defense mechanisms provides great hope to combating this age-long disease in Africa. Both experimental and computational analysis of the sequence and micro-array data of the malarial parasite is released freely into public databases as soon as they are produced. Many Africans both at home and in the diaspora in spite of their challenges are increasingly improving themselves with this technology by getting involved in some Malaria Genome Projects.

In silico antimalarial drug researchers (Osamor, 2009) are currently extracting vast amount of information from knowledge base available in microarray data for malaria treatment discovery in Africa. The time has come to concentrate on microarray analysis of the human malaria parasite with a view of eradicating the parasite in Africa and ultimately provide a home-grown solution to African problems through improved drug target and design strategies. ANDI is poised to sustain this African-led R&D innovation involving the development and use of new tools such as microarray technology.

REFERENCES

Affymetrix Inc. (2001). GeneChip Expression Analysis Algorithm Tutorial.

Alibu VP, Ewing TG (2003). Genomics Research and Malaria Control: Great Expectations. Plos Biol. 1(2) .

Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL (2003a). The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum Public Library of Science Biology, 1, E5.

Bozdech Z, Zhu J, Joachimiak PM, Cohen FE, Pulliam B, DeRisi JL (2003b). Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. Genome Biology, 4:R9. Available online at: http://genomebiology.com/2003/4/2/R9

Chen T (2006). DNA Microarrays - An Armory for Combating Infectious Diseases in the New Century. Infectious Disorders - Drug Targets 6(3): 263-279.

Coppel RL, Roos DS, Bozdech Z (2004). The genomics of malaria infection. Trends Parasitol. 20 (12).

Daily JP, Scanfeld D, Pochet N, Le Roch K, Plouffe D, Kamal M, Sarr O, Mboup S, Ndir O, Wypij D, Levasseur K, Thomas E, Tamayo P, Dong C, Zhou Y, Lander ES, Ndiaye D, Wirth D, Winzeler EA, Mesirov JP, Regev A (2007). Distinct physiological states of Plasmodium falciparum in malaria-infected patients. Nature 450: 1091-1095.

Daily JP, Le Roch KG, Sarr O, Ndiaye D, Lukens A, Zhou Y, Ndir O, Mboup S, Sultan A, Winzeler EA, Wirth DFJ (2005). In Vivo Transcriptome of Plasmodium falciparum Reveals Overexpression of Transcripts That Encode Surface Proteins. J. Infect. Dis. 191, 1196: 1203.

Draghici S (2003). Data Analysis Tools for DNA Microarrays. London: Chapman and Hall.

Esien MB, Spellman PT, Brown PO, Botstein D (1998). Cluster Analysis and Display of Genome-Wide Expression Patterns. Proc. Natl. Acad. Sci. USA 95: 14863-14868.

Enright AJ, Ouzounis CA (2001). Biolayout – an automatic graph layout algorithm for similarity visualization. Bioinformatics 17(9): 853-854.

Gaur D, Furuya T, Mu J, Jiang LB, Su XZ, Miller LH (2006). Upregulation of expression of the reticulocyte homology gene 4 in the *Plasmodium falciparum* clone Dd2 is associated with a switch in the erythrocyte invasion pathway. Mol. Biochem. Parasitol. 145: 205.

Gibson G (2003). Microarray Analysis. PLoS Biol 1(1): e15 doi:10.1371/journal.pbio.0000015

Karp P, Paley S, Romero P (2002). The Pathway Software. Bioinformatics 18(1).

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De la Vega P, Holder AA, Batalov S, Carucci DJ, Winzeler EA (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Sci. 301(5639): 1503-1508.

Li C, Wong WH (2001a). Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection, Proc. Natl. Acad. Sci. 98: 31-36.

Li C, Wong WH (2001b). Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application, Genome Biol. 2(8): research0032.1-0032.11.

Lin M, Wei L-J, Sellers WR, Lieberfarb M, Wong WH, Li C (2004). dChipSNP: Significance Curve and Clustering of SNP-Array-Based Loss-of-Heterozygosity Data. Bioinformatics 20: 1233-1240.

Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL (2006). Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. Nucleic Acids Res. 34(4):1166-1173.

Miller LH, Baruch DI, Marsh K, Doumbo OK (2002). The Pathogenic

Basis of Malaria. Nature 415: 673-679.

Osamor VC (2009). Simultaneous and Single Gene Expression Computational Analysis for Malaria Treatment Discovery. Ph.D Thesis, Covenant University, Ota, Nigeria.

Quackenbush J (2005). Using DNA Microarrays to Assay Gene Expression. In Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Third Edition, Eds Baxevanis A.D. and Ouellette B.F. F., John Wiley & Sons.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J (2003). TM4: A Free, Open-Source System for Microarray Data Management and Analysis. BioTechniques 34: 374-378.

Schena M, Shalon D, Davis RW, Brown PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. Oct 20; 270 (5235): 467-470.

Tatusov RL, Koonin EV, Lipman DJ (1997). A genomic perspective on protein families. Sci. 278: 631.

Vinciotti V, Khanin R, D'alimonte D, Liu X, Cattini N, Hotchkiss G, Bucca G, De Jesus O, Rasaiyaah` J, Smith CP, Kellam P, Wit E (2005). An experimental evaluation of a loop versus a reference design for two-channel microarrays. Bioinformatics 21(4):492-501. [Available at http://bioinformatics.oxfordjournals.org/cgi/content/ full/21/4/492].

Wang D, Urisman A, Liu YT, Springer M, Ksiazek TG, Erdman DD, Mardis ER, Hickenbotham M, Magrini V, Eldred J, Latreille JP, Wilson RK, Ganem D, DeRisi JL (2003). Microarray-based detection and genotyping of viral pathogens, Proc. Natl. Acad. Sci. U. S. A. 99: 15687-15692.

WHO (2005). Genetics, genomics and the patenting of DNA : review of potential implications for health in developing countries. World Health Organisation, Geneva.

Wosik E (2006). Oligonucleotide Array. http://cnx.org/content/m12385 /latest/ 31 Mar 2006.