Patterns of gene expression that characterize outcomes of *Plasmodium falciparum* infection

C.S. Boutlis[1], H.K. Dressman[2], E. Tjitra[3], H. Maniboey[4], J.R. Nevins[2], M.C. Levesque[2], J.B. Weinberg[2,5], N.M. Anstey[1]

1. Menzies School of Health Research (MSHR), Darwin, Australia. 2. Duke University Medical Center, Durham, NC, USA. 3. National Institute of Health Research and Development (NIHRD), Jakarta, Indonesia. 4. NIHRD & MSHR, Jayapura, Indonesia. 5. Veterans Affairs Medical Center, Durham, NC, USA.

Background

- Pathology due to *Plasmodium falciparum* is largely mediated by circulating immune cells.
- Understanding the molecular basis of malarial immunopathology has been constrained by having to study cellular responses in isolation.
- High throughput gene technologies provide opportunities for examining malaria disease phenotypes in unprecedented detail.
- We have utilised microarray technology to examine the genome-wide expression profile in mRNA from peripheral blood mononuclear cells (PBMC) in subjects with clinical malaria.

Methods

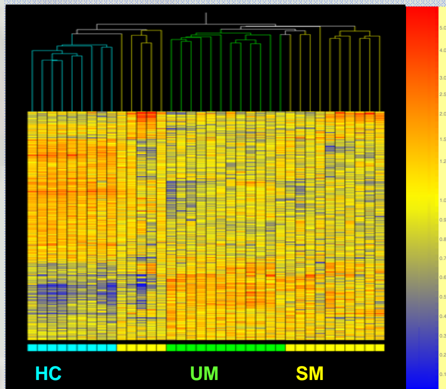
- Adults with uncomplicated (UM; *n*=12) or severe (SM; *n*=15) malaria and healthy controls (HC; *n*=9) were enrolled between 1998 and 2000 in Jayapura, Papua Province, Indonesia.
- RNA was extracted from PBMCs using Qiagen RNeasy kits and checked by a Bioanalyzer.
- All experiments utilised Affymetrix U133A gene-chips according to standard protocols; these gene-chips contain 22,283 probe sets ("genes").
- Arrays were scaled to a target intensity of 500 fluorescence units (FU) and data quality was checked using the R Program's "affy", "simpleaffy" and "affyPLM" packages[1].
- Data were normalised around a median value of 1 for each probeset in GeneSpring software.
- Data were filtered to only retain those with a high percent of "present" or "marginal" calls (≥ 50%) and a scaled intensity of ≥ 100 FU.

➢ After filtering, data were retained for 9,359 of the 22,283 genes for use in clustering and biological pathway mapping analyses.

- "Changed genes" were defined as those exhibiting significantly different levels of expression between the 3 study conditions (using the Kruskal-Wallis [groupwise] and Mann Whitney U tests [pairwise]) controlling for a 5% false discovery rate (FDR).
- Raw data from 22,283 genes were normalised using the robust multi-array average algorithm in the R Program for binary regression modelling[1].

Results

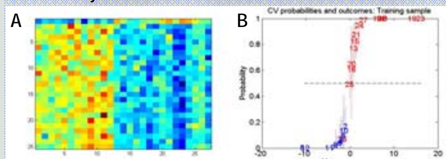
- Of the 9,359 filtered genes, there were 4,164 changed genes showing significantly different expression among the three study conditions.
- Unsupervised clustering of these changed genes showed a high level of segregation of subjects according to their study condition, with the greatest differences between HC and UM.



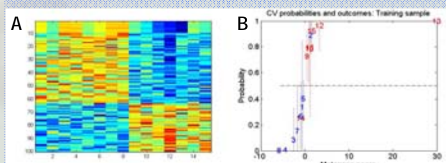
Rows are the 4,164 changed genes and columns are the 36 subjects. Median expression for each gene is normalised around 1 (yellow): increased expression = red, decreased expression = blue. For both genes and subjects, Pearson correlation was used as the joining metric.

Metagenes

- To further capture the clinical information in the expression data, supervised strategies were employed using the clinical phenotypes to drive a binary regression analysis that distinguishes between the study conditions[2].
- An aggregate pattern ("metagene") of 25 genes readily distinguished subjects with UM & SM.
- Similarly a metagene of 100 genes distinguished SM subjects with and without renal failure.



A. *Rows are 25 highly discriminatory individual genes, columns are the 12 UM subjects (left) and 15 SM subjects (right). Blue represents the lowest level of expression, followed by green, orange, yellow and red (highest).*
B. *One at a time cross-validation (C.V.) of classification probabilities from the regression analysis. Horizontal axis shows estimates of the overall score in the regression. Vertical axis are estimated classification probabilities (90% probability intervals marked as dashed lines). Samples from UM subjects are blue, and SM subjects are red.*



A. *A metagene comprising 100 genes trained to distinguish 8 SM subjects without renal failure (left) from 7 SM subjects with renal failure (RF; right).*
B. *C.V. probabilities of SM subjects without RF (blue) or with RF (red).*

Biological Pathway Mapping

- Changed genes from each of three pairwise comparisons between 2 study conditions were mapped to biological pathways curated by the Gene Ontology (GO) consortium using EASE[3].
- The reference set was the 9,359 filtered genes.
- There were 4,138 changed genes showing significantly different expression between HC and UM, and higher than expected proportions of changed genes were found to be associated with 53 GO terms (below).

Type	GO terms – predominance of down-regulated genes	Ch	M	%Ch	Up	%Up	EASE rank	EASE score
C	cytosolic ribosome (sensu Eukaryota)	27	29	93	1	4	4	0
C	ribosome	74	104	71	6	8	5	0
C	*cytosolic large ribosomal subunit (sensu Eukaryota)*	18	18	100	1	6	8	0
F	*structural constituent of ribosome*	74	107	69	4	5	9	0.0001
F	*RNA binding*	197	326	60	30	15	10	0.0002
C	large ribosomal subunit	21	25	84	2	10	16	0.0022
P	*cellular defense response*	40	56	71	19	48	18	0.0023
P	cell activation	35	49	71	12	34	23	0.0048
P	immune cell activation	35	49	71	12	34	24	0.0048
F	structural molecule activity	135	230	59	29	21	29	0.0087
C	ribonucleoprotein complex	122	207	59	10	8	31	0.0104
C	intracellular non-membrane-bound organelle	289	523	55	81	28	33	0.0125
C	non-membrane-bound organelle	289	523	55	81	28	34	0.0125
P	*T-cell activation*	17	21	81	2	12	35	0.0136
P	lymphocyte activation	30	43	70	8	27	38	0.0161
F	protein binding	594	1115	53	263	44	39	0.0186
P	*protein biosynthesis*	162	287	56	41	25	41	0.0215
P	macromolecule biosynthesis	176	315	56	49	28	45	0.0266
P	cellular biosynthesis	257	472	54	93	36	49	0.0336
P	*regulation of lymphocyte activation*	11	13	85	3	27	53	0.0469

Type	GO terms – predominance of up-regulated genes	Ch	M	%Ch	Up	%Up	EASE rank	EASE score
P	defense response	261	422	62	157	60	1	0
P	immune response	243	393	62	149	61	2	0
P	response to biotic stimulus	296	494	60	176	59	3	0
P	response to external biotic stimulus	157	249	63	106	68	6	0
P	response to pest, pathogen or parasite	155	246	63	105	68	7	0
F	receptor activity	254	433	59	126	50	11	0.0003
P	response to wounding	112	177	63	78	70	12	0.0004
F	signal transducer activity	430	768	56	225	52	13	0.0006
P	organismal physiological process	337	599	56	198	59	14	0.0010
F	cytokine binding	27	34	79	16	59	15	0.0016
P	response to external stimulus	235	411	57	148	63	17	0.0023
P	*chemotaxis*	35	48	73	26	74	19	0.0028
P	taxis	35	48	73	26	74	20	0.0028
P	*inflammatory response*	64	98	65	53	83	21	0.0030
P	*positive regulation of I-kappaB kinase/ NF-kappaB cascade*	38	54	70	24	63	22	0.0046
P	response to stimulus	428	785	55	230	54	25	0.0050
P	regulation of I-kappaB kinase/NF-kappaB cascade	39	56	70	24	62	26	0.0053
P	response to chemical substance	59	91	65	39	66	27	0.0058
P	positive regulation of signal transduction	41	60	68	25	61	28	0.0069
P	cell communication	611	1147	53	317	52	30	0.0097
P	signal transduction	545	1020	53	276	51	32	0.0116
P	cell-cell signaling	73	119	61	45	62	36	0.0138
P	I-kappaB kinase/NF-kappaB cascade	45	69	65	28	62	37	0.0150
F	*hematopoietin/interferon-class (D200-domain) cytokine receptor activity*	18	23	78	11	61	40	0.0188
C	vacuole	45	70	64	39	87	42	0.0232
P	regulation of signal transduction	58	94	62	32	55	43	0.0251
P	response to stress	265	485	55	147	55	44	0.0256
C	*lysosome*	43	67	64	37	86	46	0.0278
C	lytic vacuole	43	67	64	37	86	47	0.0278
P	*humoral immune response*	50	80	63	33	66	48	0.0289

Legend: *Type – C=Cellular component, P=Biological process; F=Molecular function; Ch=no. of changed genes and M=no. of measure genes for term; %Ch=percent changed; Up and %Up=no. and percent up-regulated in UM relative to HC; EASE rank=rank from most statistically significant (no. 1); EASE score=P value using Fisher's exact test with jackknifing procedure.* ***Note:*** *GO terms in italics are those with no lower ranking child terms in the list and could be regarded as the most informative. The background rate of changed genes (i.e., %Ch overall) was 4,138 / 9,359 = 44%.*

Conclusions

- Characteristic patterns of PBMC gene expression readily distinguished the different clinical presentations of malaria.
- Changes in expression were more pronounced in UM rather than SM relative to HC; this may be indicative of a more effective immunity in UM.
- The prominent down-regulation of protein biosynthesis and ribosome pathways in UM (and SM) was surprising and to our knowledge has no precedent in human infections.
- Our ability to define metagenes characterising different malaria disease states and RF in SM subjects provides proof-of-concept that gene expression profiles may one day be used to predict &/or target therapy toward complications.

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

BAKTI HUSADA

mshr menzies school of health research

References: 1. R foundation for Statistical Computing, Vienna, Austria
2. Pittman J et al, PNAS, 2004 Jun 1:101(22):8431-6.
3. Hosack DA et al, Genome Biology, 2003, 4:R70.