The early whole-blood transcriptional signature of dengue and features associated with progression to dengue shock syndrome in Vietnamese children and young adults

Long Truong Hoang, David J. Lynn, Matt Henn, Bruce W. Birren, Niall J. Lennon, Phuong Thi Le, Kien Thi Hue Duong, Tham Thi Hong Nguyen, Lanh Ngoc Mai, Jeremy J. Farrar, Martin L. Hibberd, Cameron P. Simmons

1 Oxford University Clinical Research Unit, Hospital for Tropical Diseases, 190 Ben Ham Tu, District 5, Ho Chi Minh City, VietNam
2 Teagasc, Animal Bioscience Department, Dunsany, Co. Meath, Ireland.
3 Broad Institute, 7 Cambridge Centre, Cambridge, MA 02142
4 Dong Thap Hospital, 144 Mai Van Khai St, Hamlet 3, My Tan village, Cao Lanh city, Dong Thap province, Viet Nam
5 Genome Institute of Singapore, 60 Biopolis Street, #02-01, Genome, Singapore 138672

Footnotes
*Corresponding authors:
Cameron Simmons
Email: csimmons@oucru.org
The authors have no financial or commercial conflicts of interest.
Abstract

Dengue is a pan-tropic public health problem. In children, dengue shock syndrome (DSS) is the most common life-threatening complication. Predicting patients who may develop DSS may improve triage and treatment. To this end, we conducted a nested case-control comparison of the early host-transcriptional features in 24 DSS patients and 56 sex-, age-, virus serotype-matched uncomplicated dengue patients. In the first instance we defined the “early dengue” profile. The transcriptional signature in acute versus convalescent samples (≤72hrs post illness-onset) was defined by an over-abundance of interferon-inducible transcripts (31% of the 551 over-abundant transcripts), and canonical gene ontology terms that included response to virus, immune response, innate immune response and inflammatory response. Pathway and network analysis identified STAT1, STAT2, STAT3, IRF7, IRF9, IRF1, CEBPB and SP1 as key transcriptional factors mediating the early response. Strikingly, the only differences in the transcriptional signatures of early DSS and uncomplicated dengue cases was the greater abundance of several neutrophil-associated transcripts in patients who progressed to DSS, a finding supported by higher plasma concentrations of several canonical proteins associated with neutrophil-degranulation (BPI, ELA2, DEF1A). Elevated levels of neutrophil-associated transcripts were independent of the neutrophil count and also the genotype of the infecting virus, as genome-length sequences of DENV-1 (n=15) and DENV-2 (n=3) viruses sampled from DSS patients were phylogenetically indistinguishable from those sampled from uncomplicated dengue patients (32 DENV-1 and 9 DENV-2). Collectively, these data suggest an hitherto unrecognised association between neutrophil activation, pathogenesis and the development of DSS and point to future strategies for guiding prognosis.
Author summary

Dengue is a public health problem for tropical and sub-tropical countries. Predicting patients who may develop DSS may improve treatment. We conducted a study to compare early gene expression by cells in the blood of children with DSS and with mild dengue. The patients were matched by sex, age and the type of infecting viruses. We found that the “early dengue” gene expression profile was characterised by an innate anti-viral response that was regulated by several proteins that regulate gene expression. Strikingly, the only difference in early gene expression between DSS patients and mild dengue patients was the greater abundance of genes that were related to neutrophil degranulation in DSS patients. This finding was supported by higher amount of several proteins in the plasma of DSS patients that are encoded by these genes. The increased expression of these genes was independent of the number of neutrophils in the blood and also the genotype of the infecting virus, as genome-length sequences of viruses sampled from DSS patients were similar from those sampled from mild dengue patients. Collectively, these data suggest an hitherto unrecognised association between neutrophil activation, pathogenesis and the development of DSS and point to future strategies for guiding prognosis.
Introduction

Dengue is the most significant mosquito-borne viral infection of humans. The dengue pandemic has spread to the extent that between 70-500 million infections occur each year in over 100 countries resulting in ~40 million clinically apparent cases and ~20,000 deaths (32). There are no licensed vaccines for prevention of dengue and treatment of severe cases is limited to supportive care.

Dengue is an acute, self-limiting systemic viral infection that affects mostly children and young adults in endemic settings. Dengue manifestations range from non-specific fever to a more severe syndrome commonly characterised by increased vascular permeability, thrombocytopenia and a bleeding diathesis. In severe cases, the increased vascular permeability results in circulatory compromise and the patient may develop potentially life-threatening dengue shock syndrome (DSS). A characteristic feature of DSS is that it manifests clinically between day 4-6 of illness, a time when the viraemia is in steep decline and the host immune response is well established. The timing of these events suggests the host pro-inflammatory response, rather than direct viral-mediated effects, mediates the vascular permeability syndrome leading to DSS. Fortunately, these physiological derangements are transient and most patients recover fully if supported with parenteral fluid therapy during the period of maximal vascular leakage. Host features implicated in susceptibility to severe disease include genetic variation (8), previous DENV infection history (7, 13, 34, 40), individual propensity for capillary leakage that itself is influenced by age and sex (5). Viral genetic traits might also be important, with some lineages of DENV being virologically (9) and epidemiologically fitter (28) than others and more frequently associated with severe disease.

There are no animal models that fully mimic the dengue vascular leakage syndrome and therefore clinical studies are of particular importance. Previous studies of the whole-blood host transcriptional response in Vietnamese, Singaporean and Thai children and adults have been relatively small in size and used non-uniform sampling times. Collectively, these studies have suggested that during the febrile phase transcripts from interferon-stimulated genes are highly prominent (11, 26, 37) but that this signature wanes rapidly with the resolution of infection. However, previous
studies have not fully explored the interconnectedness of elements of the transcriptional signature in a way that could yield insights into critical gene regulation pathways and transcriptional factors. Moreover, previous studies have not examined whether the early (≤72hrs of illness history) transcriptional signature of dengue is distinct in patients who subsequently develop clinically important complications, e.g. DSS. This is important as the identification of prognostic markers of severe disease could allow for improved patient triage and management, or in the future, rational treatment with anti-viral or immune-modulating therapies.

To this end, the aim of this study was firstly to provide a definitive profile of the early whole-blood transcriptional signature of dengue in a large number of clinically and virologically well characterised Vietnamese children and young adults with severe and mild dengue. Secondly, we sought to identify features of the whole-blood transcriptome that were associated with progression to DSS. Intriguingly, our findings identified several neutrophil-associated transcripts as more abundant in patients who progressed to DSS, suggesting a hitherto unrecognised association between neutrophil activation, pathogenesis and the development of DSS.
Materials and Methods

Study enrolment and investigations The study protocol was approved by the Scientific and Ethical committee of Dong Thap Hospital and the Oxford Tropical Research Ethical committee. Clinical and haematological assessments were performed daily. Patients underwent screening in the outpatient clinic of Dong Thap Hospital if there was a clinical suspicion of dengue and illness duration of less than 72hrs. Plasma from screened patients was tested with an NS1 antigen ELISA test (Biorad, California, USA). Patients that were NS1 ELISA positive were admitted for further observation and participation in the study protocol if written informed consent was obtained from the patients (patients >14yrs) or from the patient’s parents/guardians (patients ≤14yrs). Serial blood samples for isolation of RNA, DNA and plasma were collected at the time of enrolment (study day 1), on study day 3 and at the time of discharge. Haematological data were recorded daily and all patients received an ultrasound within 24hrs of defervesence. From June 2006 to Dec-2007, 450 patients were recruited of whom 35 patients developed DSS according to WHO criteria (pulse pressure ≤ 20 mmHg with poor peripheral perfusion and rapid, weak pulse). The remaining 415 patients were defined as having uncomplicated dengue without evidence of cardiovascular compromise. In line with recently revised WHO guidelines (31), we used the term “uncomplicated dengue” to refer to these hospitalised patients as they did not require any significant clinical interventions and were managed throughout on the general infectious diseases ward. At the acute phase (day -3 and day -2 relative to defervescence), 24 samples from DSS patients and 56 samples from UC dengue patients were used for expression microarray. For controls, 18 and 16 autologous follow-up samples were collected at 2 weeks after discharge from DSS patients and UC dengue, respectively.

Patient diagnosis

Serological investigations (IgM and IgG capture ELISAs) were performed using paired plasma samples using methods described previously (15). A serological determination of dengue was made if IgM seroconversion occurred between the acute and early convalescent plasma sample. Serology was interpreted as suggestive of secondary infection if DENV-reactive IgG was detected in the capture ELISA in the first week of illness. Plasma concentrations of DENV NS1 concentrations were measured by capture ELISA (Biorad, California, USA) using recombinant DENV-1
NS1 protein (Hawaii Biotech, Hawaii, USA) as a standard. DENV serotype-specific quantitative RT-PCR was performed to determine viral serotype and viremia in the enrolment plasma samples collected from each patient using an established method (37).

**DENV genome sequencing**

Viral genomes were sequenced using the Broad Institute’s capillary sequencing (Applied Biosystems) directed amplification viral sequencing pipeline (see http://www.broadinstitute.org/annotation/viral/Dengue). This sequencing effort was part of the Broad Institute’s Genome Resources in Dengue Consortium (GRID) project. Viral RNA was isolated from diagnostic plasma samples (QIAmp viral RNA mini kit, Qiagen) and the RNA genome reverse transcribed to cDNA with superscript III reverse transcriptase (Invitrogen), random hexamers (Roche) and a specific oligonucleotide targeting the 3’ end of the target genome sequences. cDNA was then amplified using a high fidelity DNA polymerase (pfu Ultra II, Stratagene) and a pool of specific primers to produce 14 overlapping amplicons of 1.5 to 2kb in size for a physical coverage of 2X. Amplicons were then sequenced in the forward and reverse direction using primer panels consisting of 96 specific primer pairs, tailed with M13 forward and reverse primer sequence, that produce 500-700bp amplicons from the target viral genome. Amplicons were then sequenced in the forward and reverse direction using M13 primer. Total coverage delivered post amplification and sequencing was ~ 8-fold. Resulting sequence reads were assembled de novo and annotated sing the Broad Institute’s in-house viral assembly and annotation algorithms. All genome sequences newly determined here have been deposited in GenBank and assigned accession numbers (Table S1).

**Gene expression microarray**

We used one-colour array technology on the Illumina platform (Illumina Inc, San Diego, USA) for gene expression microarray. In brief, whole-blood (2.5ml) was collected directly into PAXgene RNA tubes (Qiagen, Sussex, UK). RNA extraction was performed using Paxgene RNA kits (Qiagen). Biotinylated amplified cRNA was generated by in vitro transcription (IVT) technology using Illumina® TotalPrep™ RNA Amplification Kit (Ambion, Inc., Austin, TX) according to the manufacturer's instructions. After purification, 850ng of cRNA was hybridized to an Illumina
HumanRef-8 V2BeadChip (containing probes to 23,961 RefSeq gene sequences) at 55 °C for 18 hours following the manufacturer’s instructions (Illumina, Inc., San Diego, CA). This was followed by washing, blocking and streptavidin-Cy3 staining steps. Finally, the chip was scanned with an Illumina Bead Array Reader confocal scanner.

Normalisation of array data and analysis
Standard normalisation procedures in Genespring were used (ver.10, Agilent, Santa Clara, CA, USA). In brief, array normalisation was performed by dividing the mean of array intensity value to the 75th percentile value of all arrays. Gene normalisation was performed by dividing the mean value of the gene in each array to the 75th percentile value of the gene in all the arrays. Normalised data was then filtered based on expression data in which only transcripts with detection confidence of greater than 0.999 in at least one out of 227 samples (total samples arrayed) under analysis were used. The detection P value was calculated by Beadstudio software (Illumina). After normalization, 9,780 genes were available for further analyses. Significance Analysis of Microarray (SAM) was used to detect transcripts that were relatively more or less abundant in one group of samples. Significant genes were those with False Discovery rate of less than 5 percent and fold difference between 2 groups of at least 2 fold. Normalised intensity data and chip image files have been deposited into GEO (http://www.ncbi.nlm.nih.gov/projects/geo). Network analysis of the differentially expressed genes between acute DSS and acute uncomplicated dengue was performed using Ingenuity Pathway Analysis (IPA) (www.ingenuity.com).

Pathway & Gene Ontology analysis of DE genes using InnateDB
Illumina Probe IDs were mapped to NCBI Entrez Gene IDs and these were used to cross-reference and upload differentially expressed (DE) genes to InnateDB (www.innatedb.com) (27) along with associated gene expression data. A list of pathways mapping to the uploaded genes was returned and pathway analysis was undertaken to determine which pathways were significantly over-represented in the up- and down-regulated gene datasets using the Hypergeometric test. InnateDB simultaneously tests for over-representation of DE genes in more than 3,000 pathways from the KEGG (23), PID (http://pid.nci.nih.gov), INOH (http://www.inoh.org/), NetPath (http://www.netpath.org) and Reactome databases (21). The Benjamini and Hochberg (BH) FDR correction (3) was applied to correct for multiple testing.
Similarly, the InnateDB Gene Ontology analysis tool was used to identify Gene Ontology terms (2) that were significantly associated with DE genes.

**InnateDB molecular interaction network analysis**

InnateDB pathway analysis can be a powerful approach to provide insight in which annotated pathways are most significantly associated with DE genes. Network analysis, analysing the molecular interactions between DE genes and their encoded products, however, provides the opportunity to move beyond investigating signalling pathways as relatively simple linear cascades. Network analysis allows one to potentially identify novel relationships between DE genes and their regulators and to uncover as yet unknown signalling cascades or pathways, functionally relevant sub-networks and the central molecules, or hubs, of these networks.

InnateDB is one of the most comprehensive databases of all human and mouse experimentally-supported molecular interactions but also specifically includes annotation on more than 11,500 manually curated human and mouse innate immunity relevant interactions, many of which are not present in any other database. InnateDB was used to construct molecular interaction networks consisting of interactions between DE nodes and their interacting partners with the associated gene expression data overlaid as node attributes. Networks were analysed using the cytoHubba plugin (25) (http://hub.iis.sinica.edu.tw/cytoHubba/) for Cytoscape 2.6.3 (35) to investigate a variety of properties of a network including the identification of network hubs and bottlenecks which may represent the key regulatory nodes in the network. The jActiveModules plugin (19) was also used to identify high-scoring differentially expressed sub-networks.

**Transcriptional network analysis**

The networks discussed above mainly consist of protein-protein direct physical or biochemical interactions, with a relatively small number of experimentally-verified protein-DNA (Transcription factor-gene) interactions. This is because the production of high-throughput experimental data on protein-DNA interactions has lagged behind the data for protein-protein interactions. Although InnateDB contains ~13,000 protein-DNA interactions, this number represents less than 3% of all interactions and is a small sample of likely transcription factor-gene interactions. To overcome this
limitation, InnateDB now also contains computationally predicted transcription factor interactions using transcription factor binding site analysis data extracted from the CisRED database (www.cisred.org) (Robertson et al, 2006). InnateDB was used to construct a network of all predicted transcription factor (TF) interactions with DE genes (TF_network). Each association between a transcription factor and a gene was inferred as a protein-DNA interaction and a gene regulatory network was constructed. Given the large number of false positives associated with predicting transcription factor binding sites and therefore transcription factor-gene interactions, identifying the key regulators by ranked degree alone is not sufficient (i.e. a node may have a large number of predicted interactions with DE genes but this may expected by chance if that node is annotated to interact with a large number of genes in InnateDB). To overcome this issue, a right-sided Fisher's exact test (implemented using the Text::NSP Perl module http://search.cpan.org/dist/Text-NSP/) was used to identify nodes which interact with significantly more DE genes than expected by chance.

**ELISA assays**

Elastase 2 (ELA2) in patient plasma samples was measured using a PMN-Elastase ELISA kit (ImmunDiagnostics, Netherlands). Plasma Defensin 1 alpha (DEF1A), Bactericidal/permeability-increasing protein (BPI) and Myeloperoxidase (MPO) were measured using capture ELISA assays (Hbt, Hycult Biotechnology, Netherland). The limit of detection was 120pg/ml for ELA2, 50pg/ml for DEF1A, 250pg/ml for BPI and 0.4ng/ml for MPO. Plasma albumin concentrations were measured by capture ELISA. Briefly, ELISA plates (Maxisorp, Nunc, Denmark) were coated with anti albumin pAb (polyclonal rabbit antihuman albumin( Dako Co.) overnight and then blocked with 2% skim milk (Merck) in PBS-T 0.05% for 1h at 37°C. Next, samples and standards were added into the wells followed by addition of diluted biotinylated anti-albumin antibody solution and incubated for 1hr at 37°C. Diluted streptavidine HRP (P0397, Dako Co. Japan) was added and incubated for 1hr at 37°C before the OPD substrate was added (Sigma, P9187, USA). After the reactions was stopped by diluted sulphuric acid (10%), the OD value was measured using plate reader (Biorad, California, USA) and the result was analysed by Microplate software (Biorad, California, USA).

**Statistics**
All statistical analysis was performed using Intercooled STATA version 9.2 (StataCorp, TX). Significance was assigned at $P<0.05$ for all parameters and were two-sided unless otherwise indicated. Categorical variables between groups were compared by Fisher’s exact test. The $t$-test was used for continuous variables. Classification analysis by decision tree forest model (6) was performed to examine whether the set of differentially expressed genes between acute DSS and UC dengue are good predictors of disease outcomes. In brief, expression intensities of the 21 differentially expressed genes were used as “predictor” and the outcomes are used as “target”. This model generates an ensemble of 200 trees using randomisation of data and predictors. The tree with the most likely predicted value was chosen.

Results

Patient population
Between June 2006 and Dec-2007, 450 patients with dengue and less than 72 hrs of illness were enrolled into this prospective, hospital-based study. Thirty-five patients subsequently progressed to DSS as defined by WHO criteria (pulse pressure $\leq 20$ mmHg with poor peripheral perfusion and rapid, weak pulse) (31). The remaining 415 patients were defined as having “dengue” by the recently revised WHO criteria (31) as they did not require any clinical interventions and were managed throughout the general infectious disease ward. Samples from twenty-four DSS patients were available for gene expression microarray analysis. For each DSS patient, 1-3 patients with uncomplicated dengue and matched by sex, infecting virus serotype, age (within 2 years) and day of illness were selected as controls for virological investigations and host gene expression profiling (n=56 in total). The demographic and virological characteristics of the DSS and matched uncomplicated dengue patients is summarised in Table 1. The median day of illness was 3 days (range 2-5) and the median time to defervescence was 3 days (range 2-3) (Table 1). In patients who developed DSS, the median time from enrolment to shock was 2 days (range 1-4).

Virological comparisons between patients with DSS versus uncomplicated dengue
Patients with DSS had significantly higher plasma NS1 concentrations at the time of enrolment than matched uncomplicated dengue controls (Table 1). Plasma viraemia levels were not significantly different between the two groups (Table 1). To
understand if virus genetic traits were associated with clinical outcome, we attempted to derive the consensus genome sequence of each virus directly from plasma samples collected at study enrolment. From the 24 DSS patients, we derived consensus genome sequences (nucleotides 30-10649) of 15 DENV-1 and 3 DENV-2 viruses. These sequences were compared phylogenetically to consensus genome sequences of 32 DENV-1 and 9 DENV-2 sampled from amongst the 54 patients with uncomplicated dengue (Fig. 1-2). All DENV-1 sequences belonged to the genotype I lineage and there was no evidence of a phylogenetic structure in the maximum likelihood tree that was related to clinical outcome (Fig. 1). All DENV-2 sequences belonged to the Asian-1 lineage and similarly, consensus genome sequences from DSS patients were not phylogenetically distinct from uncomplicated dengue cases (Fig. 2).

**Differences in host gene transcript abundance between the early febrile phase of dengue and convalescence**

We undertook an expansive interrogation of the early host transcriptional signature in dengue by comparing transcriptional profiles of 9870 genes in samples collected early in the acute phase (fever day -2 or -3) from all 80 acute dengue patients (24 DSS and 56 uncomplicated dengue) with 34 convalescent control samples (18 severe dengue and 16 uncomplicated dengue). By SAM analysis with a FDR <5%, we identified 860 differentially expressed (DE) transcripts (fold change equal or greater than 2, q value ≤5%). Of the 860 differentially abundant gene transcripts, 309 were less abundant and 551 were more abundant in acute samples relative to convalescent samples. InnateDB pathway, ontology and network analysis of differentially expressed transcripts was performed to identify the molecular pathways, networks and biological processes that dominate the whole blood transcriptional profile, and in particular, the major signalling hubs involved in the early immune response.

**Pathways and Gene Ontology terms associated with up-regulated genes**

Many of the over-abundant DE transcripts are interferon-inducible, indeed of the 551 up-regulated genes, 173 (31%) were annotated as Type I interferon-inducible by the Interferome database (33). Pathways that were significantly over-represented in up-regulated genes after correction for multiple testing (FDR < 5%) included Systemic lupus erythematosus (KEGG database); Classical complement pathway (PID...
BIOSCARTA) and Complement and coagulation cascades (KEGG). The pathway Systemic lupus erythematosus is likely significant due to the overlap between this pathway and the complement pathways and because of the large number of histone genes that are up-regulated. A number of complement and coagulation related genes were found to be up-regulated (C1QA; C1QB; C1QC; C2; C3AR1; C5; F9; PLAU; PROS1; SERPINA1; SERPING1). Prior to correction for multiple testing (which is conservative given the large number of pathways tested) a range of pathways were identified as being significantly over-represented. These pathways included Toll-like receptor signaling pathway, IL27-mediated signaling events, IL12-mediated signaling events, Chemokine signaling pathway, Cytokine-cytokine receptor interaction, Lysosome, RIG-I-like receptor signaling pathway, IFN-gamma pathway and the JAK-STAT pathway and regulation pathway. Many of these are likely of biological significance (despite being below the statistical threshold) and are highlighted in Table S2. Some of the prominent up-regulated transcripts here were associated with recognition of virus ligands and downstream signalling; TLRs, RIG-I, MDA5, IRF7, AIM2 and DAI and STATs all of which make contributions to innate anti-viral sensing and downstream signalling.

The most significant GO terms associated with up-regulated genes (FDR < 5%) included response to virus, immune response, innate immune response and inflammatory response. A full list of GO terms associated with up-regulated genes is described in Table S3.

Pathways and Gene Ontology terms associated with down-regulated genes
Almost all of the significantly down-regulated pathways were related to translation (Table S4). These pathways were significant due to the down-regulation of genes encoding ribosomal subunit proteins. The pathway hemoglobins chaperone was also significantly over-represented due to the down-regulation of several genes involved in the heme biosynthesis pathway (ALAD; ALAS2; ERAF; FECH; HBB; HMBS). GO analysis paints a similar picture to the pathway analysis with a number of terms related to translation all being significant (FDR < 5%) Table S5.

Network analysis of differential gene expression profiles in acute vs. convalescent dengue
InnateDB (www.innatedb.com) (27) was used to generate molecular interaction networks involving DE genes and their encoded products (DE nodes). The first was a network consisted of only the interactions between DE nodes (AcuteDengue_DE network, Fig. 3), while the second expanded upon this network by including all non-differentially expressed interacting partners of the DE nodes (AcuteDengue_All network, Fig. S1). The AcuteDengue_DE network has 289 nodes and 429 edges, in comparison to the more extensive AcuteDengue_All network, which had 4,364 nodes and 9,094 edges.

The AcuteDengue_DE network was analysed to identify network hubs and bottlenecks which may represent the key regulatory nodes in the network. Using the “Degree” algorithm from the cytoHubba plugin (25) the top 5 hubs (i.e. genes/proteins that are highly connected to other DE genes) in this network were identified as the transcription factors STAT1 and STAT2 (2-fold up-regulated), the tyrosine kinase SRC (2-fold up-regulated), PTPN6 (SHP1) (2.5 fold down-regulated) and C1orf103 (2-fold up-regulated). The Hubba software also allows one to predict proteins that act as bottlenecks in the network. Bottlenecks are network nodes that are the key connector proteins in a network and have many “shortest paths” going through them, similar to bridges or tunnels on a highway map (43). Several of the hubs including STAT1, SRC and PTPN6 (SHP1), were identified as bottlenecks in the network, further supporting their central role in signalling. TRIP6 and JAK2 were also identified in the top 5 bottlenecks. Analysis of the AcuteDengue_All network, which consists of all interactions involving molecules encoded by DE genes (regardless of whether the interacting molecule is DE), also identified SRC, STAT1, TRIP6, SHP1, C1orf103 and JAK2 in the top 20 hub/bottleneck nodes (Fig. S2). SRC is the highest ranked hub in this network and SRC and STAT1 are the top 2 bottlenecks.

Two major differentially expressed sub-networks were identified in the AcuteDengue_DE network. The top-ranked network consisted of 23 nodes (including JAK2, JAK3, SRC, TLR2, IL2RG, SOCS1, SHP1, TRIP6 and other JAK/STAT and SRC regulators) (Fig. S3A). The second ranked sub-network, also consisting of 23 nodes, was enriched for nuclear-localised proteins (13 nodes) and in terms such as transcription corepressor activity (ATF3; DDIT3 (CEBPZ); ID3; NFIL3) and may represent an important transcriptional regulatory network (Fig. S3B).
Combining the transcriptional regulatory network and the physical interaction network

InnateDB was also used to construct a network of all predicted transcription factor (TF) interactions with DE genes (TF_network). This predicted transcriptional regulatory network was then merged with the AcuteDengue_ALL network. This combined network provides a much more comprehensive picture of the connection between signalling and transcriptional regulation (Fig. S4). Nodes with significantly more interactions with DE genes than expected by chance were highly enriched in transcription factors involved in IFN/NF-κB signalling responses including STAT1, STAT2, STAT3, IRF7, IRF9, IRF1, CEBPB and SP1. Notably, this approach enabled the identification of potentially important regulators that were not identified as DE at the time-point sampled. IRF1, for example, although not DE itself, is predicted to interact with 100 DE genes. Similarly, CEBPB is predicted to interact with 88 genes and is not itself DE. Gene expression analysis alone would not reveal the multitude of transcription factors that are likely the key regulators driving the host response to DENV infection.

Differences in early host gene transcript abundance between DSS and uncomplicated dengue patients

The previous analysis dealt with differences in the transcriptome between acute and convalescent samples. Since the identification of prognostic markers of severe dengue is an important goal, we turned our focus to identifying differences in the early acute transcriptome between severe and uncomplicated dengue cases. To this end we compared gene transcript abundance on fever day -2 or -3 in the 24 DSS patients and their 56 matched controls. By SAM analysis, twenty one transcripts were significantly enriched in DSS patients relative to uncomplicated dengue patients (FDR <5%, >2-fold difference in abundance). There were no significantly down-regulated transcripts in acute DSS patients compared to acute uncomplicated dengue patients. Table 2 summarises the list of 21 differentially abundant transcripts. Prominent amongst these were transcripts that could be linked to activated neutrophils, with 12 of the 21 differentially expressed genes associated with activated neutrophils in the shape of membrane-bound integrin receptors (CEACAM6, CEACAM8), cytokine decoy
receptor (IL1R2), secreted proteases (CTSG, ELA2), inflammatory molecules (S100A12), secreted anti-microbial proteins/peptides (DEF1, DEF4, BPI, CAMP, PGLYRP1) or oxidative enzymes (MPO). We explored possible functional relationships between the 21 differentially expressed genes using InnateDB and separately, the Ingenuity Pathway analysis (IPA) knowledge base. We describe here the IPA analysis as both methods identified networks enriched in neutrophil proteins. Unsupervised IPA network analysis identified two clusters of 35 genes each that included 18 of the 21 differentially expressed genes (Fig. 4). These two network interactions were highly unlikely to have occurred by chance ($P=10^{-31}$ and $P=10^{-21}$). The first cluster included 12 differentially expressed genes and the second cluster included 9 differentially expressed genes. The neutrophil-associated CAMP and MPO, together with the decoy receptor IL1R2, were found in both networks. Cathepsin G (CTSG) and Elastase (ELA), which were found in the list of 21 DE genes, were the key elements of the first network while TNF-α and β-estriadiol were the key elements of the second network. Strikingly, these data suggest gene expression profiles from members of two overlapping networks discriminate between patients who progress to DSS from those with an uncomplicated disease course. InnateDB identified the transcription factor CEBPB as a potential important regulator of several of the differentially expressed genes. We performed classification analysis using a decision tree forest model. In the analysis, the intensity signals of all 21 differentially expressed genes were used as “predictors” and the disease outcomes were used as “target”. From the total 200 trees built, the best misclassification rate was 15/24 (62.5%) for DSS and 10/46 (17.9%) for uncomplicated dengue patients. The result indicates that these differentially expressed genes could not be used as predictors of disease outcomes.

The differential abundance of neutrophil-associated transcripts in patients who progressed to DSS was not simply a reflection of neutrophil levels in the sample as the mean absolute count of neutrophils in each group was not significantly different from one another (median (95%CI) DSS patients $2.75*10^3$/mm$^3$ (2.22 – 3.33) versus $2.48*10^3$/mm$^3$ (2.27–3.10) in uncomplicated dengue cases). These results indicated neutrophils are phenotypically activated in children who subsequently develop DSS and suggests certain neutrophil-associated transcripts could have prognostic value in identifying patients at risk of severe disease.
Plasma concentrations of Bactericidal/permeability-increasing protein (BPI), Defensin-1α, Elastase 2 and Myeloperoxidase (MPO)

Concentrations of BPI, DEF1A, ELA2 and MPO were measured in plasma samples collected at the same time point as the RNA used for expression array analysis to determine if there was evidence suggestive of neutrophil activation. Consistent with the array findings, plasma concentrations of all four proteins were significantly higher at the time of study enrolment in those children who developed DSS relative to children with uncomplicated dengue, although the absolute difference was small (Fig. 5A-D). Plasma concentrations of BPI, DEF1A and ELA2 were also significantly higher in children who developed DSS compared to convalescent samples and healthy donor samples.

To understand if elevated concentrations of these neutrophil-associated proteins was independent of early haemoconcentration in children with DSS (i.e. reduced vascular volume leading to higher plasma protein concentrations), plasma albumin levels in all samples were measured as a surrogate marker of the plasma protein concentration (Fig. 5E). Levels of BPI, ELA2 and MPO were not significantly correlated with the plasma albumin concentration in the same sample, suggesting their elevated levels in plasma were not merely a reflection of haemoconcentration at this time point (Pearson’s correlation; ([Plasma albumin] vs. [BPI], (r=-0.04, P = 0.6), vs. ELA2 (r=-0.41, P = 0.6), or vs. MPO (r=-0.008, P = 0.9). In contrast, concentrations of DEF1A were weakly correlated with the plasma albumin concentration ([Plasma albumin] vs. DEF1A (r=-0.176, P = 0.014), suggesting at least some component of this measurement might have been influenced by the reduced vascular volume.
Discussion

This case-control study investigated the early whole blood transcriptional signature in children who subsequently developed DSS, the most common life-threatening complication of dengue in children. This study was rooted in clinical practice by focusing on the dengue syndrome that always requires a clinical intervention, often in the setting of the intensive care unit. Strikingly, we identified in the first few days of illness, two overlapping gene networks that distinguished patients who developed DSS from those with uncomplicated dengue. A feature of these networks were genes associated with neutrophil activation and degranulation, suggesting a hitherto unrecognised association of neutrophils with pathogenesis and expression of the overall disease phenotype.

Previous studies from our groups have described the whole blood transcriptional signature in dengue patients by microarray analysis (11, 26, 37). The timing of sample collection is clearly a major factor in the transcriptional signature, with samples collected during the febrile phase having a characteristic anti-viral profile, e.g. with interferon-stimulated genes highly prominent (11, 26, 37), whilst those collected during the afebrile stage having a predominantly metabolic profile (26, 37). Studies of PBMC (i.e. minus the neutrophil population) have also been described (30, 41). In short however, all previous microarray studies of the dengue host response have been relatively small in size, collected samples at heterogenous timepoints and rarely included patients with DSS, the commonest life-threatening complication in children. Against this backdrop, a strength of the current study is the matched case-control design, large sample size, early sampling prior to defervescence and cardiovascular decompensation and inclusion of detailed genomic scale information on the infecting pathogen.

Previous epidemiological, in vitro and in vivo studies have suggested phenotypic differences can exist between virus lineages of the same serotype (1, 9, 42). We determined consensus virus genome sequences and demonstrated that viruses from DSS patients in this study were not phylogenetically different from those in patients with uncomplicated dengue, implying that host factors were more important determinants of the clinical course. In addition, the viremia (as measured by qRT-PCR) was not significantly higher in children who subsequently developed DSS,
although the plasma NS1 concentration was, suggesting that antigen burden may be a better correlate of severe clinical outcomes, as has been alluded to previously (24).

Using a large sample size, we defined the major transcriptional features of the acute response to DENV infection during the febrile period. Complement, TLR and RIG-I signalling pathways, interferon-stimulated genes and cytokine/chemokines and their receptors were the major features of the transcriptional signature, consistent with previous studies in smaller numbers of febrile dengue patients (11, 26, 41). Utilising a systems biology approach that investigated transcriptional profiles in the context of their molecular interaction networks, we identified the transcription factors STAT1 and STAT2, the tyrosine kinase SRC, SHP1, TRIP6 and JAK2 as key central molecules in these networks. Furthermore, the top-ranked differentially expressed sub-network was enriched in molecules involved in cytokine signalling and JAK/STAT pathways (including JAK2, JAK3, SRC, TLR2, IL2RG, SOCS1, SHP1, TRIP6 and other JAK/STAT and SRC regulators). The top hub/bottleneck molecules in these networks form a densely connected network module, with a variety of known interactions between the nodes, which supports this being a core signalling module in the network and a central feature of the host response. Analysis of the transcriptional regulatory network also identified several STAT and IRF transcription factors as the key regulators of the transcriptional response. The importance of the STAT1 pathway in control of DENV replication in mice and mosquitoes has been demonstrated previously (36, 38) and STAT1/2 may also be targets of DENV-mediated interference in the interferon signalling pathway (20, 29).

We identified 21 genes as differentially expressed (more abundant) in patients who developed DSS compared to matched control patients. Remarkably, almost all of these genes belonged to one of two, overlapping networks, in which some of the interconnecting elements have immune response functions. That TNF-α should be a central hub in the 2nd network is striking, given that TNF-α has been repeatedly implicated in the pathogenesis of severe capillary leakage (4, 14). Although TNF-α is a major hub in the network analysis, we did not measure TNF-α plasma level because TNF-α transcripts were not significantly different between DSS and uncomplicated dengue patients at the time of enrolment. The suggestion instead from our work is that there may only be small differences in the acute response that leads to a “tipping
point" and the physiological derangement of normal vascular endothelial cell functions. Furthermore, the data sheds light on previously unrecognised aspects of the acute response to DENV infection and provides an impetus for further investigations of the role of neutrophils in pathogenesis. Clearly, further studies are needed to understand why these two networks, and some of their differentially expressed members, are associated with progression to DSS.

Neutrophil activation and degranulation was a most prominent theme in the DSS-associated differentially expressed gene list. A gene signature consistent with neutrophil activation has also recently been described in the whole blood of Cambodian children with DSS (10). We verified that plasma proteins concentrations of CTSG, BPI, ELA2 and MPO were also higher in early DSS than in control patients, albeit the absolute difference was small and unlikely to be useful for prognosis. Of the differentially expressed genes associated with neutrophil degranulation, ELA2, CTSG and the defensins (DEFA1, DEFA4) are of particular interest. ELA2 and CTSG are serine proteases that can cleave vascular endothelial cadherin and thereby compromise the integrity of the vascular endothelium (18). These proteases might conceivably play a similar role in perturbing the endothelium in capillary beds during DENV infection. Accordingly, ELA2 has also been detected previously at higher concentrations in serum of patients with DSS compared with patients without shock (22). DEF1A and DEF4A are neutrophil-associated defensins with anti-viral activity (12, 16, 17). These innate anti-microbial peptides may also functionally participate in the innate anti-viral response to DENV infection.

The dengue capillary leakage syndrome begins in children with secondary infections within the first few 1-2 days of illness and can be measured by ultrasound as early as the 3rd day of illness (39). Commencement of capillary leakage in infants with primary dengue likely occurs with similar kinetics and can lead to DSS between illness days 4-6 (37). The triggering of capillary permeability early in the disease evolution in both these clinical settings might be mediated by robust innate immune responses rather than acquired responses, particularly since pre-existing immunity does not exist in infants with primary infection. We propose that activated, degranulating peripheral blood neutrophils could contribute to an early triggering of capillary permeability. In this model, neutrophils (which express Fc receptors) are
activated by immune complexes and/or by high early virus antigen burdens in blood and tissues, where they secrete cytokines and chemotactic molecules. High viral antigen burdens in vivo could be arrived at by antibody dependent enhancement in secondary infection or in primary infection of infants born to dengue immune mothers. Neutrophil activation and degranulation alone is highly unlikely to be sufficient to drive capillary leakage to the extent seen in patients with DSS. Instead, neutrophil adherence to endothelial cells and the secretion of soluble mediators of vascular permeability may represent a small step in the inflammatory cascade that synergises with other adaptive host responses to mediate capillary permeability in severe secondary infection. In infants with primary infection and severe dengue, an innate response that includes neutrophil activation, together with an inherently permeable vascular endothelium, might be sufficient to trigger clinically significant vascular leak. Further studies of neutrophils and their response to DENV infection are warranted.
Acknowledgements

This work was supported by The Wellcome Trust and A*STAR Singapore. InnateDB is funded by Genome Canada and Genome BC through the Pathogenomics of Innate Immunity (PI2) project, and by the Foundation for the National Institutes of Health and the Canadian Institutes of Health Research under the Grand Challenges in Global Health Research Initiative (Grand Challenges ID: 419). Continued development of InnateDB is also supported by Teagasc (RMIS 6018).
References


expression in peripheral blood mononuclear cells indicate a significant role of
the innate responses in progression of dengue fever but not dengue
Hayes, and S. B. Halstead. 1999. Failure of secondary infection with
American genotype dengue 2 to cause dengue haemorrhagic fever. Lancet
43. Yu, H., P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein. 2007. The
importance of bottlenecks in protein networks: correlation with gene
Table 1: Summary characteristics of patients included in microarray study

<table>
<thead>
<tr>
<th>Variable</th>
<th>DSS (n = 24)</th>
<th>Uncomplicated dengue (n = 56)</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (range)</td>
<td>11 (2 – 30)</td>
<td>10.5 (2 - 29)</td>
<td></td>
</tr>
<tr>
<td>Gender (Male (%))</td>
<td>16/24 (67%)</td>
<td>34/56 (61%)</td>
<td></td>
</tr>
<tr>
<td>Median day of illness (range)</td>
<td>3 (2 - 5)</td>
<td>3 (2 - 5)</td>
<td></td>
</tr>
<tr>
<td>Fever day (range)</td>
<td>-3 (-3 to -2)</td>
<td>-3 (-3 to -2)</td>
<td></td>
</tr>
<tr>
<td>Received i.v. fluids</td>
<td>24 (100%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Median array sample date relative to date of shock (range)</td>
<td>-2 (-3 – 0)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Primary infection Unknown</td>
<td>1 (4%)</td>
<td>1 (1.8%)</td>
<td></td>
</tr>
<tr>
<td>DENV Serotype</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DENV1</td>
<td>16 (67%)</td>
<td>41 (73%)</td>
<td>0.7</td>
</tr>
<tr>
<td>DENV2</td>
<td>7 (29%)</td>
<td>15 (27%)</td>
<td></td>
</tr>
<tr>
<td>DENV3</td>
<td>0 (0%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>DENV4</td>
<td>0 (0%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (4%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Viremia Median (range)</td>
<td>3.47E+7</td>
<td>(Missing = 14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.11E+4 – 2.15E+9)</td>
<td>4.18E+07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.27E+5 – 1.46E+9)</td>
<td></td>
</tr>
<tr>
<td>NS1 ng/ml Median (range)</td>
<td>1002 (0.01 – 3842)</td>
<td>(Missing= 17)</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>117 (0.01 – 3415)</td>
<td></td>
</tr>
<tr>
<td>Neutrophils (x10^3 cells/ul)</td>
<td>(Missing= 1)</td>
<td>2.48</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>2.75</td>
<td>(0.8 – 5.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.7 – 8.4)</td>
<td></td>
</tr>
</tbody>
</table>

* Fever day is the number of days relative to the time point when the patient became afebrile, defined as fever day 0.
* Mann-Whitney test P value.
Table 2: Differentially expressed transcripts in acute DSS patients relative to uncomplicated dengue patients at fever day -2 or -3

<table>
<thead>
<tr>
<th>Symbol</th>
<th>SAM analysis</th>
<th>Acute DSS</th>
<th>Acute UC dengue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fold Change</td>
<td>q-value (%)</td>
<td># Samples detected (%)</td>
</tr>
<tr>
<td>DEF4A</td>
<td>6.7</td>
<td>0</td>
<td>22 (92)</td>
</tr>
<tr>
<td>CEACAM6</td>
<td>6.5</td>
<td>0</td>
<td>12 (50)</td>
</tr>
<tr>
<td>CTSG</td>
<td>6</td>
<td>2.6</td>
<td>17 (71)</td>
</tr>
<tr>
<td>CEACAM8</td>
<td>5.3</td>
<td>0</td>
<td>22 (92)</td>
</tr>
<tr>
<td>ARG1</td>
<td>4.5</td>
<td>0</td>
<td>20 (83)</td>
</tr>
<tr>
<td>M5A4A3</td>
<td>4.5</td>
<td>0</td>
<td>20 (83)</td>
</tr>
<tr>
<td>BPI</td>
<td>3.8</td>
<td>0</td>
<td>21 (88)</td>
</tr>
<tr>
<td>ELA2</td>
<td>3.8</td>
<td>0</td>
<td>23 (96)</td>
</tr>
<tr>
<td>PGLYRP1</td>
<td>3.7</td>
<td>0</td>
<td>24 (100)</td>
</tr>
<tr>
<td>RNASE3</td>
<td>3.6</td>
<td>0</td>
<td>12 (50)</td>
</tr>
<tr>
<td>M5A4A3</td>
<td>3.2</td>
<td>4.4</td>
<td>20 (83)</td>
</tr>
<tr>
<td>LOC72835B</td>
<td>2.7</td>
<td>0</td>
<td>24 (100)</td>
</tr>
<tr>
<td>HP</td>
<td>2.5</td>
<td>0</td>
<td>24 (100)</td>
</tr>
<tr>
<td>S100P</td>
<td>2.3</td>
<td>2.4</td>
<td>24 (100)</td>
</tr>
<tr>
<td>CCNA1</td>
<td>2.1</td>
<td>4.7</td>
<td>22 (92)</td>
</tr>
<tr>
<td>S100A12</td>
<td>2.1</td>
<td>0</td>
<td>24 (100)</td>
</tr>
<tr>
<td>CAMP</td>
<td>2.4</td>
<td>0</td>
<td>24 (100)</td>
</tr>
<tr>
<td>TFF3</td>
<td>4.9</td>
<td>0</td>
<td>11 (46)</td>
</tr>
<tr>
<td>ORM1</td>
<td>4.5</td>
<td>0</td>
<td>9 (38)</td>
</tr>
<tr>
<td>MPO</td>
<td>2.2</td>
<td>4.3</td>
<td>24 (100)</td>
</tr>
<tr>
<td>IL1R2</td>
<td>2</td>
<td>2.1</td>
<td>21 (88)</td>
</tr>
</tbody>
</table>

Note: Std: Standard deviation; DEF4A: Defensin, alpha 4, corticostatin; CEACAM6: Carcinoembryonic antigen-related cell adhesion molecule 6; CTSG: Cathepsin G; CEACAM8: Carcinoembryonic antigen-related cell adhesion molecule; BPI: Bactericidal permeability-increasing protein; ELA2: Elastase 2, neutrophil; PGLYRP1: Peptidoglycan recognition protein 1; RNASE3: Ribonuclease, RNase A family, 3 (eosinophil cationic protein); DEF1A: Defensin alpha 1; HP: Haptoglobin; S100P: S100 calcium binding protein P; CCNA1: Cyclin A1; S100A12: S100 calcium binding protein 12; CAMP: Cathepsin G; TFF3: TFF3: Trefoil factor 3 (intestinal); ORM1: Orosomucoid 1; MPO: Myeloperoxidase; IL1R2: Interleukin 1 receptor, type II.

28
Figure legends

Figure 1: Phylogenetic tree of DENV-1 consensus genome sequences from patients in this study. The tree (Neighbor Joining method) contains consensus genome sequences deduced from the plasma of 32 DENV-1 infected patients with uncomplicated dengue (grey highlighted tip labels) and 15 genomes sampled from DENV-1 infected patients with DSS (black highlighted tip labels). The tree is mid-point rooted and contains sequences from other DENV-1 viruses for reference only (black tip labels). Bootstrap values are shown on major branches.

Figure 2: Phylogenetic tree of DENV-2 consensus genome sequences from patients in this study. The tree (Neighbor Joining method) contains genome sequences deduced from the plasma of 9 DENV-2 infected patients with uncomplicated dengue (grey highlighted tip labels) and 3 genomes sampled from DENV-2 infected patients with DSS (black highlighted tip labels). The tree is mid-point rooted and contains sequences from other DENV-2 viruses for reference only (black tip labels). Bootstrap values are shown on major branches.

Figure 3: AcuteDengue_DE_network - A Network of known protein-protein and protein-DNA interactions encoded by genes differentially expressed in 80 acute Dengue patients in comparison to 34 convalescent samples. Nodes encoded by up-regulated genes are shown in red; down-regulated in green. This network was generated using InnateDB (www.innateDB.com) and was visualised using the Cerebral 2.0 plugin for Cytoscape 2.6.2 which was developed as part of the InnateDB project. This network has 289 nodes and 429 edges. The top 5 hubs (i.e. genes/proteins that are highly connected to other DE genes) in this network were identified as the transcription factors STAT1 and STAT2 (2X up-regulated), the tyrosine kinase SRC (2X up-regulated), PTPN6 (SHP1) (2.5X down-regulated) and C1orf103 (2X up-regulated).

Figure 4: Ingenuity Pathway Analysis (IPA) of differentially expressed genes between 24 DSS and 56 uncomplicated dengue patients. Twenty one differentially expressed transcripts were analyzed using IPA. Two significant networks were identified: A) Cancer, cell cycle, cell-mediated immune response (network 1, score 31) and B) antigen presentation, cell-mediated immune response, humoral immune response (Network 2, score 21). The lines between genes represent known interactions, with solid lines representing direct interactions and dashed lines representing indirect interactions. Differentially expressed genes are highlighted in red, non-highlighted genes were identified by IPA. The high scores associated with these networks indicate they were highly unlikely to be formed by chance.
Figure 5: Concentrations of secreted neutrophil-associated proteins in plasma. Concentrations of A) BPI, B) DEF1A, C) ELA2, D) MPO and E) albumin in acute and convalescent dengue cases, and for reference, from patients with other febrile illness and healthy donor plasma samples. The box and whisker plots represent median and interquartile ranges.
Figure 1
Figure 2
Figure 3