Human Conjunctival Transcriptome Analysis Reveals the Prominence of Innate Defense in *Chlamydia trachomatis* Infection

Angels Natividad,‡ Tom C. Freeman, David Jeffries, Matthew J. Burton, David C. W. Mabey, Robin L. Bailey, and Martin J. Holland

Department of Clinical Research, London School of Hygiene and Tropical Medicine, University of London, London WC1E 7HT, United Kingdom; The Roslin Institute, R(D)SVS, University of Edinburgh, Roslin Biocentre, Midlothian EH25 9PS, United Kingdom; and Medical Research Council Laboratories, Viral Diseases Programme, Fajara, Banjul, the Gambia

Received 3 August 2010/Returned for modification 17 August 2010/Accepted 24 August 2010

Trachoma is the leading infectious cause of blindness and is endemic in 52 countries. There is a critical need to further our understanding of the host response during disease and infection, as millions of individuals are still at risk of developing blinding sequelae. Infection of the conjunctival epithelial cells by the causative bacterium, *Chlamydia trachomatis*, stimulates an acute host response. The main clinical feature is a follicular conjunctivitis that is incompletely defined at the tissue-specific gene expression and molecular levels. To explore the features of disease and the response to infection, we measured host gene expression in conjunctival samples from Gambian children with active trachoma and healthy controls. Genome-wide expression and transcript analysis network identification signatures characteristic of the expected infiltrating immune cell populations, such as neutrophils and T/B lymphocytes. The expression signatures were also significantly enriched for genes in pathways which regulate NK cell activation and cytotoxicity, antigen processing and presentation, chemokines, cytokines, and cytokine receptors. The data suggest that in addition to polymorph and adaptive cellular responses, NK cells may contribute to a significant component of the conjunctival inflammatory response to chlamydial infection.

*Chlamydia trachomatis* is an obligate intracellular bacterium that infects millions of people worldwide. Infection of conjunctival epithelial cells causes trachoma, which is the world’s leading infectious cause of blindness, affecting over 40 million people in the developing world (50). *C. trachomatis* is also the world’s most common bacterial sexually transmitted infection (STI), with an estimated 92 million new cases of *C. trachomatis* urogenital infections occurring annually (74). Asymptomatic infection is common, and untreated cases are at risk of developing complications related to fertility and pregnancy (6). A vaccine to prevent *C. trachomatis* infection or disease progression would be of great value, but the development of such a vaccine is handicapped by the fact that the immunological correlates of protective immunity and pathogenesis are not well understood.

The immune and inflammatory responses initiated by *C. trachomatis* infection, although important for successful control and resolution of infection, are thought to be at least partly responsible for tissue damage and its sequelae (10). Some progress has been made in dissecting correlates of protective immunity and immunopathology in humans (14), but reports are dominated by studies using the mouse as a model system (reviewed in reference 54). The extrapolation of results obtained from murine experimental models needs cautious interpretation. Several papers have demonstrated the exquisite and often subtle adaptations of chlamydial parasites to their natural host tissues and the specificity of the molecular pathways that control intracellular replication (53, 58, 67).

Ocular *C. trachomatis* infection is readily accessible to examination and investigation. As a result, the clinical and epidemiological features of trachoma and the phases of disease are well documented in many populations (49). Trachomatous inflammation (follicular) (TF) and/or trachomatous inflammation (intense) (TI) is characterized by *C. trachomatis*-driven conjunctival inflammation and is most prevalent during childhood (<10 years). Conversely, the sequelae of these earlier stages, i.e., trachomatous conjunctival scarring (TS), trachomatous trichiasis (TT), and corneal opacity (CO), are most prevalent in adults (>45 years) and require repeated or persistent episodes of infection and inflammation.

Gene expression profiling studies have used a variety of array platforms to study the host response to chlamydial infection (2, 18, 32, 43, 44, 47, 64, 66, 70, 86, 89, 93). With the exception of two recent studies (43, 64), each has described the *in vitro* response in experimentally infected cell lines or cells isolated from tissue. Numerous genes and pathways have been implicated as being important in the innate response to infection. Thus far, there have been no transcriptome-defining studies of human tissues that are infected or diseased as a result of natural chlamydial infection. To gain a better understanding of the immune and inflammatory responses to ocular *C. trachomatis* infection in humans, we collected conjunctival swabs.
from the upper tarsal conjunctiva of Gambian children with active trachoma and examined their transcriptomes using gene-wide expression arrays.

The extraction of biological meaning from microarray data is challenging and complex. This has led to the development of many new computational tools and methods for their analysis. We used statistical methods to define differential gene expression and a graph theoretic method to define networks of co-expressed transcripts (27). The latter method analyzes the degree of correlation (coexpression) between transcripts (79) and can help define the transcriptional networks which are characteristic of the cell types present in conjunctival samples.

**Materials and Methods**

**Ethical approval.** The study was approved by the Gambia Government/Medical Research Council Joint Ethics Committee (reference L2006/47) and by the Ethics Committee of the London School of Hygiene and Tropical Medicine (LSHTM). Written informed consent was obtained from all study subjects and their guardians. Children diagnosed with clinical signs of active trachoma were treated according to National Eye Care Programme guidelines with topical azithromycin (20 mg/kg).

**Study subjects.** Infants and children from the North Bank, Western, and Central River regions in the Gambia were screened for clinical signs of trachoma. Sterile polyester-tipped swab (Hardwood Products, ME) samples were collected from the upper tarsal conjunctiva from 200 children with clinical signs of inflammatory trachoma and from control subjects without clinical signs of active trachoma. Clinical diagnosis was made according to the WHO simplified trachoma grading system (30); each subject was graded based on follicular score (F0 to F3), papillary score (P0 to P3), and conjunctival scattering score (C0 to C3), with the presence of clinically active trachoma indicated by a follicular score of 2 or 3 (F2/F3) or a papillary score of 3 (P3). These are equivalent to trachomatous inflammation (follicular) (TF) and trachomatous inflammation (intense) (TI) of the WHO simplified trachoma grading system, respectively (81).

Anesthetic eye drops (Minims Prosymetacaine 0.5%; Chauvin Pharmaceuticals, Romford, United Kingdom) were administered prior to swab collection. Swab collection was performed in a standardized manner (12). Two swabs per eye were collected; the first swab (swab A) was collected into RNAlater (Ambion, Huntingdon, United Kingdom), and the second swab (swab B) was collected in a dry polypropylene tube. Both samples were stored on ice until frozen (−20°C) in the laboratory on the same day. Samples were then shipped on dry ice to the London School of Hygiene and Tropical Medicine (LSHTM).

**Processing of conjunctival swab samples for human gene expression studies.** Total RNA was extracted from swab A samples using the RNeasy micro kit (QIAGEN, Crawley, United Kingdom) according to the manufacturer’s instructions. A modification of the protocol was used. Total RNA was isolated from 200 children with clinical signs of active trachoma and examined their transcriptomes using gene-wide expression arrays.

**Processing of conjunctival swab samples for diagnosis and quantification of C. trachomatis.** DNA was extracted from dry conjunctival swabs (swab B) for diagnosis and quantification of C. trachomatis as described by Solomon et al. (72). The Amplicor CT/NG diagnostic kit (Roche Molecular Systems, Branchburg, NJ) was used as a qualitative PCR for C. trachomatis. Both Amplico-reactive and -negative samples were then purified using DNA extraction spin columns (QIAamp DNA mini kit; Qiagen, Crawley, United Kingdom) and tested by a C. trachomatis quantitative PCR (qPCR) which amplifies a 123-bp fragment within the 1% level (8). Large-scale differential gene expression was estimated in two ways: first by calculation of B values (log odds of p [transcript is differentially expressed]/p [transcript is not differentially expressed]), where a B value of ≥ 0 indicates that transcripts are more likely than not to be differentially expressed, and second by fold change expression, defined as the ratio of the adjusted P value of <0.01 and log2 fold change (FC) of <−0.7 (FC < 0.6) and >0.7 (FC > 1.6) or <−1.0 (FC < 0.5) and >1.0 (FC > 0.2).

**Network analysis.** A sample-to-sample Pearson correlation matrix of normalized total gene expression was calculated using the cor function within R. A graph was then generated using the matrix file in BioLayout Express3D (www.biolayout.org) (27). An overall gene expression graph was then generated using a standard gene expression data set and was subjected to an interquartile range (IQR) filtering, in which array features with an IQR of <0.5 log2 intensity unit were discarded, removing outliers and weakly expressed probe sets. Group mean differences in average expression for each transcript were tested through empirical Bayes moderated t tests (Bioconductor package limma). The simultaneous testing of a large number of variables was accounted for by controlling the false-discovery rate (FDR) at the 1% level (8). Large-scale differential gene expression was estimated in two ways: first by calculation of B values (log odds of p [transcript is differentially expressed]/p [transcript is not differentially expressed]), where a B value of ≥ 0 indicates that transcripts are more likely than not to be differentially expressed, and second by fold change expression, defined as the ratio of the adjusted P value of <0.01 and log2 fold change (FC) of <−0.7 (FC < 0.6) and >0.7 (FC > 1.6) or <−1.0 (FC < 0.5) and >1.0 (FC > 0). Independent confirmation of the transcriptional networks which are characteristic of the cell types present in conjunctival samples.

**Microarray data processing and statistical analysis.** Array data files (.cel) were run through Bioconductor R (http://www.bioconductor.org) in order to perform quality assessment and normalization. The data set was normalized using the Robust Multichip Average (RMA) algorithm (40). Normalized data were then subjected to an interquartile range (IQR) filtering, in which array features with an IQR of <0.5 log2 intensity unit were discarded, removing outliers and weakly expressed probe sets. Group mean differences in average expression for each transcript were tested through empirical Bayes moderated t tests (Bioconductor package limma). The simultaneous testing of a large number of variables was accounted for by controlling the false-discovery rate (FDR) at the 1% level (8). Large-scale differential gene expression was estimated in two ways: first by calculation of B values (log odds of p [transcript is differentially expressed]/p [transcript is not differentially expressed]), where a B value of ≥ 0 indicates that transcripts are more likely than not to be differentially expressed, and second by fold change expression, defined as the ratio of the adjusted P value of <0.01 and log2 fold change (FC) of <−0.7 (FC < 0.6) and >0.7 (FC > 1.6) or <−1.0 (FC < 0.5) and >1.0 (FC > 0.2).

**Network analysis.** A sample-to-sample Pearson correlation matrix of normalized total gene expression was calculated using the cor function within R. A graph was then generated using the matrix file in BioLayout Express3D (www.biolayout.org) (27). An overall gene expression graph was then generated using a standard gene expression data set and was subjected to an interquartile range (IQR) filtering, in which array features with an IQR of <0.5 log2 intensity unit were discarded, removing outliers and weakly expressed probe sets. Group mean differences in average expression for each transcript were tested through empirical Bayes moderated t tests (Bioconductor package limma). The simultaneous testing of a large number of variables was accounted for by controlling the false-discovery rate (FDR) at the 1% level (8). Large-scale differential gene expression was estimated in two ways: first by calculation of B values (log odds of p [transcript is differentially expressed]/p [transcript is not differentially expressed]), where a B value of ≥ 0 indicates that transcripts are more likely than not to be differentially expressed, and second by fold change expression, defined as the ratio of the adjusted P value of <0.01 and log2 fold change (FC) of <−0.7 (FC < 0.6) and >0.7 (FC > 1.6) or <−1.0 (FC < 0.5) and >1.0 (FC > 0.2).

**Network analysis.** A sample-to-sample Pearson correlation matrix of normalized total gene expression was calculated using the cor function within R. A graph was then generated using the matrix file in BioLayout Express3D (www.biolayout.org) (27). An overall gene expression graph was then generated using a standard gene expression data set and was subjected to an interquartile range (IQR) filtering, in which array features with an IQR of <0.5 log2 intensity unit were discarded, removing outliers and weakly expressed probe sets. Group mean differences in average expression for each transcript were tested through empirical Bayes moderated t tests (Bioconductor package limma). The simultaneous testing of a large number of variables was accounted for by controlling the false-discovery rate (FDR) at the 1% level (8). Large-scale differential gene expression was estimated in two ways: first by calculation of B values (log odds of p [transcript is differentially expressed]/p [transcript is not differentially expressed]), where a B value of ≥ 0 indicates that transcripts are more likely than not to be differentially expressed, and second by fold change expression, defined as the ratio of the adjusted P value of <0.01 and log2 fold change (FC) of <−0.7 (FC < 0.6) and >0.7 (FC > 1.6) or <−1.0 (FC < 0.5) and >1.0 (FC > 0.2).

**Network analysis.** A sample-to-sample Pearson correlation matrix of normalized total gene expression was calculated using the cor function within R. A graph was then generated using the matrix file in BioLayout Express3D (www.biolayout.org) (27). An overall gene expression graph was then generated using a standard gene expression data set and was subjected to an interquartile range (IQR) filtering, in which array features with an IQR of <0.5 log2 intensity unit were discarded, removing outliers and weakly expressed probe sets. Group mean differences in average expression for each transcript were tested through empirical Bayes moderated t tests (Bioconductor package limma). The simultaneous testing of a large number of variables was accounted for by controlling the false-discovery rate (FDR) at the 1% level (8). Large-scale differential gene expression was estimated in two ways: first by calculation of B values (log odds of p [transcript is differentially expressed]/p [transcript is not differentially expressed]), where a B value of ≥ 0 indicates that transcripts are more likely than not to be differentially expressed, and second by fold change expression, defined as the ratio of the adjusted P value of <0.01 and log2 fold change (FC) of <−0.7 (FC < 0.6) and >0.7 (FC > 1.6) or <−1.0 (FC < 0.5) and >1.0 (FC > 0.2).

**Network analysis.** A sample-to-sample Pearson correlation matrix of normalized total gene expression was calculated using the cor function within R. A graph was then generated using the matrix file in BioLayout Express3D (www.biolayout.org) (27). An overall gene expression graph was then generated using a standard gene expression data set and was subjected to an interquartile range (IQR) filtering, in which array features with an IQR of <0.5 log2 intensity unit were discarded, removing outliers and weakly expressed probe sets. Group mean differences in average expression for each transcript were tested through empirical Bayes moderated t tests (Bioconductor package limma). The simultaneous testing of a large number of variables was accounted for by controlling the false-discovery rate (FDR) at the 1% level (8). Large-scale differential gene expression was estimated in two ways: first by calculation of B values (log odds of p [transcript is differentially expressed]/p [transcript is not differentially expressed]), where a B value of ≥ 0 indicates that transcripts are more likely than not to be differentially expressed, and second by fold change expression, defined as the ratio of the adjusted P value of <0.01 and log2 fold change (FC) of <−0.7 (FC < 0.6) and >0.7 (FC > 1.6) or <−1.0 (FC < 0.5) and >1.0 (FC > 0.2).
RESULTS

C. trachomatis infection. Age and gender distributions were comparable among the clinical groups (see Table S1 in the supplemental material). The median age of participants was 6 years (range, 1 to 13 years). The C. trachomatis infection loads ranged from <19 to 185,270 C. trachomatis ompA copies per swab. A positive result by Amplicor and increasing conjunctival load were significantly associated with clinical severity (F score) (see Table S2 in the supplemental material) (36). High-load infection was detected in the majority (11/20) of Amplicor-positive participants. The high-load infections accounted for all but one of the participants with the highest clinical severity scores. The quantified load in a minority of Amplicor-positive participants amplified with a reverse primer which binds genovar B ompA strain types more efficiently than genovar A ompA strain types, and these samples are indicated in Table S1 in the supplemental material (35).

Clustering of study subjects by global conjunctival gene expression correlation. Quality control of Affymetrix gene expression data demonstrated that all data were of a consistent and high quality. The pairwise Pearson correlation coefficient of global gene expression was used as the distance metric to construct a matrix from which the individual samples were hierarchically clustered using an agglomerative clustering algorithm with complete linkage (Fig. 1a). This clustered the samples into three main branches. Branch 1 (n = 26) contains 95% of the samples from healthy subjects. Of the remaining samples in this branch, seven had F scores of ≥2 and a single sample had evidence of current intermediate-level infection (134 copies per swab). One participant also had papillary inflammation (P2). Branch 2 (n = 21) contained the single remaining sample from a subject with a normal conjunctiva. The remaining 20 samples in this branch were collected from subjects with follicular scores of ≥F2 or papillary scores of ≥P3. Fifteen of these had no or low-level infection (<34). Three samples had intermediate infection loads, and three had high-load infection (>390 copies per swab). Branch 3 (n = 13) contained subjects with the highest infection loads and clinical grades of disease (≥F2 or ≥P2). Eighty-four percent of the participants in branch three were infected, of whom over half (8/13) had high (≥390) infection loads. It is not surprising that the overall range of correlation between these 60 arrays is narrow (between 0.9 and 1.0), since the overwhelming majority of genes do not have altered expression between the clinical and infection states. Pairwise correlation of global expression stratified by disease severity or infection load was also explored using BioLayout Express3D (Fig. 1b), and this confirms that the expression profiles cluster with clinical and infection phenotypes.

Identification of genes differentially expressed during infection and disease. Large numbers of transcripts were both up- and downregulated when healthy and disease groups (with and without current infection) were compared (Table 1). There was a >2-fold change in the level of expression in 2,316 transcripts in samples derived from individuals with clinical disease and infection compared to healthy, infection-free individuals. However, the presence of disease signs alone induced differential expression of only 421 transcripts compared to healthy individuals. Among individuals with disease, the additional presence of C. trachomatis infection induced differential expression in 341 transcripts (the full gene list can be found in Table S3 in the supplemental material).

The major and fine biology of the differentially regulated genes were explored using the DAVID v6 database. Each of the differentially regulated gene lists showed significant gene enrichment for numerous annotation terms. The top-ranked gene ontology (GO) terms were typical of immune system activation, epithelial cell integrity, apoptosis, cell death, leukocyte migration, and interleukin receptor activity (Table 2). In particular, genes involved in pathways that control cell adhesion, chemokine signaling, antigen (Ag) processing and presentation, and regulation of the actin cytoskeleton were strongly represented.

Identification of differentially regulated transcripts for classification. The probe sets that were differentially expressed (>2-fold change and adjusted P value of <0.01) were separated in a Venn diagram that identified probe sets with significant differential expression unique to each comparison (Fig. 2). The samples from participants who were diseased and infected contained the vast majority of the transcripts with greatest changes in expression compared to normal conjunctiva. Other comparisons had relatively few unique transcripts (N versus D = 23 and D versus DI = 23). The principal newly identified GO terms uncovered included enrichment of genes characteristic of neutrophils and mast cell biology during disease and infection episodes. The use of a set of 63 probe sets defined as uniquely expressed in each comparison classified the samples with 75% accuracy across the three clinical groups (Fig. 3; see Table S4 in the supplemental material). The true class and the KNN class are shown in Fig. 3 with a heat map of expression intensity. Samples are ordered by clinical category and within-group similarity of expression.

Comparison of Affymetrix array gene expression results with previous independent studies of gene expression in trachoma. We compared the data obtained from this set of samples and arrays with data from previous studies in which participants were of equivalent age range and in which the diagnosis of current infection was made using the same PCR test (CT/NG Amplicor). Data were available from an independent set of samples from the same population in which gene expression was measured using Affymetrix HG-focused target arrays. The comparative results are shown in Tables S5a to S5d.
in the supplemental material. Overall there was strong correlation of fold change for all probe sets common to each array platform. This became very strong when genes with high levels of fold change were considered. Some probe sets, such as CXCL13 and S100A7, had high levels of fold change on one array platform (U133 plus 2.0) and low levels on the second platform; however, in each case the fold changes were statistically significant. Previous studies using qRT-PCR to estimate gene expression with conjunctival samples from children with active trachoma were also correlated with gene expression by microarray analysis. The highest degree of correlation was always obtained from the comparison of group N with group DI in all studies and platforms.

Transcription networks in the conjunctiva. The undirected network graph based on a Pearson correlation ($r$) threshold of $\geq 0.85$ contained 9,993 nodes (probe sets) representing approximately 8,359 genes connected by 245,457 edges. MCL clustering partitioned the network into 577 clusters of coexpressed genes. These clusters ranged in size from 1,148 to 4 transcripts and accounted for 7,719 (77.3%) of the probe sets in the original network. Probe sets that formed clusters containing $<4$ members that were part of the network were not assigned.
a cluster number (no class) and accounted for remaining 2,274 probe sets. Clusters established by MCL were assigned a number according to the number of probe sets they contained; MCL1 contained the largest number of probe sets (1,148), while MCL577 contained four probe sets. Each MCL cluster was assigned an arbitrary color, and the graph (Fig. 4) was explored for evidence of functional enrichment of gene clusters and the relationships between genes and clusters (the full list of transcripts and cluster assignment is available in Table S6 in the supplemental material). The graph (Fig. 4) shows the interrelationships and overlapping nature of the main large clusters and the discrete separation of other, smaller clusters. A number of small but interesting clusters confirm the power of this approach in identifying and grouping coexpressed genes. For example, MCL33 and -47 are derived from the probes for the Affymetrix labeling controls (24 probe sets) and the Affymetrix hybridization controls (19 probe sets), respectively. MCL37 was made up entirely of transcripts derived from the Y chromosome, which are expressed only in males. The clusters MCL12, -13, and -22 were all highly enriched with ribosomal genes.

The partitioned graph of 577 clusters contained three basic classes of clusters: genes in which expression was unchanged across all samples, genes whose expression was increased during infection and disease, and genes whose expression was downregulated during infection and disease. The gene/transcript content of each cluster and their associated biological function are provided in Table S6 in the supplemental material. The major and fine biologies of the members of each of the major transcriptional networks with the numbers of differentially regulated genes are summarized in Tables 3 and 4.

Individuals with active disease and infection exhibit altered expression profiles that are enriched with genes indicative of T cell and NK cell activation. The largest cluster of upregulated genes was MCL2. Manual inspection of this cluster and DAVID v6 analysis indicated that this cluster contained genes typical of NK cells, T cells, and macrophages. Of particular interest was the upregulation of genes associated with NK and T cell ligand-receptor complexes (PTPN6, NCR3-CD3Z, CD244-CD48, NKG2D-DAP10, CD86, CD8, CD28, CD2, CD45, CD4, CD3D, and CD3E), intracellular signal-transduction mediators of the T-cell receptor (TCR) cascade, and nuclear proteins mediating expression of genes controlling the functions of Th1 lymphocytes (TXK and NFATC) and activation of major histocompatibility complex (MHC) class II transcription (RFX5). This cluster also contains members of the killer cell immunoglobulin-like receptors (KIRs), some of which regulate the cytotoxic function of NK cells via MHC class I ligands. These include inhibitors of NK-med-
TABLE 2. Principal biological annotation and functional terms identified by gene set enrichment analysis of gene lists made up only of genes with a high fold change in expression (adjusted P < 0.01)

<table>
<thead>
<tr>
<th>Gene List Comparison</th>
<th>No. of functional annotation terms</th>
<th>No. of functional annotation terms with enrichment score &gt;1.3</th>
<th>Principal GO terms (cellular clusters)</th>
<th>Fold enrichment</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N vs D</td>
<td>27 (46/62)</td>
<td>166</td>
<td>Principal canonical pathways</td>
<td>2.00</td>
<td>1.17E-12</td>
</tr>
<tr>
<td>N vs DI</td>
<td>97</td>
<td>97</td>
<td>Principal canonical pathways</td>
<td>2.00</td>
<td>1.17E-12</td>
</tr>
<tr>
<td>D vs DI</td>
<td>12 (15/43)</td>
<td>77</td>
<td>Principal canonical pathways</td>
<td>2.00</td>
<td>1.17E-12</td>
</tr>
</tbody>
</table>

FIG. 2. Identification of genes with high-degree changes in expression unique to each clinical and infection category, based on >2-fold changes. The biology of each of the gene lists was investigated by testing for enrichment or membership of pathways using DAVID v6, and the major GO terms are listed. The main features reflect those identified based only on high-degree differential expression (Table 2). Influx and activation of cells important in innate immune responses, such as neutrophils and mast cells, are prominent during disease and infection. Humoral and cytokine/chemokine responses typical of adaptive responses appear as expected and are shared between groups. Genes unique to each category were used to generate a small list of genes that could be used to classify participants based on expression profiles and to identify either biomarkers or important expression pathways.

At the genetic level, many of the same genes responsible for innate immune responses appear as expected and are shared between groups. Genes unique to each category were used to generate a small list of genes that could be used to classify participants based on expression profiles and to identify either biomarkers or important expression pathways.

Gene lists were also compared to a published series of breast cancer gene expression profiles. In addition, we identified overrepresentation of transcripts with a role in NK cell biology. Therefore, the fold change expression values for each probe set from MCL2 were imported into the KEGG database defined pathway map for NK cell-mediated cytotoxicity (Fig. 5). This provides a visual representation of the canonical pathway for the activation of NK cells. The greatest changes in expression intensity are evident in the subjects with infection and disease signs. Since analysis based purely on differential expression (2-fold changes) (Tables 1 and 2) also identified enrichment of genes in this pathway, we submitted differential expression (2-fold changes) (Tables 1 and 2) also identified enrichment of genes in this pathway, we submitted
the list of differentially regulated genes to an expression perturbation analysis tool (Onto-tools Pathway Express). The results indicated that the largest predicted effects based on enrichment and fold changes were on the same MCL2-enriched pathways and provided further supporting evidence for stimulation of NK cell-mediated cytotoxicity.

Transcripts typical of neutrophils and interferon signaling form overlapping clusters. MCL3 consists of 271 transcripts, and the average profile of expression is similar to that of MCL2; this can be visualized in Fig. 4, which shows the proximity of these two clusters in the coexpression network. The cluster is dominated by genes typical of leukocyte biology and innate host defense (cytokines, chemokines, and Toll-like receptors [TLRs]). Several members of the Toll-like receptor family (TLR1, TLR2, TLR4, and TLR8) were present in this cluster. The expression signature of cytokine receptors involved in cell-cell signaling is characteristic of the innate immune response that induces transcription of inflammation-related genes (IL8RA, IL8RB, CSF3R, CSF2RB, IL1B, IL1A, IL1R2, and IL6R) as well as those involved in humoral and adaptive immunity. MCL3 also contains receptors for granulocyte-macrophage colony-stimulating factors (CSF3R and CSF2RB) and mediators of neutrophil migration to sites of inflammation (CXCL5 and MMP25). Consequently, genes involved in many of the processes of phagocytosis, cell-mediated cytotoxicity, chemotaxis, or cellular activation of NK cells, monocytes, and neutrophils are represented (CR1, C5AR1, and CFP), including those genes modulating iron metabolism (SLC11A1) and intracellular superoxide production (FPR1, NCF2, and NCF4). MCL3 also has upregulation of genes that

FIG. 3. Classification of individuals based on differentially expressed transcripts. Transformed expression levels are indicated by the color scale, with red representing relatively high expression and green indicating relatively low expression compared to the median expression for each gene across all participants. Hierarchical clustering of 63 gene classifiers obtained from selecting genes with >2-fold changes in expression which are unique to each participant category is shown. Selection of a small subset from the largest group of uniquely expressed genes (1,644 in the disease and infection class) was based on the fold change and absence of correlation of expression with the genes identified for the other unique classes (23 genes in each). Selection of a relatively small list of genes as potential biomarkers is a requirement for the application of computationally efficient methods. A supervised learning algorithm (kNN) was used to test the capacity of the classifier subset of genes to discriminate three classes of participants. Leave-one-out cross-validation of the training set with 63 genes classified the samples with 75% accuracy. The predicted class is indicated by light-color solid rectangles (green for healthy participants, amber/yellow for participants with clinical disease signs only, and red for participants with disease signs with current ocular C. trachomatis infection). Participants from all categories are misclassified; however, misclassified participants are those with disease signs only (9/20). Two patients with clinical disease and infection were classified as healthy uninfected participants, and no healthy participants were classified as diseased and infected.
balance inflammatory responses, such as SLA, LILRB3, SOCS3, and protease inhibitors (SERPINA1). Manual annotation and enrichment analysis therefore suggests that this cluster is comprised of genes expressed predominately in neutrophils.

The expression profiles of MCL8 and -10 are highly similar, are upregulated during disease, and are further upregulated with high C. trachomatis loads (Fig. 6). These clusters have a signature of type I and II interferon (IFN)-regulated genes. These include assembly of MHC class I molecules and the presentation of endogenous peptides to CD8+ T cells (TAP1 and TAP2), induction of apoptosis in infected cells presenting antigen–MHC-I complexes by cytotoxic T lymphocytes (CTL) (PRF1, GZMA, and GZMB), and expression of CTL cell adhesion molecules (ITGAL and CD226). There is a chemokine and cytokine pattern indicative of cell-mediated adaptive immune responses (IL2RA, IL15RA, JAK2, STAT1, STAT4, and IL-22) and IFN-inducible genes (IFN1/2/3/5/7, STAT1/2, and IRF1/7). Other well-established markers of the IFN-γ response, such as GBP1/2/5, were upregulated in this cluster, as, importantly, was INDOL1. INDOL1 is an IFN-γ-induced enzyme that is structurally and functionally similar to the immunomodulatory and antimicrobial indoleamine 2,3-dioxygenase (IDO) (IDO, although not part of the MCL8 coexpression network, was strongly upregulated in those with disease and infection). Further evidence of activation of other genes in known IFN-γ-induced anti-C. trachomatis pathways was ATF5 and WARS.

MCL8 was also enriched with genes which regulate or oppose the activation of Th1 cells and IFN-γ-induced genes, such as SOCS1, TFEC, PTGER2, HAVCR2, and TRAFD1 (FLN29). Negative regulators of CTL and NK cell effector mechanisms such as the serine protease inhibitor SERPINB9 and cell surface receptors (CTLA4 and KLRD1) were also induced. Although similar in profile, MCL10 is enriched with genes associated with type I IFN signaling. These include
interferon-induced genes (IFIH1, IFIT2, IFIT3, IFI15, and ISG20), the interferon regulatory transcription factor IRF7, and genes associated with host defense against intracellular infections (MX1, MX2, OAS1, OAS2, OAS3, OASL, DDX58, RIG-I, and PML).

MCL7 and -9 were comprised of 219 transcripts that were strongly upregulated during disease with infection. MCL9 contained genes characteristic of the B cell lineage. Genes coding for members of the B cell antigen receptor complex (FCRL1, FCRL2, FCRL4, FCRL5, and FCRLA), immunoglobulins, and the CD79A/CD79B heterodimer were present in this cluster, as well as the classical B cell marker CD19. Other immunoglobulin-related genes included those involved in signal transduction pathways (BLK, SPIB, BACH2, TECs, and PTKs) triggered by the B cell antigen receptor (BCR and BRDG1). On the other hand MCL7 was highly enriched with immunoglobulin genes predominately expressed in plasma cells, e.g., IGH@, IGAH1, IGHD, IGHG1, IGHM, IGHV1, IGKC, IGKL@, and IGJ3.

To illustrate the effect of clinical status on the levels of expression in healthy, diseased, and diseased and infected individuals, the geometric mean fluorescence intensities of all the transcripts which belong to a particular MCL or cluster were calculated. The raw intensity levels and the differences in these levels between the clinical groups show the effect of either disease or disease with infection on the level of expression (Fig. 6). The largest changes in expression across a cluster of genes are seen in the diseased-with-infection group. Only rarely did current infection have little effect on expression (e.g., MCL5).

Other clusters significant in the biology of trachoma. MCL4 was made up almost entirely of genes encoding proteins involved in the S phase of the cell cycle, which were upregulated with high infection loads. MCL22 is a cluster of upregulated transcripts which function in lipid metabolism and/or trafficking (APOE, APOC1, SCARB1, and SLC16A1) and innate intracellular defense. Genes in MCL29 were upregulated during disease and infection and included 24 genes coding for extracellular matrix proteins of connective tissue collagens (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, COL5A2, COL6A2, and COL6A3) and extracellular matrix (ECM) glycoproteins (TNC and ASPN). MCL133 was composed of nine transcripts covering eight genes that are characteristic of conjunctival goblet cells, such as the trefoil factors (TFF1 and TFF3) and mucins (MUC5AC and CAPN8). All the genes in this cluster were downregulated in disease and infection.

Downregulated clusters. Genes that were downregulated in disease and infection clustered together in the network (MCL1). MCL1 is the largest single cluster in the network and was enriched for transcripts in the Wnt signaling pathway. Other clusters in this downregulated category include MCL11, -16, -19, -21, -23, and -24.

**DISCUSSION**

In trachoma, unresolved or repeated ocular *C. trachomatis* infection and inflammation causes a progressive fibrotic response that can ultimately result in blindness. The conjunctival scarring observed is more common in those with a severe inflammatory response to infection. This study describes tis-
sue-specific transcriptional networks associated with the response to ocular C. trachomatis infection and inflammation. Previous data suggest that in communities with endemic trachoma, individuals who are at different positions in the disease and infection cycle can be distinguished based on their host response profile (11, 24, 25). We now show that global transcription profiles can also cluster individuals into the same types of disease and infection categories. Furthermore, a re-

FIG. 5. Gene expression within cluster MCL2 is enriched in genes affecting NK cell-mediated cytotoxicity. Fold change in gene expression from MCL2 is shown on the KEGG pathway map (modified from map04650, natural killer cell-mediated cytotoxicity, at the Kyoto Encyclopedia of Genes and Genomes [KEGG] website http://www.genome.jp/kegg/kegg1.html, with permission). KIR2DL, phosphatidylinositol 3-kinase (PI3K), and phospholipase Cγ (PLCγ) depicted in this pathway are represented by multiple gene transcripts in MCL2; in these cases genes are colored according to the transcript with the largest fold change in expression within a multigene or transcript family (KIR2DL4, PI3KCD, and PLCG2). Participants with disease and infection had the greatest changes in gene expression (shown on the map by increasing intensities of red). Onto-tools Pathway Express expression perturbation analysis based only on differentially regulated genes (<0.5 and >2.0 at an adjusted P value of <0.01) comparing healthy subjects with diseased and infected subjects also identifies NK cell-mediated cytotoxicity as substantially affected. An overlapping effect on antigen processing and presentation related primarily to NK cell activity was also observed. The impact factor is the estimated effect on each pathway as a result of changes in transcript expression in addition to enrichment.
Differences in expression intensity of each co-expression cluster

- MCL1
- MCL2
- MCL3
- MCL4
- MCL5
- MCL6
- MCL7
- MCL8
- MCL9
- MCL10

Raw fluorescence intensity units (RMA adjusted)
duced subset of genes that were statistically differentially regulated was also able to separate individuals with increased accuracy into the same categories. This suggests that discrete expression patterns are associated with these different biological states and illustrates the potential to identify an expression signature for those at most risk of severe inflammation and the likely development of scarring sequelae.

An initial wave of polymorphonuclear leukocyte influx is a well-recognized feature of ocular and genital chlamydial infection that is documented in animal models and in human infection (69, 92). The prominent neutrophil gene signature identified in the conjunctiva (MCL3) supports this observation and suggests that the network and enrichment analyses are robust. Indeed, signatures typical of granulopoiesis and neutrophil activation have frequently been identified using genome expression profiling in response to other classes of infection (9, 30, 63). In some of these infections, in which the focus of research has been adaptive “protective” T cells, transcriptome signatures have identified previously hidden type I interferon signaling pathways present in neutrophils and have suggested that they are crucial in the control of infection (9). In some chlamydial infections, depletion of neutrophils delayed the clearance of infection and increased chlamydial shedding in the murine genital tract (5). Neutrophils were found to be a requirement for the recruitment of T cells, particularly CD8+ T cells, to the site of infection (19), and in the lung, an increased influx of neutrophils was associated with a greater chlamydial burden of infection in a susceptible mouse strain relative to a resistant strain (41). The influx of neutrophils in the conjunctiva could be attributed to chemokines such as IL-8 and CXCL1, -2, -5, and -6, all of which were upregulated. Extensive work on the mechanism of neutrophil influx into the cornea following infection with *Onchocerca volvulus* has demonstrated the dependence of this influx on multiple factors originating from cells that are resident in and infiltrate the tissue (28, 46). The resulting neutrophil influx and activation was then responsible for keratitis and corneal haze (29). It therefore seems likely that the control of neutrophil recruitment and activation either by adaptive CD4+ Th cells (e.g., Th17 cells) or by chemokines secreted by infected epithelial cells will be crucial, and this requires further investigation.

We found strong induction of gene expression for IFN-γ (MCL2) and IDO in active trachoma and *C. trachomatis* infection. The production of IFN-γ has a pivotal role in chlamydial disease via control of pathogen growth and replication (68). IFN-γ-induced IDO can inhibit proliferation of *C. trachomatis in vivo* through consumption of the essential amino acid Trp (65, 80). Previously we have also shown upregulation of IDO expression in the conjunctiva of subjects with increasing loads of ocular *C. trachomatis* infection. IDO also has recognized immunoregulatory properties in both human and murine cells (61). Therefore, IDO could control the balance between T cell subset differentiation and local DC priming, suggesting that *C. trachomatis* might exploit IDO expression to induce immunoregulation (5). The identification of inflammatory cells (neutrophils: MCL3) and the upregulation of IFN-γ with T cell receptor signaling pathways (MCL2) together provide evidence that the arrays reveal elements of the expected and previously known cellular and gene expression patterns observed in *C. trachomatis* disease and infection. We suggest that this expected result supports the novel observation that transcripts associated with NK cells and NK cell cytotoxicity, which are found in MCL2, are over represented and suggests an important contribution of NK cells in the response to *C. trachomatis* infection and disease.

There are a limited number of studies in which NK cells have been reported to have a demonstrable impact on chlamydial disease or infection (33, 34), yet depletion of NK cells exacerbated the course of disease and infection in mice (83). Using different gene enrichment approaches, we consistently found evidence for the contribution of NK cell activation and cytotoxicity in the conjunctiva of participants with trachoma. In addition to cytotoxic effects, NK cells could also be a major source of crucial cytokines such as IFN-γ and IL-22. Thus, while NK cells may not be essential for the resolution of infection, they may be critical in the inflammatory process and in the bridge between the adaptive and innate responses. Work with murine models suggests that adaptive CD4+ Th1 cells which produce IFN-γ are required for the resolution of primary infections and that in secondary responses other immune cells can contribute but are not an absolute requirement (87, 90, 91). NK cells can polarize the CD4+ T cell response via dendritic cells (DC), which results in an amplification of IFN-γ production by T cells. NK cells can be helped to produce IFN-γ by other innate cells, such as neutrophils, or by chemokines derived from infected epithelial cells, such as IL-12 and IL-18. It is also well recognized that T cell-derived IL-2 can activate NK cells, and recent evidence suggests that Ag-specific IL-2 from effector memory T cells can in effect immediately “recall” NK cells which degranulate and secrete IFN-γ (35, 36). Thus, in the presence of IL-2, IL-12, and IL-18, the local inflammatory responses are directed toward strong type I and IFN-γ responses (1, 26, 73). Therefore, the strong expression signature of NK cells observed in these conjunctival samples, which reflects those seen after several natural ocular challenge infections or episodes of disease, could be explained by the boosting effect of antigen-specific effector memory T cells. We suggest that this interaction warrants further in vitro study and investigation.

The regulation of NK cell activity is complex; epistatic effects between HLA ligands and KIRs control the activity of NK cells. The level of diversity in the KIR gene system reflects its coevolution with MHC Class I (71), which encodes the ligands of some KIRs (15–17, 31). This diversity is generated by haplotypic, allelic (84, 88), and transcriptional (3) variation that results in a unique KIR expression repertoire. The unusual

FIG. 6. Mean expression levels of all the transcripts in a coexpression cluster for each array are summarized as a single point for each array. Differences in the expression level for each transcript cluster between each clinical and infection category was then tested by analysis of variance. Where differences between the groups are significant, they are indicated with the appropriate *P* value, and this indicates where the changes of expression are attributable.
nature of KIR polymorphism and expression can confound the interpretation of microarray expression studies with respect to individual KIR alleles or genes. Nevertheless, KIR2DL4, which is found on all NK cells, can be regarded as a KIR “framework” locus present (with only rare exceptions) in all KIR haplotypes (62). Since we find this marker and several other markers of NK cells to be highly expressed in the conjunctiva, we suggest that their activity in inflammatory trachoma is significant.

The majority of cells from conjunctival swabs are epithelial cells, and it is well established that infected host epithelial cells are the source of many of the initiating factors that drive inflammation (20, 21). This led Stephens to suggest an alternative paradigm for the pathogenesis of chlamydial diseases (75). We found strong induction of many chemokines, pattern recognition genes (TLRs and NOD [inflammasome] [NOD2, NLRP3, and NLRP7]), and mediators of inflammation (IL1B). Clustering of coexpressed genes and annotation of the gene content of the clusters suggests that infiltrating cells, largely neutrophils, are a major cellular source of many of these factors. The largest fold changes in expression were seen for CXCL5 (epithelial cell-derived neutrophil-activating protein [ENA-78]), -11, and -13. Strong induction of Cxcl13 has been described in the development of murine salpingitis, and this tors. The largest fold changes in expression were seen for

The control of innate responses appears at the center of several networks, and expression of cytokines such as IL-12 and IL-23 is at the fulcrum of protective acquired immune responses. Recent work with mice, which differ in susceptibility to chlamydial pulmonary infection, has shown that the IL-12/IL-23 balance is altered in DC isolated from susceptible BALB/c strains. The excessive IL-23 production observed was suggested to favor the later development of Th17 cells, which were associated with a larger burden of infection in the lungs (41). In our study, the expression pattern of IL23A in the conjunctiva is contained in a transcription network that is characteristic of human epidermal keratinocytes (MCL6). IL-23 promotes inflammatory responses that include upregulation of MMP9 (45), polymorphism of which is associated with scarring trachoma. The overall effect of this polymorphism is complex, since several genes involved in the immune (IFNG and IL-10) and inflammatory (IL-8 and CSF2) responses alter its protective or risk-associated effects (55, 57). Although IL-23 is not involved in Th17 differentiation, it is thought to play an important role in maintaining Th17 effector function (76, 77) and therefore in local tissue inflammation. IL-22 expressed by Th17 or Th22 (22, 82) or NK (13) cells was upregulated in active disease with C. trachomatis infection, whereas its decoy receptor IL22RA2 was upregulated during active disease episodes free of infection. IL-22 allows cross talk between the immune system and epithelial cells, and it has been suggested to have an important role both in host defense and in the pathogenesis of inflammatory skin diseases such as psoriasis. Th22 cells, which are a CD4+ subset that home to the skin and are important in the repair of the epithelial barrier (23), therefore would appear to be important in C. trachomatis infection and immunity. However, the means to identify a clear and separate CD4+ T helper subtype signature within the background of the transcriptome expression profiles are not yet available.

Interpretation of expression studies of disease versus control tissue is often confounded by the very dramatic differences in the cell populations present. An immune response is orchestrated by the activity of numerous leukocyte populations, with each cell expressing genes specific to that population, and in principle the level of these genes in each sample contributes to the overall transcriptional signature. Identification of clusters of coexpressed transcripts that are indicative of these cell types offers the opportunity to attribute some of the differences in expression to the cellular content of the tissue. Furthermore, unlike experimental models of infection where genetically homogeneous animals or cell lines are challenged with equivalent levels of infectious organisms of the same strain, naturally observed infection in the human population results in considerable variation, and this must be taken into account as we interpret the data.

Several observations suggest that transcription network-based analysis provides a powerful approach and an added advantage that allows the identification of genes expressed by particular cell types or those under the influence of the same transcriptional activators. Combining this with pairwise differential gene expression, we show that the major networks of coexpressed and highly regulated genes in the conjunctiva of participants with active trachoma and C. trachomatis infection are dominated by genes involved in innate immune responses and IFN-γ-mediated signaling. We have demonstrated the
prominence of innate responses (NK cells and neutrophils), underpinned by the balance of IL-12/IL-23 and expression characteristic of several CD4+ helper phenotypes (Th1, Th17, and Th22 cells).

ACKNOWLEDGMENTS

We thank the study participants, field workers, and laboratory staff at the Medical Research Council Laboratories in the Gambia for their assistance.

This work was supported by the Wellcome Trust (079246/Z/06/Z) and MRC (United Kingdom) core budget, the Gambia.

REFERENCES

6. Berry, M. P. R., C. M. Graham, F. W. McNab, Z. Xu, S. A. A. Bloch, T. Oni, F. W. McNab, Z. Xu, S. A. A. Bloch, T. Oni, B. Ondondo, T. Kinyari, N. R. Mugo, R. Nguti, and R. C. Brunham. 2004. Chlamydial infection and MRC (United Kingdom) core budget, the Gambia. We thank the study participants, field workers, and laboratory staff at the Medical Research Council Laboratories in the Gambia for their assistance. This work was supported by the Wellcome Trust (079246/Z/06/Z) and MRC (United Kingdom) core budget, the Gambia.

REFERENCES

6. Berry, M. P. R., C. M. Graham, F. W. McNab, Z. Xu, S. A. A. Bloch, T. Oni, F. W. McNab, Z. Xu, S. A. A. Bloch, T. Oni, B. Ondondo, T. Kinyari, N. R. Mugo, R. Nguti, and R. C. Brunham. 2004. Chlamydial infection and MRC (United Kingdom) core budget, the Gambia. We thank the study participants, field workers, and laboratory staff at the Medical Research Council Laboratories in the Gambia for their assistance. This work was supported by the Wellcome Trust (079246/Z/06/Z) and MRC (United Kingdom) core budget, the Gambia.

REFERENCES

6. Berry, M. P. R., C. M. Graham, F. W. McNab, Z. Xu, S. A. A. Bloch, T. Oni, F. W. McNab, Z. Xu, S. A. A. Bloch, T. Oni, B. Ondondo, T. Kinyari, N. R. Mugo, R. Nguti, and R. C. Brunham. 2004. Chlamydial infection and MRC (United Kingdom) core budget, the Gambia. We thank the study participants, field workers, and laboratory staff at the Medical Research Council Laboratories in the Gambia for their assistance. This work was supported by the Wellcome Trust (079246/Z/06/Z) and MRC (United Kingdom) core budget, the Gambia.

REFERENCES


Editor: R. P. Morrison