Sequencing, annotation and characterization of the influenza ferret infectome

Alberto J. Leóna,b,c, David Bannerb, Luoling Xub, Longsi Ranb, Zhiyu Pengc, Kang Yib, Chao Chenb, Fengping Xu, Jinrong Huangc, Zhen Zhaoa, Zhen Linb, Stephen H.S. Huangb,e, Yuan Fangab,ce, Alyson A. Kelvinfo, Ted M. Rossd, Amber Farooquiak, David J. Kelvinb,ce,h, #

aDivision of Immunology, International Institute of Infection and Immunity, Shantou University Medical College, Shantou, Guangdong, China
bDivision of Experimental Therapeutics, Toronto General Research Institute, University Health Network, Toronto, Ontario, Canada
cGuangdong Provincial Key Laboratory of Infectious Diseases and Molecular Immunopathology, Shantou, Guangdong, China
dResearch & Cooperation Division, BGI-Shenzhen, Shenzhen, Guangdong, China
eDepartment of Immunology, University of Toronto, Toronto, Ontario, Canada
foImmune Diagnostics & Research (IDR), Toronto, Ontario, Canada
gCenter for Vaccine Research and Department of Microbiology and Molecular Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
hDipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Sardinia, Italy
#Address correspondence to: David J. Kelvin, dkelvin@jidc.org

Running title: characterization of the ferret infectome
Abstract

Ferrets have become an indispensable tool in the understanding of influenza virulence and pathogenesis. Furthermore, ferrets are the preferred preclinical model for influenza vaccine and therapeutic testing. Here we characterized the influenza “infectome” during the different stages of the infectious process in ferrets with and without prior specific immunity to influenza. RNA from lung tissue and lymph nodes from infected and naïve animals were subjected to next-generation sequencing, followed by de novo data assembly and annotation of the resulting sequences; this process generated a library comprising 13,202 ferret mRNAs. Gene expression profiles during pdmH1N1 influenza infection were analyzed by Digital Gene Expression and solid support microarrays. As expected during primary infection, innate immune responses were triggered in the lung tissue; meanwhile, in the lymphoid tissue, genes encoding antigen presentation and maturation of effector cells of adaptive immunity increased dramatically. After 5 days post-infection, the innate immune gene expression was replaced by the adaptive immune response which correlates with viral clearance. Reinfection with homologous pdm influenza virus resulted in a diminished innate immune response, early adaptive immune gene regulation, and a reduction in clinical severity. The fully annotated ferret infectome will be a critical aid to the understanding of the molecular events that regulate disease severity and host-influenza virus interactions among seasonal, pandemic, and highly pathogenic avian influenzas.
Introduction

Influenza virus is responsible for a great burden of disease in humans. The infection follows distinct epidemiological patterns, including the yearly influenza season (38), off-seasonal pandemic outbreaks (11, 15), and sporadic cases with swine flu (10) or avian influenza (42). The clinical severity of influenza infection is highly variable and is determined by multiple factors; these include the pathogenic potential of the virus strain (19), the immune memory (14, 28), age (18), interactions with the microbial environment (20, 39), and also, genetic host factors that confer higher susceptibility (5, 13).

Experimental influenza infection in ferrets has a close resemblance with the human disease (4) and it is widely used to study the pathogenicity of newly isolated strains (19, 22) and to test vaccine efficacy (3, 8). Last but not least important, infection in ferrets is also used as a disease model to study the host responses. Previously, microarray analysis allowed us to profile the gene signatures in ferrets experimentally infected with different influenza subtypes (6, 7), determine gene expression profiles after direct treatment with antiviral interferons (9), and also, to characterize the shift from the innate to the adaptive immune responses in influenza infected ferrets (35). A better understanding of the pathogenic mechanisms of influenza requires in-depth profiles of previously known and unknown components by integrating transcriptome-wide gene expression profiles with pathological outcomes.

We propose the use of the term *infectome* to designate an integrated view of the regulated genes within the context of the host functional components of which they are part during the infectious process. This term was previously applied to the gene regulation in the infected cells alone (30); however, here the concept of *infectome* will also include non-infected cells in the diseased tissue, which are under the influence of the same local regulatory factors as the infected ones, and other body compartments that actively participate in the immune responses, such as blood, lymph nodes or spleen.

In this paper, we aimed to characterize the *infectome* during pdmH1N1 infection in experimentally infected ferrets. First, we generated an annotated library of ferret sequences, which can be used to perform gene expression analysis by next-generation sequencing and/or to design PCR primer sets of ferret genes not previously characterized. Next, we employed digital gene expression (DGE) (29) to analyze differentially expressed genes in the lung tissue and lymph nodes from ferrets infected with A/Mexico/4108/2009; also, we studied the gene expression during influenza infection in ferrets that underwent a prior infection and presented protective immune memory. Microarray data was included to track the evolution of gene categories during the time-course of the infection. Finally, we integrated...
the gene expression profiles within the functional components of the host immune responses that lead
to the resolution of the infectious process, either through a process with clinical manifestations or in a
completely asymptomatic manner.

Materials and Methods

Experimental influenza infection in ferrets, sample collection and RNA extraction

In a first experiment, male ferrets purchased from Triple F Farms (Sayre, PA, USA) were infected with
1×10^6 50% egg infectious doses (EID50) of influenza A/California/07/2009 in an ABSL-3 facility at the
University of Pittsburgh (Pittsburgh, PA, USA), as previously described (34). Animals were sacrificed on
days 1, 3, 5, 7 and 14 days post infection (d.p.i.), together with a group of four uninfected ferrets. Lungs
and lymph nodes were excised and stored in RNAlater (Ambion, Austin, TX, USA) and the RNA was
purified using TriPure (Roche, Indianapolis, IN, USA). A second round of ferret infections (in-house ferret
breeding) was performed using 1×10^6 EID50 of A/Mexico/4108/2009 in a BSL2+ facility at University
Health Network (Toronto, Ontario, Canada). Lung tissues and lymph nodes were collected from 3 ferrets
5 d.p.i. and also from 3 uninfected animals. Finally, a group of ferrets were infected with 1×10^6 EID50 of
A/Mexico/4108/2009 and re-challenged 21 days later using the same virus strain and infectious dose;
lungs and lymph nodes were collected on days 0, 1, 3, 5, and 7 after infection. The RNA from the
experiments with A/Mexico/4108/2009 was purified with the RNAEasy Mini kit (Qiagen, Toronto, ON,
Canada). All procedures were approved by the Institutional Animal Care committee and the Biosafety
committee of each institution.

High-throughput Illumina sequencing, De novo assembly, annotation of contigs and generation of a
database of ferret mRNA sequences

To characterize the sequences of the ferret transcriptome, 7 RNA samples from 6 different ferrets from
the California/07 infections were subjected to deep sequencing: control (lymph nodes), day 1 (lung), day
3 (lung and lymph nodes), day 5 (lung), day 7 (lymph nodes), day 14 (lymph nodes). Although initially a
larger number of RNA samples were selected, only those meeting the quality requirements were
sequenced. Construction of the cDNA libraries and Illumina sequencing for RNA-seq and digital gene
expression (DGE) were performed at BGI-Shenzhen (http://en.genomics.cn). RNA-seq was performed as
previously reported (45). Briefly, polyadenylated RNA was isolated using Sera-Mag Magnetic Oligo (dT)
Particles (Thermo Scientific, Rockford, IL, USA). Fragmentation of the mRNA was performed by
incubation with divalent cations at 94℃ for 5 minutes. The double-stranded cDNA was synthesized with
the SuperScript Double-Stranded cDNA kit (Invitrogen Life Technologies, Carlsbad, CA, USA) using
random hexamer primers (Illumina, San Diego, CA, USA), and subjected to end-repair and
phosphorylation by incubation with a mix containing T4 DNA polymerase, T4 PNK and Klenow Enzyme.
Adenylation of 3’ ends was performed by Klenow exo treatment (3’ to 5’ exo minus) and followed by
adaptor ligation (Illumina). The products were subjected to agarose electrophoresis; gel bands of
200±25 bp were excised and the products were enriched in a 15-cycle PCR reaction using the PCR
primers PE 1.0 and PE 2.0 (Illumina). The quality of the cDNA library was assessed on an Agilent
Technologies 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA). Paired-end 90bp sequencing of the cDNA
library was performed in an Illumina Genome Analyzer IIx according to the procedures provided by the
manufacturer.
The resulting 90bp sequencing data was filtered using an Illumina’s chastity threshold of 0.6 on the first
25 cycles and a quality filter of 10% Q<20. De novo assembly was performed with SOAPdenovo
(http://soap.genomics.org.cn/soapdenovo.html) software, using the default settings except for a K-mer
value of 29, which was found to be the optimal setting. Next, paired-end information was used to merge
contigs followed by a final step of gap filling. To select biologically meaningful contigs and determine the
correct orientation, the resulting contigs were aligned with BLASTx with a similarity threshold of
e_value<10^{-5} using NCBI’s non-redundant (Nr) protein database (ftp://ftp.ncbi.nih.gov/blast/db), Swiss-
Prot protein database (http://www.uniprot.org) and Cluster of Orthologous Groups database
(http://www.ncbi.nlm.nih.gov/COG). To annotate the contigs, the mRNA sequences from Canis
familiaris (domestic dog), Ailuropoda melanoleuca (giant panda bear), Homo sapiens (human), Bos
taurus (cow) and Sus scrofa (pig) were extracted from the Reference Sequence mRNA database
were generated using the mRNA from these species, and de novo assembled contigs were aligned using
BLASTn (version 2.2.25+), e_value<10^{-12} and word_size=18; the annotations were considered valid only
when having coincidental gene orthologs from at least two different species. During the alignment
process, some sequences were trimmed to remove non-matching terminal regions of the contigs, which
were incorporated during the de novo assembly process and possibly derived from immature mRNAs.
For those mRNA sequences that were covered by multiple non-contiguous contigs, Iliad Assembler
(http://www.ferretscience.org/2012/02/iliad-assembler.html) was used to place the contigs in the
correct relative position and fill the unresolved gaps with NNN spacer sequences. The final mRNA ferret
library included the newly generated sequences; additionally, some of these were replaced by previously published ferret sequences that were more informative.

**Analysis of gene expression profiles by DGE**

Digital Gene Expression (DGE) was used to analyze the gene expression profiles in the lung tissue and lymph nodes from ferrets infected with influenza A/Mexico/4108/2009, 5 d.p.i., with respect to uninfected controls. Day 5 was chosen because this time point had previously shown significant lung pathology, weight loss, and lung viral load during infection with A/Mexico/4108/2009 in ferrets (19). To generate the tag libraries, RNA binding to oligo(dT) magnetic beads and double-stranded cDNA synthesis were performed as mentioned previously. Next, tag generation and adaptor ligation were performed by using the Digital Gene Expression Tag Profiling Kit (Illumina) according to the manufacturer’s instructions. Briefly, double-stranded cDNA chains were incubated with NlaIII endonuclease and adaptor 1 was ligated to the free 5' end. Digestion with Mmel endonuclease detached the tags from the beads. Single-end 49bp sequencing of the cDNA library was performed in an Illumina HiSeq2000 according to the procedures provided by the manufacturer.

The short-read aligner Bowtie v0.12.7 (25) ([http://bowtie-bio.sourceforge.net/index.shtml](http://bowtie-bio.sourceforge.net/index.shtml)) was used to align the DGE tags with the genes contained in the mRNA ferret library. First, a Bowtie index was built using the ferret mRNA sequences. Next, alignments were performed with Bowtie using the parameters -a --best --strata and -S. The resulting tag counts per gene were loaded into Multi Experiment Viewer (MeV) version 4.7.4 (36) and the differential gene expression was computed using DEGseq with MA-plot-based method with Random Sampling model (DEGseq-MARS) (44). Differences in gene expression were considered statistically significant only when the sum of tag counts per gene ≥25 (including 3 controls and 3 infected samples 5 d.p.i.) and DEGseq’s p<0.05 and fold-increase >1.5. Significantly regulated genes were subjected to hierarchical clustering with MeV and functional gene classification using the DAVID annotation tool version 6.7 ([http://david.abcc.ncifcrf.gov/](http://david.abcc.ncifcrf.gov/)) (17). Microarray analysis of tissues from ferrets infected with A/California/07/2009 was performed as indicated in the supplementary methods. Datasets are publicly available at NCBI’s Gene Expression Omnibus (accession numbers GSE17079 and GSE36019).

**Analysis of gene expression by real-time PCR**

PCR primers for the ferret mRNA CCL5, CD79A, CD8A, CD80, CD86, IGHG and IGHM were designed using Roche Probe Designer 2.0 (Roche Applied Sciences) software; additionally, previously published primer sets for ferret CXCL10, beta-actin and the influenza segment 7 (MP) were used (suppl. Table S1). Reverse
transcription was performed with ImProm-II Reverse Transcription System (Promega, Madison, WI, USA). Real-Time PCR was performed in triplicate, using the Power SYBR Green kit (Applied Biosystems, Foster City, CA, USA) in an ABI-Prism 7900HT (Applied Biosystems). Results were normalized to the house-keeping gene beta-actin. Statistical analysis of real-time PCR results was performed using Graph-Pad Prism 5.0.

Results

Annotation of the Ferret Influenza Infectome

To identify genes that comprise the influenza infectome, RNA was extracted from the lung tissue (1, 3, 5 and 7 days post-infection (d.p.i)) and lymph nodes (3 and 14 d.p.i., and uninfected controls) from ferrets infected with A/California/07/2009. Illumina sequencing resulted in the generation of 20 million 90bp paired-end sequences per sample; the sequencing data was submitted to the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra/?term=SRP010102). The resulting short-read sequences, which were derived from the analysis of 7 RNA samples from lung tissue and lymph nodes at different time-points after infection with A/California/07/2009, and also a mock-infected ferret, were subjected jointly to de novo assembly; the resulting contigs were aligned with the protein sequences from several databases to select biologically meaningful contigs. To annotate these contigs, they were aligned with respect to the reference mRNA from several mammal species (supplementary methods); the resulting data set containing 32,350 contigs was submitted to GenBank’s Shotgun Transcriptome Assembly (TSA) database (http://www.ncbi.nlm.nih.gov/nuccore?term=JP004807:JP037156[PACC]). Of these, 19,184 were annotated contigs and 13,166 were non-annotated; the mean length was 654.6bp, minimum length of 200bp and maximum of 8,016bp (Fig. 1A). 9,501 ferret mRNAs were covered by a single annotated contig, and 9,683 annotated contigs belonged to genes covered by two or more non-contiguous contigs. The quality of the newly assembled sequences was assessed by comparing them with a number of previously published mRNA (suppl. Table S2) and genomic (suppl. Table S3) ferret sequences. Finally, a ferret mRNA library containing 13,202 annotated sequences was generated; 13,109 mRNA sequences were derived from the newly assembled contigs, including sequences formed by multiple non-contiguous contigs separated by NNN spacers; additionally, 93 previously published ferret mRNA sequences that were more informative than the newly generated ones were incorporated into
the library. A FASTA file containing the mRNA ferret library is publicly available
(data: http://www.ferretscience.org/2012/02/ferret-transcriptome-project.html).

Biological breadth of the ferret transcriptome

Next, the consistency and the biological breadth of the ferret mRNA library were tested. The sequences
were aligned with the Refseq_rna database and the percent homology values were obtained. From all
the species whose sequences were present in the refseq_rna database, Ailuropoda melanoleuca (giant
panda bear) and Canis lupus familiaris (domestic dog) showed the highest degree of homology with
ferret, with median homology values of 94.7% and 93.57%, respectively (Fig. 1B); these homology values
are not surprising given that these three species belong to the same order (Carnivora) and suborder
(Canifornia). Although the precise taxonomic relationships among the families in the order Carnivora are
still under investigation, previous phylogenetic studies showed that ferrets have a closer relationship
with giant panda than with dog (47, 48). Next, the homology values between ferret and giant panda
were subdivided by protein domains (Fig. 1B); this analysis revealed which domains are under a higher
or lower evolutionary pressure than the average for the total transcriptome. We also obtained an
overview of the biological breadth of the ferret mRNA library. From all the genes contained in the ferret
mRNA library, 11,528 gene symbols were mapped by the DAVID annotation tool. Fig. 1C shows an
overview of the Gene Ontology analysis covering the biological process and molecular function
categories. To further evaluate the level of detail with which the ferret mRNA library was able to cover
relevant biological signaling pathways, the gene identifiers were loaded into Ingenuity Pathway Assist
(release version Winter-2011); the software was able to map 11,580 gene identifiers, and these were
able to cover comprehensively the gene members of different pathways (Fig. 1D).

Study of gene expression in lung tissue and lymph nodes by DGE

To perform the analysis of the gene expression profiles during influenza infection, lung tissue was
obtained from ferrets infected with A/Mexico/4108/2009, 5 d.p.i (n=3), and also from control animals
(n=3). The clinical signs observed during the infection were featured by very mild fever during days 1
and 2 post-infection, marked hypothermia during days 7-10, and a mere 3.4% weight loss on day 2
(suppl. Fig. S1). The severity of the clinical signs with A/Mexico/4108/2009 in the present study is lower
than that previously reported for this strain (19); this discrepancy might be attributed to viral
attenuation that could have occurred in some viral stocks, or to other unidentified sources of
experimental variability. The RNA was purified and subjected to DGE Tag sequencing and the resulting
sequencing data files were deposited in the SRA (http://www.ncbi.nlm.nih.gov/sra/?term=SRP011016).

An average of $11.98 \times 10^6$ 20bp-tags per sample were produced (maximum $12.38 \times 10^6$, minimum $11.46 \times 10^6$). Tags were aligned to the sequences in the ferret mRNA library and the tag counts per gene were used to compute differences in gene expression by DEGseq analysis. The overall gene expression is markedly upregulated in the lung tissue of ferrets infected with A/Mexico/4108/2009, 5 d.p.i.; 2,926 genes were significantly upregulated and 637 genes were downregulated (suppl. Table S4). Interferon-stimulated genes were markedly upregulated, including CXCL10, OAS1, IRF1 and RSAD2 (Fig. 2A), which have a prominent role in the antiviral responses. The increased levels of different cytokines (IL1B, IL6, IL8, IL27 and IFNG) and chemokines (CXCL11, CXCL16, CCL3, CCL4, and CCL5) indicate the presence of active cell recruitment, inflammation, and immunomodulatory processes in the infected lung tissue. Additionally, DGE showed the presence of a rich variety of immune related cell receptors (Gene Ontology: immune system process plasma membrane receptor activity), such as CD2, CD3E, CD4, CD5, CD8A, CD8B, CD27, CD40, CD79A, CD80 and CD86 (Fig. 2A). DGE analysis of lymph nodes from ferrets infected with A/Mexico/4108/2009, 5 d.p.i. shows markedly downregulated gene expression with 2,970 down-regulated genes that covered a wide array of general and immune-related categories. Only 425 genes were upregulated, which included interferon-stimulated genes (suppl. Table S4 and Fig. 2B).

Evolution of gene category enrichment during the time-course of influenza infection

To characterize the overall evolution of the gene expression, the enrichment of functional gene categories was analyzed. Ferrets were infected with A/California/07/2009 and the gene expression in the lung tissue and lymph nodes was analyzed by Affymetrix GeneChip Canine Genome 2.0 microarrays (supplementary methods). In the lung tissue, upregulated genes part of the “cellular process” category showed a steady increase during the first part of the process, reaching a plateau 5 d.p.i. that was maintained through the rest of the process (Fig. 3A). The functional category “Fc gamma R-mediated phagocytosis” was significantly enriched on 5 d.p.i. and declined shortly thereafter; the presence of these upregulated genes might be an indication of higher numbers of neutrophils, macrophages and monocytes in the lung tissue, and this peak might correlate with the evolution of the inflammatory response (Fig. 3B). On the contrary, the “T cell receptor signaling pathway” category was significantly enriched on 7 and 14 d.p.i. (Fig. 3C), indicating that the onset of the adaptive immune response in the lung tissue occurred at a late stage of the process. The situation observed in the lymph nodes was quite different, evolving from marked enrichment of upregulated genes on 3 and 5 d.p.i., to marked...
downregulation on 7 and 14 d.p.i. (Fig. 3D); this trend was also observed in other functional categories (Fig. 3 and f).

Evolution of lung host gene expression during influenza infection and reinfection

During influenza epidemics and outbreaks, individuals display different degrees of immune memory acquired from past influenza seasons or sequential reinfections during a short period of time. In an attempt to model the effect of protective immunity, we performed an infection-reinfection protocol with homologous pdmH1N1. To determine the impact of influenza infection and reinfection on temporal changes in host immune gene expression, we generated primer sets for genes associated with select innate and adaptive immune responses (suppl. Table S1 and supplementary methods) and monitored expression of these genes over time in ferret lung and lymph nodes from influenza infected and re-infected animals. CXCL10, CCL5, IGHM, IGHG, CD8A, CD79A, CD80, and CD86 expression levels in lung and lymph nodes were determined at 0, 1, 3, 5, and 7 days following initial infection with pdm H1N1.

Three weeks following the primary infection, an additional group of ferrets were homologously reinfected with pdmH1N1 and expression levels of innate and adaptive immune genes were monitored in lung and lymph nodes (Fig. 4). As shown in Fig. 4, lung tissue expression of CXCL10 and CCL5 increased as early as day one and was maintained through day 7 following initial infection with pdmH1N1; these results are similar to the observations reported in previous studies (35). Adaptive immune gene (IGHM, IGHG, and CD8A) expression was increased on days 5 and 6 post infection and correlated with a decrease in viral load. A very different temporal pattern of host immune gene expression was observed following reinfection. CXCL10 and CCL5 expression was delayed until day 5 and the levels were substantially reduced compared with the expression levels observed following primary infection for these two genes. Furthermore, Peak levels for expression of IGHM, IGHG, and CD8A were observed earlier (day 5) following reinfection compared to peak levels observed in primary infection (day 7) (Fig. 4).

In the lymph nodes, significant increases in CXCL10 and CCL5 could be observed within one day following primary infection (Fig. 4C and E). Significant increases in expression of CD8A and IGHM could be observed in lymph nodes by day 3 post infection, whereas IGHG expression was significant on day 7 (Fig. 4B and F). Interestingly, reinfection resulted in no significant increase in gene expression of CXCL10, CCL5, IGHM, IGHG, and CD8A. There was, however, a decrease associated with IGHM and CD8A.

Increases in the expression of CD80 and CD86 were observed in both lungs and lymph nodes of infected animals but no significant increases were observed in re-infected animals (suppl. Fig. S2). None of the genes analyzed showed significant increases in the lymph nodes during the reinfection process. This
data indicates that homologous re-challenge of pdmH1N1 in the lungs of ferrets results in an attenuated interferon/innate response and an efficient T cell/B cell driven response.

Discussion

Characterization of the ferret transcriptome

Analysis of gene expression is a key approach to the study of pathogenic processes during influenza infection in ferrets; however, the characterization of the ferret at the molecular level has occurred at a slower pace than with other more commonly used animal models. A number of ferret mRNAs were sequenced independently as part of different studies. The first initiative involving high-throughput sequencing of ferret mRNAs consisted of sequencing a library that comprised 4,209 cDNA clones from ferret (GenBank accessions GD180851-185059, non-annotated gene products); given the considerable length of those sequences and the richness in immune-related genes, this library can still be considered a valuable resource. More recently, Bruder and colleagues have performed Roche 454 sequencing analysis on the ferret mRNA (GenBank accessions EZ456440-EZ516573, non-annotated gene products) (6). In the present study, we generated a library of ferret mRNA sequences that broadly covers the biological processes that take part during influenza infection; these sequences can be accessed as contigs (GenBank accessions JP004807-JP037156, annotated gene products), or as a ferret mRNA library (supplementary methods) (http://www.ferretscience.org/2012/02/ferret-transcriptome-project.html). These newly-generated resources will allow others to interpret next-generation sequencing reads from ferret experiments and translate them into biologically meaningful gene expression profiles.

At present, three different approaches to analyze transcriptome-wide gene expression in ferrets have been described: Affymetrix canine microarrays (7), Agilent microarrays with ferret-specific probes (6), and the next generation sequencing approach employed in the present study. Although a direct comparison of the performance of the three platforms is yet to be made, we predict that NGS will prevail in the mid-term because of its platform independence and, since this technique does not rely on chemically pre-defined probes, NGS will probably present the highest levels of specificity for genes that show high sequence variability among ferret breeds. Nonetheless, wide adoption of NGS to analyze gene expression in ferrets requires further characterization of the ferret transcriptome and the standardization of a unique set of reference sequences.

In this study, RNA samples were analyzed by next-generation sequencing (46) and the resulting 90bp paired-end sequences were subjected to de novo assembly (45); this method resulted in the generation
of sequence contigs by searching for overlapping regions among the short-reads and using the paired-end information to further organize these contigs (26). The newly generated contigs were annotated by searching the gene orthologs of highest sequence similarity in other mammal species. To achieve broad representation of the different functional components of the immune system, the mRNA sequences were derived from infected lung tissue at different time-points after infection; furthermore, the inclusion of lymph nodes from animals undergoing the infectious process ensured the abundance of leukocyte-specific genes. We proved that the resulting library of ferret mRNA sequences had sufficient biological breadth to depict in detail the functional categories that participate in the infectious process (Fig. 1B and C); this library allows one to perform broad gene expression analysis and statistically sound category enrichment of the main components of the immune system. More comprehensive characterization of the ferret mRNA sequences will require additional studies to reveal the genes that are tissue-restricted and also those that only participate in highly specific biological processes. Further characterization of blood-expressed mRNAs in ferrets would be of interest given the relevance of this body compartment as a source for biological markers of disease; also, better knowledge of the sequences of genes expressed in the brain tissue would facilitate the study of neuro-invasive forms of influenza infection (31). The study of the splicing variants is also necessary to increase the level of completeness of the ferret infectome, not only of the naturally occurring isoforms expressed under immunological stress, but also the variations in the mRNA caused by the presence of the virus (41).

Host immune responses in infected ferrets

Virus levels in the lung tissue peaked as soon as day 1 post-infection, were maintained until day 5, and finally, declined abruptly thereafter as a consequence of the host responses (Fig. 4A and Fig. 5A). Previously, we proposed a bi-phasic model of immune responses in ferrets infected with pdmH1N1, that featured an early activation of the innate immune response that is superseded upon activation of the adaptive immune response in a later stage (34). As part of the first phase, the host mounts a strong antiviral response whose hallmark is the expression interferon-stimulated genes (ISGs) such as OAS1, IRF1, ISG15 and RSAD2 (Fig. 2A and Fig. 5); this antiviral response is critical to locally control the spread of the virus (37) and also participates in peripheral cell activation and lymphocyte priming (33). Failure to assemble an effective interferon response has detrimental effects for the host, as shown by in vitro and in vivo studies (24). Chemokines induce cell trafficking towards the infection sites of different immune cell types such as monocytes, neutrophils, NK cells and different lymphocyte subsets, in a selective manner (2). Ioannidis et al. found that chemokines CCL5, CXCL9, CXCL10 and CXCL11 were
expressed in primary human airway epithelial cells (hAECs) infected in vitro with influenza virus (21), which may closely portray the early chemokine response of the infected cells in the lung tissue. Once the inflammatory process has been established, not only the infected cells but also highly activated immune cells produce chemokines, including the abovementioned CCL5, CXCL9, CXCL10 and CXCL11, and in addition, CCL3, CCL4, CCL7, CCL28 and CXCL16 (Fig. 2A). Better characterization of the host chemokine responses using different strains of influenza virus (23) and during different stages of the process (32) would shed light on the implication of different cell types in the context of a dynamic process; furthermore, the modulation of the chemokine profiles has great therapeutic potential for the treatment of severe cases of influenza by using selective blockers of chemokine receptors (7) or immunomodulators that directly affect the chemokine profile (1). Later activation of the adaptive immune response at 7 d.p.i. was previously supported by increased mRNA levels in blood cells of granzyme A, B and H, and the presence of anti-pdmH1N1 antibodies in serum (34). Here, we found that the local mRNA expression of immunoglobulin heavy Mu (IGHM) and CD8A were significantly upregulated as early as 5 d.p.i. and they increased steadily thereafter (Fig. 4B and D), which suggests that the migration of mature antigen-specific cytotoxic lymphocytes and antibody-producing lymphocytes in the infected lung tissue begins 5 d.p.i., hence initiating the effector adaptive responses (Fig. 5).

Lymph nodes play a key role in the adaptive immune responses. Upon infection, dendritic cells become loaded with viral antigens at the infection sites and migrate to the lymph nodes, where they present these antigens to CD8+ cells and B lymphocytes; consequently, these cells undergo activation, maturation and expansion to initiate the effector phase of the adaptive response (43). The presence of chemotactic signals can explain the migration of circulating lymphocytes to the lymph nodes (12), which in our model of infection, occurs as early as day 1 post-infection and lasts throughout the whole process, including but not limited to CXCL10 and CCL5 (Fig. 4C and E). Upregulation of CD8A and IGHM genes (Fig. 4B and D), which is significant on day 3 and thereafter, is an indication of early accumulation of lymphocytes in the lymph nodes. On the other hand, IGHG (immunoglobulin heavy Gamma) is significantly increased at 7 d.p.i., which is consistent with the time required for the switch of immunoglobulin isotypes (Fig. 4F). In terms of global gene expression, we found that lymph nodes of ferrets infected with A/California/07/2009 undergo broad gene upregulation followed by a plateau stage, and finally, nearly complete gene downregulation is observed at 7 d.p.i. (suppl. Table S4 and Fig. 3D). This decay of the gene activity in the lymph nodes may correspond with the end of antigen presentation processes and the departure of mature CD8+ and B lymphocytes towards the infection site.
On the other hand, infection with A/Mexico/4108 at 5 d.p.i. shows overall gene downregulation while maintaining prominent ISGs and chemokine expression (suppl. Table S4 and Fig. 2B), which is suggestive of a more progressive transition towards downregulation. Taken together, these results are suggestive of variability in the time-frame of lymph node activity, which might be related to differences in the progression of the infectious process. Additionally, differences in the chemokine profile are likely to affect lymphocyte taxis towards the lymph nodes and later release, also referred as the “stop and go” model (12). A complete understanding of gene expression profiles in lymph nodes and lungs from infected animals will require the analysis of cellular composition and trafficking in each organ as well as peripheral blood. Currently the lack of antibodies directed at ferret leukocyte and lymphocyte cell surface markers makes these types of analyses difficult. Our group, however, is engaged in generating a panel of reagents that will aid in trafficking studies during influenza infection.

To complete the immunological picture, it is necessary to consider not only the infection in naive animals but also in those with previous exposure to influenza antigens. As expected, during homologous re-challenge with A/Mexico/4108/2009, ferrets failed to develop observable clinical signs (suppl. Fig. S1) and the levels of virus remained low throughout the infection process, although marginally detectable levels were present at 7 d.p.i. (Fig. 4A). These results suggest that pre-acquired immunity was able to contain the viral spread during the first stages but was insufficient to prevent the establishment of the infectious process. During the first stage of the disease the levels of the ISGs, CXCL10 and CCL5 in the lungs remained unchanged, suggesting that upregulation of ISGs and the inflammatory infiltrate were completely absent, respectively; nonetheless, during the second stage, CXCL10 and CCL5 together with markers of the effector adaptive immune response such as IGHM and CD8A, were upregulated, demonstrating the essential role of the adaptive response to clear the virus, even in scenarios of infection with efficient pre-established antibody responses (Fig. 5). Interestingly, the expression of key genes in lymph nodes remained unchanged or even was downregulated during the reinfection (Fig. 4). These genes included immunological markers, suggesting a complete lack of involvement in the rechallenge immune process; hence effector adaptive responses in the lung may have originated in peripheral compartments other than the lymph nodes, or they may reflect the expansion of the resident CD8+ cells (27) and B lymphocytes in the lungs themselves.

In this study, we generated an annotated library of ferret mRNA sequences which opens the door to genome-wide expression profiles by next-generation sequencing in a platform-independent manner. Additionally, DGE data was able to confirm the implication of different signaling and executioner proteins.
molecules which were previously known to participate in the process; the picture was further
complemented with a rich collection of surface markers (Fig. 2). Here we propose a multi-component
view of the immune processes that takes place in the host during an influenza infection. This model will
allow the integration of additional factors that determine the severity of influenza infections, and also
provide a better understanding of the effects of different immunization profiles. Comparative analysis
using different host species, viral agents, presence of bacterial co-infections and other co-morbidities,
and the use of small perturbation methods to profile steady states and gene networks (16), will help to
better characterize the components and the behaviour of the different functional processes that
comprise the host responses.

Acknowledgements

This work has been financially supported by the Li Ka-Shing Foundation, the Guangdong Provincial Key
Laboratory of Infectious Diseases and Molecular Immunopathology and grants from IDR. The funding
sources had no influence on the analyses or interpretation of the results presented in the paper. We
would like to thank the staff from the Animal Facility of University Health Network for their help with
the animal experiments.
References


Figure legends

Figure 1. Overview of the ferret mRNA library generated by next-generation sequencing and de novo assembly. A) Distribution of sequences length in the collection of de novo assembled contigs (left) and the ferret mRNA library (center). The percent coverage of the genes in the ferret mRNA library was estimated with respect to canine mRNA sequences from the refseq_rna database (right). *Unresolved “N” positions were excluded when calculating the length of ferret mRNA sequences. B) Percent homology of mRNA sequences by group. Left panel: transcriptome-wide analysis of ferret sequences with respect to giant panda bear, domestic dog, human and mouse; ***p<1E-12 (one-sided T test, hypothesis value=100). Right panel: percent homology between ferret and giant panda bear (full transcriptome median value=94.7), genes were classified according to protein domains present in their mRNA sequences; *p<1E-2 and **p<1E-4 (one-sided T test, hypothesis value=94.7). The number of sequences included in each analysis is shown in the column legends. Boxes represent median and interquartile range and whiskers represent 10th and 90th percentile. C) Gene Ontology classification of the genes present in the library. DAVID annotation tool was used to perform the classification. D) Proportion of sequences of the newly-generated ferret mRNA library that were assigned to the Leukocyte extravasation signaling and T cell receptor signaling pathways, respectively, by Ingenuity Pathway Assist (IPA) analysis. The number of mapped genes and the total number of pathway components are indicated between brackets; green-filled boxes represent components covered by one or more genes present in the ferret mRNA library, grey-filled boxes represent non-covered components.

Figure 2. Digital Gene Expression (DGE) analysis during influenza H1N1 infection in ferrets. The RNA was extracted from A) the lung tissue and B) lymph nodes from ferrets infected with 10^6 EID_{50} of A/Mexico/4108/2009 at 5 days post infection (d.p.i). Genes were considered significantly up- or down-regulated when DEGseq’s p<0.05, biological significance filter >1.5-fold and the sum of tag counts per gene ≥25. Significantly regulated genes were subjected to functional classification, and, for each group of genes, hierarchical clustering was performed.

Figure 3. Evolution in the gene enrichment of functional categories during influenza infection. RNA was extracted from A-C) the lung tissue and D-F) lymph nodes from ferrets infected with 10^6 EID_{50} of A/California/07/2009 at different time-points (three animals per time-point). Gene expression was analyzed by using Affymetrix GeneChip Canine Genome 2.0 microarrays and genes were considered significantly regulated if Student T test’s p<0.05 and fold variation >±1.5. The lists of up and down-
regulated genes on each time-point were loaded in DAVID annotation tool for functional gene classification. Statistical significance of gene enrichment assessed by the Fisher exact test; **p<0.01.

Figure 4. Real-time PCR analysis during primary infection or homologous reinfection with A/Mexico/4108/2009. RNA was purified from the lung tissue and lymph nodes and the expression levels were assessed for the following genes: A) Influenza-NP, B) CD8A, C) CCL5, D) IGHM, E) CXCL10 and F) IGHG. Results were normalized to the house-keeping gene beta-actin. Bars represent mean with SD. n=3/group. Statistically significant increase respect to the control groups is indicated. *p≤0.05, Mann-Whitney U test.

Figure 5. Stylized representation of the biological components that participate during influenza H1N1 infection. During a primary influenza infection with little or no prior specific immunity (left), a combination of antiviral interferon responses and cellular effectors of the innate immune response are triggered in a first attempt to contain the spread of the virus. Meanwhile, antigen presenting cells get loaded with influenza antigens at the infection site, they migrate to the regional lymph nodes and maturation of cellular effectors of the adaptive response takes place. Approximately 5 days post-infection, the innate immunity is replaced by the adaptive responses, which are able to reduce the vial loads in an exponential manner. During a H1N1 reinfection process (right), pre-existing antibodies neutralize the virus and limit the viral spread. Also, the early adaptive immune response, which is greatly contributed by mature lymphocytes residing in the lung, is able to resolve the infectious process and without triggering any clinical signs.
Figure 3.