



Data in Brief

DNA microarray global gene expression analysis of influenza virus-infected chicken and duck cells



Suresh V. Kuchipudi*, Stephen P. Dunham, Kin-Chow Chang

School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, LE12 5RD, UK

ARTICLE INFO

Article history:

Received 9 March 2015

Accepted 10 March 2015

Available online 20 March 2015

Keywords:

DNA microarray

Gene expression analysis

Influenza virus

Chicken

Duck

Cross-species hybridization of arrays

ABSTRACT

The data described in this article pertain to the article by Kuchipudi et al. (2014) titled “Highly Pathogenic Avian Influenza Virus Infection in Chickens But Not Ducks Is Associated with Elevated Host Immune and Pro-inflammatory Responses” [1]. While infection of chickens with highly pathogenic avian influenza (HPAI) H5N1 virus subtypes often leads to 100% mortality within 1 to 2 days, infection of ducks in contrast causes mild or no clinical signs. The rapid onset of fatal disease in chickens, but with no evidence of severe clinical symptoms in ducks, suggests underlying differences in their innate immune mechanisms. We used Chicken Genechip microarrays (Affymetrix) to analyse the gene expression profiles of primary chicken and duck lung cells infected with a low pathogenic avian influenza (LPAI) H2N3 virus and two HPAI H5N1 virus subtypes to understand the molecular basis of host susceptibility and resistance in chickens and ducks. Here, we described the experimental design, quality control and analysis that were performed on the data set. The data are publicly available through the Gene Expression Omnibus (GEO) database with accession number GSE33389, and the analysis and interpretation of these data are included in Kuchipudi et al. (2014) [1].

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications

Platform organism	<i>Gallus gallus</i>
Sample organism	<i>Anas platyrhynchos</i> ; <i>Gallus gallus</i>
Sequencer or array type	GeneChip® chicken genome array (Affymetrix)
Data format	Normalized data [Robust Multichip Averaging (RMA) transformed]
Experimental factors	Influenza virus- or mock-infected primary lung cells
Experimental features	Microarray gene expression profiling of influenza virus (H2N3, H5N1 50–92, or H5N1 ty-Ty) or mock-infected chicken and duck cells at 24 h after infection.
Consent	NA
Sample source location	NA

Direct link to deposited data

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33389>

Influenza A virus (IAV) is one of the major causes of economic loss to global poultry industry. In addition, avian IAVs, in particular, Eurasian lineage HPAI H5N1 viruses, are major zoonotic pathogens [2] causing severe disease in humans with a fatality rate of around 60% [3]. Avian IAVs contributed to the emergence of viruses that caused all past human influenza pandemics [4]. Aquatic birds such as ducks serve as

the reservoir for most influenza A viruses [5]. HPAI H5N1 virus strains cause severe disease in chickens, turkeys and quails often with up to 100% mortality within days [6,7], but ducks infected with most HPAI viruses show little or no clinical signs [8–10]. To establish the molecular basis for the contrasting clinical outcomes of H5N1 HPAI virus infection in chickens and ducks, we examined differences in host gene expression between HPAI H5N1 virus-infected primary chicken and duck lung cells.

Experimental design, materials and methods

Cells and viruses

Primary lung cells derived from White Leghorn chicken (*Gallus gallus*) and Pekin ducks (*Anas platyrhynchos*), a low pathogenicity avian influenza virus (A/mallard duck/England/7277/06, referred as LPAI-H2N3), a classical HPAI virus H5N1 strain (A/turkey/England/50-92/91, referred as H5N1-50-92) and a contemporary Eurasian lineage clade 2.2.1 H5N1 virus (A/turkey/Turkey/1/05, referred as H5N1-ty-Ty) were used in this study.

Experimental design

Primary chicken and duck lung cells grown in 6-well cell culture plates (Corning) were infected with avian H2N3 or H5N1 50–92 or

* Corresponding author.

E-mail address: suresh.kuchipudi@nottingham.ac.uk (S.V. Kuchipudi).

H5N1 ty-Ty at multiplicity of infection (MOI) of 1.0 or mock infected. Three wells of each cell type were used for each of the three viruses. Mock-infected controls were performed in triplicate wells for each cell type without virus infection. Total RNA from cells was extracted using RNeasy Mini-QIAshredder Kit (Qiagen) following the manufacturer's instructions at 24 h after infection. Total RNA samples were hybridized to GeneChip® chicken genome array (Affymetrix), and a total of 16 array chips were used for the study.

Microarray expression analysis

Extracted total RNA samples were analyzed for their suitability for further analysis using Agilent RNA 6000 nano kit (Agilent) following the manufacturer's instructions. RNA targets for labelling were prepared using the GeneChip® 3' IVT Express Kit (Affymetrix), which was based on linear RNA amplification and employed T7 *in vitro* transcription technology. This method also known as the Eberwine [11] or reverse transcription-IVT (RT-IVT) method is considered the 'gold standard' for target preparation for gene expression analysis. Set of poly-A RNA controls were used as exogenous positive controls to monitor the entire target labelling process. GeneChip chicken genome array used in this study contains probe sets for *B. subtilis* genes (*dap*, *lys*, *phe*, *thr*) that are absent in the eukaryotic samples. Target RNA samples were mixed with the poly-A RNA controls, which were then amplified and labelled together. Examination of hybridization intensities of poly-A RNA controls helped to monitor the labelling process independently from the quality of the starting RNA samples.

First-strand cDNA was synthesized from RNA samples by reverse transcription reaction primed with T7 oligo (dT) primer to synthesize cDNA containing a T7 promoter sequence. First-strand cDNA samples were used for second-strand cDNA synthesis using DNA polymerase and RNase H to simultaneously degrade the RNA and synthesize second-strand cDNA. The samples were then subjected to *in vitro* transcription to synthesize multiple copies of biotin-modified amplified RNA (aRNA) from the double stranded cDNA templates.

The aRNA were then purified to remove unincorporated NTPs, salts, enzymes, and inorganic phosphate to improve the stability of the biotin-modified aRNA. The fragmentation of aRNA targets was carried out before hybridization onto GeneChip probe array, which was critical in obtaining optimal assay sensitivity. Hybridization of labelled target on to GeneChip probe arrays was carried out using GeneChip® Hybridization, Wash and Stain Kit (Affymetrix) following manufacturer's instructions. After placing the probe array in the hybridization oven, temperature was set to 45 °C and hybridized for 16 h with rotation at 60 rpm, the probe array was removed from the oven and the hybridization cocktail was extracted with a micropipette. Probe arrays were then washed and stained before scanning using a GeneChip® Scanner 3000 with AGCC scan control software (Affymetrix). After scanning, the software aligned a grid on the image to identify the probe cells and computed the probe cell intensity data. The probe intensity data from each array were generated (.cel file) and analyzed using GenespringGx10 software (Agilent).

Microarray data analysis

Microarray expression analysis was carried out using the GeneSpring GX10 expression analysis software (Agilent Technologies). The *Advanced Workflow* option was used for data analysis in the GeneSpring GX 10, which provided many options for summarization algorithms, normalization routines, etc., depending on the technology used. Probe summarization was carried out by Robust Multichip Averaging (RMA) summarization algorithm [12,13]. The RMA algorithm conducts background correction, followed by quantile normalization and probe summarization. Subsequent to probe set summarization, baseline transformation of the data was performed with the option of baseline to median of all samples. The software calculated the log-

summarized values from all the samples for each probe and calculated the median and subtracted from each of the samples. Experimental grouping was done by defining four groups which were uninfected control, H2N3 infected, 50–92 infected and ty-Ty infected with 2 replicate arrays in each group. An interpretation was created to specify grouping of samples based on treatment as the experimental condition using the *create interpretation* function.

Quality control on arrays

Quality control check on all samples was carried out using the principal component analysis (PCA), and the scores were visually represented in a 3D scatter plot. PCA analysis showed that the replicate arrays in each treatment group were clustered together indicating good quality of the samples and hybridization (Fig. 1A and B). Correlation analysis across arrays was carried out by the Pearson correlation coefficients which showed high correlation between the replicates in each group. Correlation coefficients of each pair of arrays were between 0.98 and 1.0, and the results were displayed in visual form as a heatmap (Fig. 1C and D).

The internal controls represented RNA sample quality by showing 3'/5' ratios for a set of specific probe sets which included the actin and GAPDH probe sets. For good quality samples, the ratios for actin and GAPDH should be no more than 3. The internal control analysis of arrays showed actin and GAPDH ratios less than 3 for all the samples, indicating good sample quality. The hybridization controls represented the hybridization quality, which were composed of a mixture of biotin-labelled cRNA transcripts of *bioB*, *bioC*, *bioD* and *cre* prepared in staggered concentrations (1.5, 5, 25 and 100 pm, respectively). This mixture was spiked into the hybridization cocktail. *BioB* was at the level of assay sensitivity and should be present at least 50% of the time, whereas *bioC*, *bioD* and *cre* must be present all of the time and must appear in increasing concentrations. The hybridization controls showed the signal value profiles of these transcripts (only 3' probe sets are taken), where the x-axis represented the biotin-labelled cRNA transcripts and the y-axis represented the log of the normalized signal values. We checked that the hybridization controls of our arrays showed the signal value profiles as expected indicating good hybridization quality.

Statistical analysis

Statistical analysis was carried out by analysis of variance (ANOVA) with a *p* value cutoff of 0.05 by asymptotic *p* value computation algorithm with no multiple testing correction. The entities satisfying the significance analysis were passed on for the fold change analysis. Fold change analysis was used to identify genes with expression ratios of treatment and control samples that are outside of cutoff of 1.3. Fold change was calculated between mock and each virus-infected group separately.

Cross-species hybridization—using chicken GeneChip arrays for duck transcriptome analysis

As there was no high-density microarray platform available for duck, chicken gene chip was utilized for duck transcriptome analysis. A well-established gDNA-based probe selection method was used for increasing the sensitivity of chicken GeneChip to study the transcriptome of duck [14]. Briefly, Pekin duck (*A. platyrhynchos*) genomic DNA from cells was biotin-labelled and hybridized to the Chicken (*G. gallus*) GeneChip® array and a probe intensity data file (.cel) was generated as described above. Probe sets on the chicken chip were selected for subsequent duck transcriptome analyses if the probe set was represented by perfect match (PM) probes with duck gDNA hybridization intensities above an experimentally set threshold. Selection was performed using a .cel file parser script written in the Perl programming language

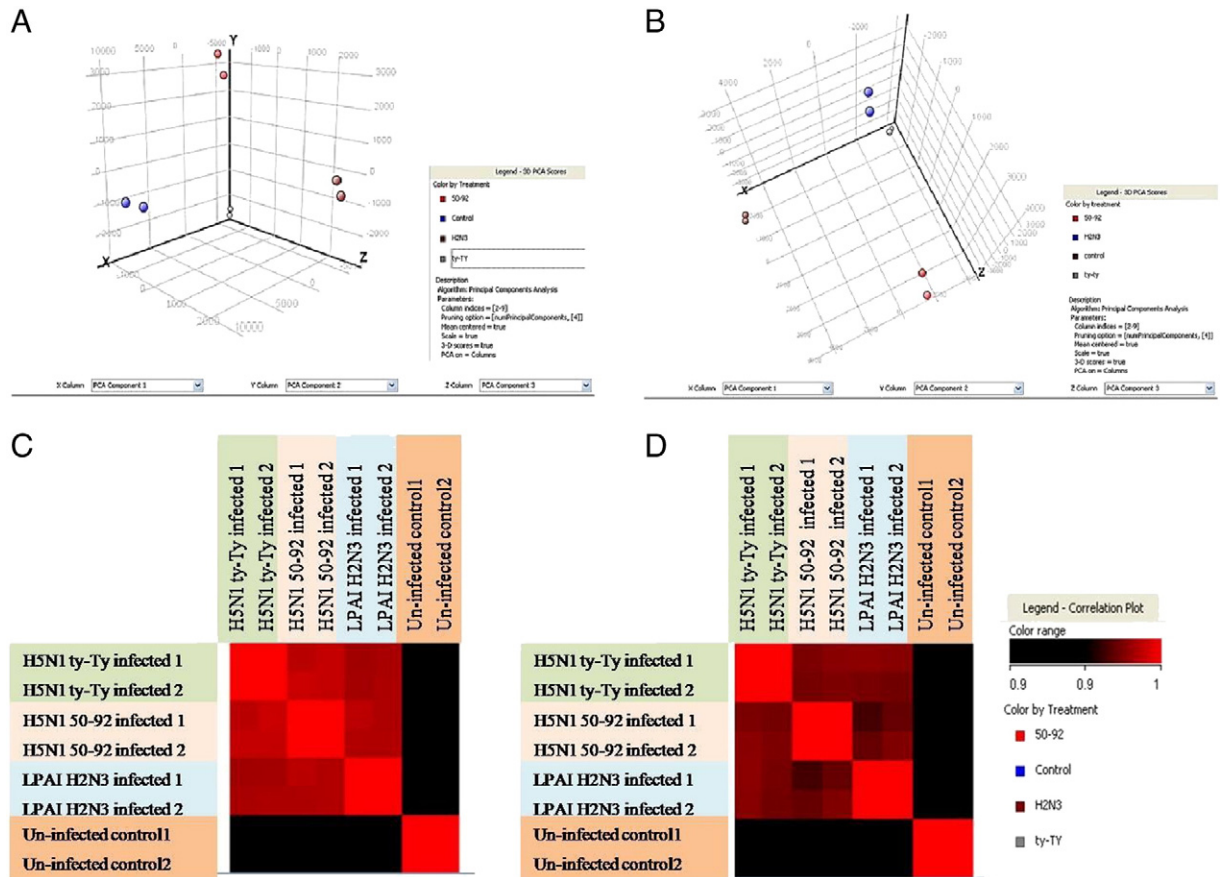


Fig. 1. Quality control of arrays. Principal component analysis (PCA) plots showing arrays hybridized with chicken (A) and duck (B) virus- and mock-infected samples. Each point representing one array with replicate samples in each group represented by the same colour clustered together. Correlation analysis of chicken (C) and duck (D) samples showing high degree of correlation between each pair of arrays in infected and control groups (Pearson correlation coefficient values ranging from 0.98 to 1.0).

(X-species Version 2.1, <http://affymetrix.arabidopsis.info/xspecies/>). After installing Active Perl software for windows (Active Perl version 5.10.1.1007 for Windows), *CDF_masking.zip* was downloaded and unzipped to a chosen location on the computer (http://affymetrix.arabidopsis.info/xspecies/CDF_masking.zip). Original CDF file for chicken chip downloaded from Affymetrix website (<http://www.affymetrix.com>) and duck gDNA hybridization *cel* file were copied to CDF-masking folder. In the CDF-masking folder, *easy_script.pl* was run to generate a series of probe mask (CDF) files for duck with a range of threshold values. After executing *easy_script.pl*, the desired gDNA hybridization intensity threshold value was needed to be mentioned to generate a probe masking file with a particular intensity threshold. Using this method, 20 probe mask files were generated with gDNA hybridization intensity thresholds ranging from 20 to 2000.

PM probes of chicken genome array hybridized extensively to the *A. platyrhynchos* genomic DNA (Fig. 2A). When the gDNA hybridization intensity threshold was increased from 20 to 2000, probe pair retention in the probe mask files decreased rapidly. However, the retention of whole probe sets, representing transcripts, was less sensitive to the increase in gDNA hybridization intensities during probe mask file generation. This was because only a minimum of one probe pair was required to retain a probe set. For example, probe mask file generated using a gDNA hybridization intensity threshold of 20 retained 100% *G. gallus* probe pairs and probe sets (i.e. 423199 and 38473 respectively). The probe mask file generated with a gDNA intensity threshold of 100, masked over 50% of probe pairs, while only 2.5% of *G. gallus* probe sets were masked (retaining 97.5 % probe sets).

Differentially regulated genes in duck cells were analyzed by comparing the treatment group against control using technologies created with each of the 11 gDNA intensities from 40 to 450. Out of

the 11 gDNA intensities analyzed, a threshold of 200 gave the highest number of genes with a significant differential regulation ($p < 0.05$) (Fig. 2B). Similarly, a threshold of 200 gave the highest number of genes regulated at a fold change of ± 2 ($p < 0.05$) (Fig. 2 C and D). Based on these findings, technology created with a gDNA threshold of 200 was used for further transcriptome analysis of all the duck samples. Technology created for duck represented 32896 transcripts out of the total 38535 transcripts represented in the original chicken GeneChip technology (Fig. 3).

Gene ontology analysis

All the genes that were significantly regulated ($p < 0.05$) by a fold change difference of ± 1.3 were grouped into gene ontology (GO) terms using the GO analysis function in GeneSpring software. Genes were grouped into cellular component, biological process and molecular function GO terms (Fig. 4). The percentage of differentially regulated genes that fitted into each of these GO terms was determined from the gene expression profiles of chicken and duck cells.

Discussion

Microarray global gene expression analysis is a useful tool to investigate effects of virus infection on host gene expression [15]. High-density microarray platforms can be used for cross-species hybridization to study the global gene expression of heterologous species [15,16]. We showed that the Chicken GeneChip arrays could be used for the analysis of duck transcriptome and can be used for gene expression analysis in other avian species [17]. While these data sets are highly valuable to explore host response to influenza virus infection

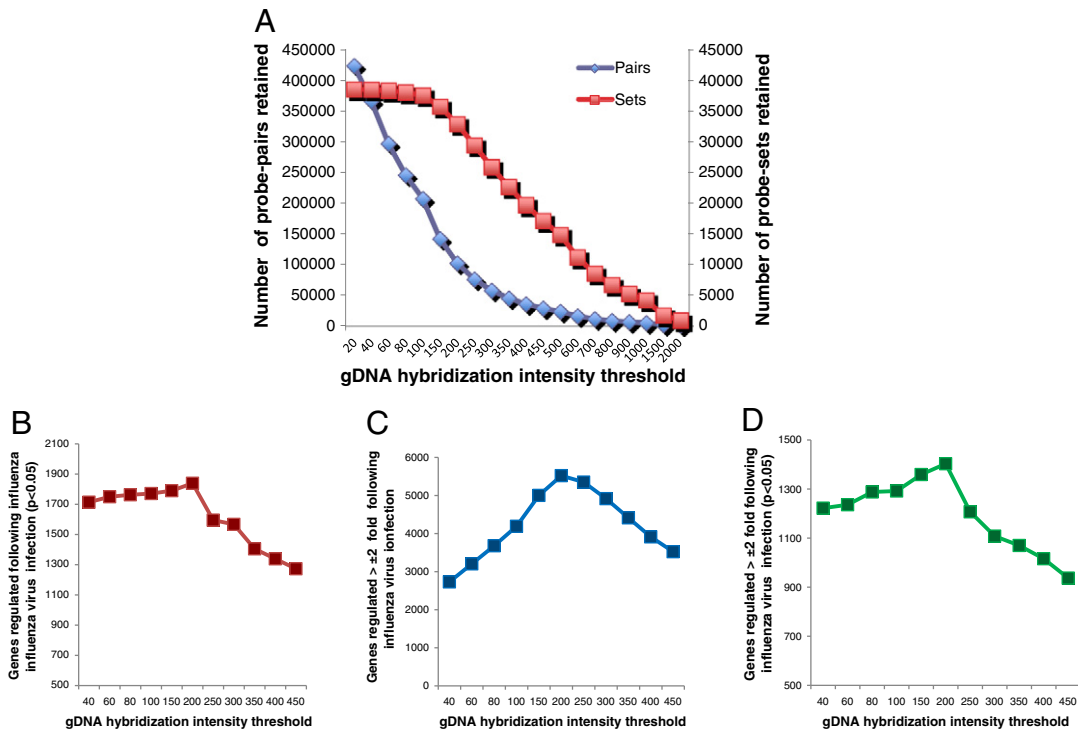


Fig. 2. Genomic DNA (gDNA) based probe selection to improve the sensitivity of chicken Genechip for duck transcriptome analysis. (A) *Anas platyrhynchos* genomic DNA (gDNA) hybridization intensity thresholds used to generate the probe mask files is shown. Data were obtained by hybridizing duck gDNA on chicken Genechip. Number of *Gallus gallus* probe pairs and probe sets from the chicken GeneChip® array retained across a range of gDNA intensity threshold is shown. Probe pairs retained (data in blue) is scaled to the left hand y-axis, while number of probe sets retained (data in red) are scaled to the to the right-hand y-axis). Intensity threshold of 200 gave highest number of genes differentially regulated following 24 h of infection with influenza viruses (H2N3, 50–92 and ty-Ty) compared to mock-infected controls. (B) All the genes significantly differentially regulated ($p < 0.05$). (C) Genes regulated $> \pm 2$ -fold following infection. (D) Genes significantly regulated ± 2 -fold ($p < 0.05$).

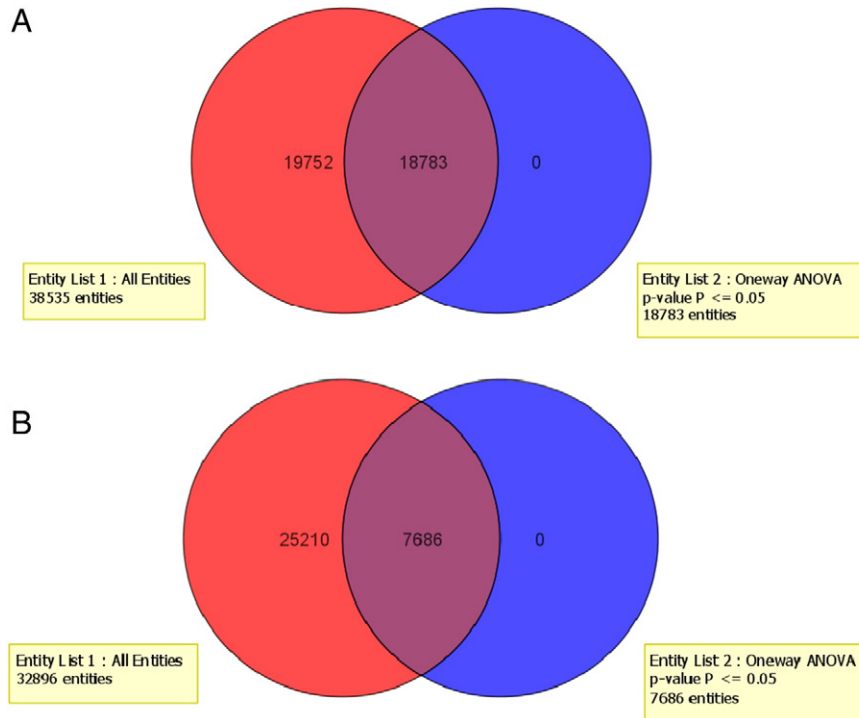


Fig. 3. Comparison of gene expression profiles of virus-infected and mock-infected samples showing differentially regulated genes after influenza virus infection in chicken (A) and duck (B) cells with a p value cut off of 0.05. In chicken cells, 48.74% transcripts were differentially regulated, whereas in duck cells, 23.36% of the transcripts were differentially regulated compared to control. Red circles represent all the entities (transcripts) on the array, and blue circles represent significantly ($p < 0.05$) differentially regulated genes derived by analysis of variance (ANOVA).

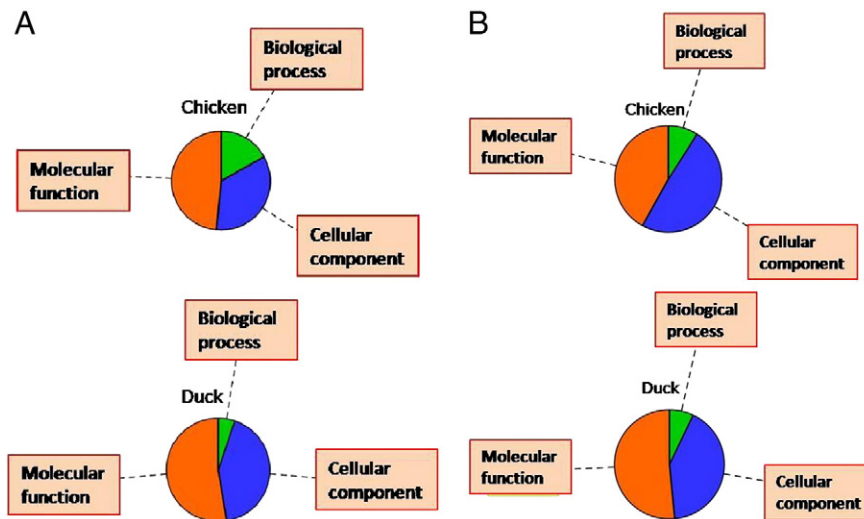


Fig. 4. Gene ontology analysis of gene expression profiles of chicken and duck cells at 24 h post-infection with H5N1 50–92 virus (A) or H5N1 ty-Ty virus (B). Significantly differentially expressed genes ($p < 0.05$) with a fold change difference of ± 1.3 between virus- and mock-infected samples were categorized into three major gene ontology terms. Each coloured fraction of the Venn diagram represents the percentage of all the differentially regulated genes that fits into the particular ontology term.

in chicken and ducks, users should be aware that the data were a snapshot of gene expression changes to IAV infection *in vitro* and any significant observations must be validated. The analysis and interpretation of these data are included in Kuchipudi et al. (2014) [1].

Acknowledgements

This work was part funded by the BBSRC grant BB/E010849/1, the University of Nottingham (pump prime grant to SVK and SPD). The authors are grateful to Prof. Ian Brown and colleagues at Animal and Plant Health Agency (APHA), Weybridge, for the kind supply of influenza viruses, technical support and access to high containment laboratory to perform the HPAI infection studies.

References

- [1] S.V. Kuchipudi, et al., Highly pathogenic avian influenza virus infection in chickens but not ducks is associated with elevated host immune and pro-inflammatory responses. *Vet. Res.* 45 (2014) 118.
- [2] A.J. Gibbs, J.S. Armstrong, J.C. Downie, From where did the 2009 'swine-origin' influenza A virus (H1N1) emerge? *Virol. J.* 6 (2009) 207.
- [3] A. Thitithayanont, et al., Antiviral immune responses in H5N1-infected human lung tissue and possible mechanisms underlying the hyperproduction of interferon-inducible protein IP-10. *Biochem. Biophys. Res. Commun.* 398 (4) (2010) 752–758.
- [4] J.M. Katz, et al., The public health impact of avian influenza viruses. *Poult. Sci.* 88 (4) (2009) 872–879.
- [5] G.B. Sharp, et al., Coinfection of wild ducks by influenza A viruses: distribution patterns and biological significance. *J. Virol.* 71 (8) (1997) 6128–6135.
- [6] O.M. Jeong, et al., Experimental infection of chickens, ducks and quails with the highly pathogenic H5N1 avian influenza virus. *J. Vet. Sci.* 10 (1) (2009) 53–60.
- [7] T. Saito, et al., Pathogenicity of highly pathogenic avian influenza viruses of H5N1 subtype isolated in Thailand for different poultry species. *Vet. Microbiol.* 133 (1–2) (2009) 65–74.
- [8] N. Isoda, et al., Pathogenicity of a highly pathogenic avian influenza virus, A/chicken/Yamaguchi/7/04 (H5N1) in different species of birds and mammals. *Arch. Virol.* 151 (7) (2006) 1267–1279.
- [9] H. Kida, R. Yanagawa, Y. Matsuoka, Duck influenza lacking evidence of disease signs and immune response. *Infect. Immun.* 30 (2) (1980) 547–553.
- [10] L.E.L. Perkins, D.E. Swayne, Pathogenicity of a Hong Kong-origin H5N1 highly pathogenic avian influenza virus for emus, geese, ducks, and pigeons. *Avian Dis.* 46 (1) (2002) 53–63.
- [11] J. Phillips, J.H. Eberwine, Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. *Methods* 10 (3) (1996) 283–288.
- [12] R.A. Irizarry, et al., Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31 (4) (2003) e15.
- [13] R.A. Irizarry, et al., Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4 (2) (2003) 249–264.
- [14] J.P. Hammond, et al., Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods* 1 (1) (2005) 10.
- [15] D.E. Moody, Z. Zou, L. McIntyre, Cross-species hybridisation of pig RNA to human nylon microarrays. *BMC Genomics* 3 (1) (2002) 27.
- [16] P.W. Anderson, B.C. Tennant, Z. Lee, Cross-species hybridization of woodchuck hepatitis virus-induced hepatocellular carcinoma using human oligonucleotide microarrays. *World J. Gastroenterol.* 12 (29) (2006) 4646–4651.
- [17] T.M. Crowley, et al., Application of chicken microarrays for gene expression analysis in other avian species. *BMC Genomics* 10 (Suppl. 2) (2009) S3.