and produces a set of regularized logical structures representing the meaning of each sentence. The ResNet mammalian database stores information harvested from the entire PubMed, including over 715,000 relations for 106,139 proteins, 1220 small molecules, 2175 cellular processes and 3930 diseases. The focus of this database is solely on human, mouse and rat.

We used the list of differentially expressed genes to build a gene regulatory network without including any additional genes not found in microarray experiments resulting in a raw connected graph of 125 nodes and 255 interactions of known effect (positive or negative). To build the network we included only literature evidences of gene expresion regulation (directed and signed interactions) and is therefore smaller and sparser than it would have been if all possible known interactions had been included (i.e. undirected protein-protein interactions, indirect interactions). The expression patterns of the DEGs were checked in the Prion Database (http://prion.systemsbiology.net) to compare with topology and associations' logic leading to removal of inconsistent 15 nodes and 81 edges. Additionally, discovery of few errors in text mining process lead us to further validation of the network. To avoid false associations we took all sentences used by Pathway Studio (Ariadne Genomics) to determine gene associations and searched for co-occurrence of specific words: modifiers of sentence meaning, indicating increased risk of false interpretation. In the next phase we checked manually highly uncertain sentences and found two clearly wrong associations: CD86 –+> TGFB1 and CEBPA –+> CASP8. In summary, we obtained a final graph of 106 nodes and 169 edges we used for fragmentation analysis in this paper. References for both raw and global network interactions are included in the Additional file 1.

c) Determining the core regulatory network.

Given that only genes with incoming interactions are relevant to the stability analysis, we had to identify genes involved in regulatory feed-back loops, or circuits, and genes regulated by them. For the first task we looked for strongly connected components (SCCs) in the raw network using Binom plugin [48] in Cytoscape [49].

An SCC is a network of nodes, where each node can be accessed directly or indirectly from every other node within the network or, in other words, if there exists a path from each node in the network to every other node. Due to the specific connectivity in a SCC, the information can flow from one node to any other in the structure following at least one path. Such a path has to respect the sense of the interactions (otherwise the component is not strongly but weakly connected). Therefore, the state of any node in the SCC can directly or indirectly affect the state of any other node. This mutual influence between any pair of nodes within the SCC indicates that the SCC may be a relevant stability-related structure. We obtained a single SCC with 16 nodes. After that, we expanded these cores iteratively by adding first neighbors regulated by the SCC until no further neighbors could be added. This yielded a network consisting of the SCC and genes that are directly or indirectly regulated by genes in the SCC (we call this the "core network"). The core network including 74 nodes and 125 interactions (all nodes with incoming interactions) was used for stability and centrality analysis in this paper. References for the interactions of the core network are included in the Additional file 1.

Stability and perturbation analysis

For the stability analysis we used the SQUAD software package [50], creating a discrete dynamical system that allowed us to identify all the stable states of the system with an asynchronous updating scheme [51] using a binary decision diagram based algorithm [52]. Subsequently, a continuous dynamical system was created to identify the stable states in this continuous model which are located near to the stable states of the discrete system, according to the method described by Mendoza *et al.*, 2006 [53], where the stable states of a Boolean model are taken as initial conditions in the continuous model. Gene perturbations were simulated in the continuous model changing the expression values of specific genes. We also calculated the stability of the SCC in isolation, as well as the stability of the core network. More details about stable states computation and perturbation simulations are included in the Additional file 1.

Network properties

Fragmentation, betweenness centrality and inter-modular participation measurements were employed to compare the properties of SCC genes with other genes in the network, and to determine key genes that might be potential candidates for experimental validation.

In order to test the importance of the SCC in the network's connectivity, we examined the fragmentation effect of removing the 16 nodes belonging to the SCC in comparison with the fragmentation effect of 1000 different randomized removals of 16 nodes in the global network of 106 nodes. The giant component is the biggest connected subgraph found in the network for the given fragmentation and thus a good measure for evaluating such fragmentation [54,55].

Betweenness centrality was computed for all genes in the network. The higher the value, the more central the gene is in the network of reference, i.e. other genes are more likely to be connected along the pathway involving these genes [54] (see Additional file 1).

Modules in the global network were defined by functional and pathological process annotation of genes. The participation coefficient P is a measure quantifying inter-modular connections of genes. For any gene in question P is greater than 0, if odds of inter-modular degree to total degree of the gene is less than 1, which means it has to have at least one connection within its own and neighboring modules. Together with measure of within-module connectivity, participation allows to define role a node in the network ranging from most influential global hub till peripheral node (global hub, connector hub, provincial hub, kinless node, satellite connector, peripheral node and ultra-peripheral node). Such genes connect various functional pathways and might therefore be considered key regulators of cellular processes [56].

Functional analysis

Hwang *et al.* described four pathological features, which were derived from GO attributes: (1) PrP^{Sc} replication and accumulation, (2) microglia/astrocyte activation (which we are calling immune response), (3) synaptic degeneration and (4) neuronal cell death. We mapped these pathological features on the nodes in our core network and examined how the genes in our network may relate to disease progression.

Additional file

Additional file 1: Supplementary material.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

IC participated in network reconstruction and analysis, the design of the study and drafted the manuscript, KR participated in the network reconstruction and analysis, the design of the study and drafted the manuscript, WJ participated in the network reconstruction and analysis, participated in the design and coordination of the study and drafted the manuscript, HK conceived of the study, AdS conceived of the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Manuel Buttini for helpful discussions and for editing the manuscript. We also thank Dr. Maria Biryukov for helping us with text mining to curate the network.

Author details

[1]Luxembourg Center for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, 7, avenue des Hauts fourneaux, Luxembourg L-4362, Luxembourg. [2]The Systems Biology Institute, Tokyo 108-0071, Japan.

Received: 10 May 2012 Accepted: 17 September 2012
Published: 15 October 2012

References

1. Colby DW, Prusiner SB: **Prions.** *Cold Spring Harbor perspectives in biology* 2011, **3**:a006833.
2. Soto C: **Unfolding the role of protein misfolding in neurodegenerative diseases.** *Nat Rev Neurosci* 2003, **4**:49–60.
3. Soto C, Satani N: **The intricate mechanisms of neurodegeneration in prion diseases.** *Trends in molecular medicine* 2010, **17**:14–24.
4. Hwang D, Lee IY, Yoo H, Gehlenborg N, Cho JH, *et al*: **A systems approach to prion disease.** *Mol Syst Biol* 2009, **5**:252.
5. Lucin KM, Wyss-Coray T: **Immune activation in brain aging and neurodegeneration: too much or too little?** *Neuron* 2009, **64**:110–122.
6. Amor S, Puentes F, Baker D, van der Valk P: **Inflammation in neurodegenerative diseases.** *Immunology* 2010, **129**:154–169.
7. Hughes MM, Field RH, Perry VH, Murray CL, Cunningham C: **Microglia in the degenerating brain are capable of phagocytosis of beads and of apoptotic cells, but do not efficiently remove PrPSc, even upon LPS stimulation.** *Glia* 2010, **58**:2017–2030.
8. Rezaie P, Lantos PL: **Microglia and the pathogenesis of spongiform encephalopathies.** *Brain Res Brain Res Rev* 2001, **35**:55–72.
9. Wojtera M, Sikorska B, Sobow T, Liberski PP: **Microglial cells in neurodegenerative disorders.** *Folia Neuropathol* 2005, **43**:311–321.
10. Williams A, Lucassen PJ, Ritchie D, Bruce M: **PrP deposition, microglial activation, and neuronal apoptosis in murine scrapie.** *Exp Neurol* 1997, **144**:433–438.
11. Moes M, Le Bechec A, Crespo I, Laurini C, Halavatyi A, *et al*: **A novel network integrating a miRNA-203/SNAI1 feedback loop which regulates epithelial to mesenchymal transition.** *PLoS One* 2012, **7**:e35440.
12. Shiraishi T, Matsuyama S, Kitano H: **Large-scale analysis of network bistability for human cancers.** *PLoS Comput Biol* 2010, **6**:e1000851.
13. Sanz J, Cozzo E, Borge-Holthoefer J, Moreno Y: **Topological effects of data incompleteness of gene regulatory networks.** *BMC Syst Biol* 2012, **6**:110.
14. de Silva E, Thorne T, Ingram P, Agrafioti I, Swire J, *et al*: **The effects of incomplete protein interaction data on structural and evolutionary inferences.** *BMC Biol* 2006, **4**:39.
15. Soto C, Satani N: **The intricate mechanisms of neurodegeneration in prion diseases.** *Trends Mol Med* 2010, **17**:14–24.
16. Pirot P, Eizirik DL, Cardozo AK: **Interferon-gamma potentiates endoplasmic reticulum stress-induced death by reducing pancreatic beta cell defence mechanisms.** *Diabetologia* 2006, **49**:1229–1236.
17. Malhotra JD, Kaufman RJ: **The endoplasmic reticulum and the unfolded protein response.** *Semin Cell Dev Biol* 2007, **18**:716–731.
18. Kim I, Xu W, Reed JC: **Cell death and endoplasmic reticulum stress: disease relevance and therapeutic opportunities.** *Nat Rev Drug Discov* 2008, **7**:1013–1030.
19. Opazo C, Barria MI, Ruiz FH, Inestrosa NC: **Copper reduction by copper binding proteins and its relation to neurodegenerative diseases.** *Biometals: an international journal on the role of metal ions in biology, biochemistry, and medicine* 2003, **16**:91–98.
20. Singh N, Singh A, Das D, Mohan ML: **Redox control of prion and disease pathogenesis.** *Antioxid Redox Signal* 2010, **12**:1271–1294.
21. Aguzzi A, Calella AM: **Prions: protein aggregation and infectious diseases.** *Physiol Rev* 2009, **89**:1105–1152.
22. Tengku-Muhammad TS, Hughes TR, Ranki H, Cryer A, Ramji DP: **Differential regulation of macrophage CCAAT-enhancer binding protein isoforms by lipopolysaccharide and cytokines.** *Cytokine* 2000, **12**:1430–1436.
23. Xu Y, Saegusa C, Schehr A, Grant S, Whitsett JA, *et al*: **C/EBP{alpha} is required for pulmonary cytoprotection during hyperoxia.** *Am J Physiol Lung Cell Mol Physiol* 2009, **297**:L286–L298.
24. Li R, Strohmeyer R, Liang Z, Lue LF, Rogers J: **CCAAT/enhancer binding protein delta (C/EBPdelta) expression and elevation in Alzheimer's disease.** *Neurobiol Aging* 2004, **25**:991–999.
25. Sorensen G, Medina S, Parchaliuk D, Phillipson C, Robertson C, *et al*: **Comprehensive transcriptional profiling of prion infection in mouse models reveals networks of responsive genes.** *BMC Genomics* 2008, **9**:114.
26. Hamilton JA, Whitty G, White AR, Jobling MF, Thompson A, *et al*: **Alzheimer's disease amyloid beta and prion protein amyloidogenic peptides promote macrophage survival, DNA synthesis and enhanced proliferative response to CSF-1 (M-CSF).** *Brain Res* 2002, **940**:49–54.
27. Bate C, Boshuizen R, Williams A: **Microglial cells kill prion-damaged neurons *in vitro* by a CD14-dependent process.** *J Neuroimmunol* 2005, **170**:62–70.
28. Konno T, Morii T, Hirata A, Sato S, Oiki S, *et al*: **Covalent blocking of fibril formation and aggregation of intracellular amyloidgenic proteins by**

transglutaminase-catalyzed intramolecular cross-linking. *Biochemistry* 2005, 44:2072–2079.
29. Iismaa SE, Mearns BM, Lorand L, Graham RM: Transglutaminases and disease: lessons from genetically engineered mouse models and inherited disorders. *Physiol Rev* 2009, 89:991–1023.
30. Minami M, Satoh M: Role of chemokines in ischemic neuronal stress. *Neuromolecular Med* 2005, 7:149–155.
31. AstraZeneca: *Thiazolopyrimidine derivatives and use thereof as antagonists of the cx3cr1 receptor AstraZeneca AB*. Sverker Hanson: Gunnar Nordvall; 2007.
32. Cunningham C, Boche D, Perry VH: Transforming growth factor beta1, the dominant cytokine in murine prion disease: influence on inflammatory cytokine synthesis and alteration of vascular extracellular matrix. *Neuropathol Appl Neurobiol* 2002, 28:107–119.
33. Goh AX, Li C, Sy MS, Wong BS: Altered prion protein glycosylation in the aging mouse brain. *J Neurochem* 2007, 100:841–854.
34. Spudich A, Frigg R, Kilic E, Kilic U, Oesch B, *et al*: Aggravation of ischemic brain injury by prion protein deficiency: role of ERK-1/-2 and STAT-1. *Neurobiol Dis* 2005, 20:442–449.
35. Frigg R, Wenzel A, Samardzija M, Oesch B, Wariwoda H, *et al*: The prion protein is neuroprotective against retinal degeneration *in vivo*. *Exp Eye Res* 2006, 83:1350–1358.
36. Hashimoto Y, Suzuki H, Aiso S, Niikura T, Nishimoto I, *et al*: Involvement of tyrosine kinases and STAT3 in Humanin-mediated neuroprotection. *Life Sci* 2005, 77:3092–3104.
37. Sponne I, Fifre A, Koziel V, Kriem B, Oster T, *et al*: Humanin rescues cortical neurons from prion-peptide-induced apoptosis. *Mol Cell Neurosci* 2004, 25:95–102.
38. Khoo JJ, Forster S, Mansell A: Toll-like receptors as interferon-regulated genes and their role in disease. *Journal of interferon & cytokine research: the official journal of the International Society for Interferon and Cytokine Research* 2011, 31:13–25.
39. Zeytun A, van Velkinburgh JC, Pardington PE, Cary RR, Gupta G: Pathogen-specific innate immune response. *Advances in experimental medicine and biology* 2007, 598:342–357.
40. Trudler D, Farfara D, Frenkel D: Toll-like receptors expression and signaling in glia cells in neuro-amyloidogenic diseases: towards future therapeutic application. *Mediators Inflamm* 2010, 497987:12.
41. Hennessy EJ, Parker AE, O'Neill LA: Targeting toll-like receptors: emerging therapeutics? *Nature reviews Drug discovery* 2010, 9:293–307.
42. Babior BM, Lambeth JD, Nauseef W: The neutrophil NADPH oxidase. *Arch Biochem Biophys* 2002, 397:342–344.
43. Chong ZZ, Maiese K: The Src homology 2 domain tyrosine phosphatases SHP-1 and SHP-2: diversified control of cell growth, inflammation, and injury. *Histol Histopathol* 2007, 22:1251–1267.
44. Thams S, Oliveira A, Cullheim S: MHC class I expression and synaptic plasticity after nerve lesion. *Brain Res Rev* 2008, 57:265–269.
45. Kauffman S: A proposal for using the ensemble approach to understand genetic regulatory networks. *J Theor Biol* 2004, 230:581–590.
46. Novichkova S, Egorov S, Daraselia N: MedScan, a natural language processing engine for MEDLINE abstracts. *Bioinformatics* 2003, 19:1699–1706.
47. Daraselia N, Yuryev A, Egorov S, Novichkova S, Nikitin A, *et al*: Extracting human protein interactions from MEDLINE using a full-sentence parser. *Bioinformatics* 2004, 20:604–611.
48. Zinovyev A, Viara E, Calzone L, Barillot E: BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics* 2008, 24:876–877.
49. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T: Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011, 27:431–432.
50. Di Cara A, Garg A, De Micheli G, Xenarios I, Mendoza L: Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinforma* 2007, 8:462.
51. Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G: Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics* 2008, 24:1917–1925.
52. Garg A, Xenarios I, Mendoza L, DeMicheli G: An efficient method for dynamic analysis of gene regulatory networks and in silico gene perturbation experiments research in computational molecular biology. In *Lecture Notes in Computer Science, Vol. 4453*. Edited by Speed T, Huang H. Heidelberg: Springer Berlin; 2007:62–76.
53. Mendoza L, Xenarios I: A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theoretical biology & medical modelling* 2006, 3:13.
54. Newman M, Barabási AL, Watts DJ: *The Structure and Dynamics of Networks*. Princeton, New Jersey: Princeton University Press; 2006.
55. Dobrin R, Beg QK, Barabasi AL, Oltvai ZN: Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network. *BMC Bioinforma* 2004, 5:10.
56. Guimera R, Sales-Pardo M, Amaral LA: Classes of complex networks defined by role-to-role connectivity profiles. *Nat Phys* 2007, 3:63–69.

doi:10.1186/1752-0509-6-132
Cite this article as: Crespo *et al.*: Gene regulatory network analysis supports inflammation as a key neurodegeneration process in prion disease. *BMC Systems Biology* 2012 6:132.

BMC Systems Biology

RESEARCH ARTICLE Open Access

Integrated network analysis reveals a novel role for the cell cycle in 2009 pandemic influenza virus-induced inflammation in macaque lungs

Jason E Shoemaker[1], Satoshi Fukuyama[1†], Amie J Eisfeld[2†], Yukiko Muramoto[3], Shinji Watanabe[1,2], Tokiko Watanabe[1,2], Yukiko Matsuoka[1,2,4], Hiroaki Kitano[1,4,5,6,7*] and Yoshihiro Kawaoka[1,2,3,8,9*]

Abstract

Background: Annually, influenza A viruses circulate the world causing wide-spread sickness, economic loss, and death. One way to better defend against influenza virus-induced disease may be to develop novel host-based therapies, targeted at mitigating viral pathogenesis through the management of virus-dysregulated host functions. However, mechanisms that govern aberrant host responses to influenza virus infection remain incompletely understood. We previously showed that the pandemic H1N1 virus influenza A/California/04/2009 (H1N1; CA04) has enhanced pathogenicity in the lungs of cynomolgus macaques relative to a seasonal influenza virus isolate (A/Kawasaki/UTK-4/2009 (H1N1; KUTK4)).

Results: Here, we used microarrays to identify host gene sequences that were highly differentially expressed (DE) in CA04-infected macaque lungs, and we employed a novel strategy – combining functional and pathway enrichment analyses, transcription factor binding site enrichment analysis and protein-protein interaction data – to create a CA04 differentially regulated host response network. This network describes enhanced viral RNA sensing, immune cell signaling and cell cycle arrest in CA04-infected lungs, and highlights a novel, putative role for the MYC-associated zinc finger (MAZ) transcription factor in regulating these processes.

Conclusions: Our findings suggest that the enhanced pathology is the result of a prolonged immune response, despite successful virus clearance. Most interesting, we identify a mechanism which normally suppresses immune cell signaling and inflammation is ineffective in the pH1N1 virus infection; a dyregulatory event also associated with arthritis. This dysregulation offers several opportunities for developing strain-independent, immunomodulatory therapies to protect against future pandemics.

Keywords: Influenza, Host response, Microarray, pH1N1, Systems biology

Background

In April of 2009, a new pandemic H1N1 (pH1N1) influenza virus strain emerged in Mexico [1], and the subsequent global pandemic claimed more than 18,000 lives [2]. Human infections with pH1N1 tended to be mild with typical clinical symptoms (e.g., fever, sore throat, vomiting, and occasional diarrhea) [1]. However, in instances where death occurred, autopsies showed necrotizing bronchitis, a symptom that has also been observed in previous pandemics [3,4]. Loss of life during the 2009 pandemic was minor compared to other pandemics (e.g., the 1918 Spanish influenza pandemic, which caused an estimated 50 million deaths worldwide [5]); however, 2009 pH1N1 infections have resulted in more severe illness in young, previously healthy individuals, which is atypical for seasonal influenza virus isolates [6-8].

Currently licensed antiviral compounds against influenza virus (e.g., oseltamavir) interfere with specific functions of viral proteins and are one defense against seasonal and newly emerging pandemic influenza viruses. However,

* Correspondence: kitano@sbi.jp; kawaokay@vetmed.wisc.edu
†Equal contributors
[1]ERATO Infection-Induced Host Responses Project, Saitama 332-0012, Japan
[2]School of Veterinary Medicine, Department of Pathobiological Sciences, Influenza Research Institute, University of Wisconsin-Madison, Madison, WI, USA
Full list of author information is available at the end of the article

the sudden increase in oseltamivir-resistant seasonal H1N1 virus strains in 2007–2008 [9] and the existence of resistant pH1N1 virus isolates [10] strongly underscore the urgent need for the development of novel strategies to alleviate influenza virus-induced disease. As an alternative to directly inhibiting viral proteins, aspects of the host response that strongly correlate with host pathology could be targeted to reduce the overall severity of infection. Such therapies may be more robust against the emergence of drug-resistant strains because they do not target influenza virus-specific gene products, which are subject to drug-mediated selective pressure. Therefore, a better understanding of the mechanisms contributing to influenza virus pathogenesis is needed to provide the basis for the development of novel host-modulatory therapies. We have reported that a pH1N1 isolate (influenza A/California/04/2009 [H1N1], referred to as CA04) is more pathogenic than a seasonal H1N1 isolate in several mammalian models, including cynomolgus macaques [11]. Specifically, CA04-infected macaques exhibited higher body temperatures, more efficient virus replication in both the upper respiratory tract and lungs, increased production of pro-inflammatory cytokines (e.g., CCL2 and IL6), and more severe lung lesions at both early (3 days) and late (7 days) times post-infection. In addition, we observed an increased presence of inflammatory infiltrates in alveolar spaces and more abundant viral antigen staining of both type I and type II pneumocytes. These pathological characteristics are similar to those that distinguish highly pathogenic avian H5N1 and 1918 influenza viruses from lower pathogenicity isolates [12-14], suggesting a common host response theme that may contribute to increased pathogenicity phenotypes for diverse influenza viruses. However, the specific mechanisms underlying the severity of CA04 virus pathogenicity remain to be fully elucidated.

In this study, we undertook a unique bioinformatics approach to identify host processes that might be responsible for the enhanced pathology of CA04 in macaques. Other studies have used microarray analysis in conjunction with bioinformatics tools (e.g., Ingenuity Pathway Analysis (IPA) or DAVID [15]) to dissect the response to influenza virus infection, primarily at the level of the outcome of host signaling (i.e., differentially expressed transcripts) [13,14,16-18]. However, these studies typically have used such tools in isolation, and largely have not attempted to predict causative regulatory influences mediated by factors existing outside of the primary dataset. Here, we have combined process-level and cell-specific functional enrichment analyses of highly differentially expressed transcripts in CA04-infected lung tissues, identification of putative transcription factors involved in regulating differential expression, canonical pathway enrichment, and protein-protein interactions to develop an integrated pathogenicity-associated host response network. This novel approach revealed not only differentially regulated pathways and differences in immune cell populations that correlate with enhanced CA04 pathogenicity, but importantly, also identified critical regulatory intermediaries between these correlates. We suggest that the resulting network can be systematically explored for novel therapeutic intervention.

Results

CA04 may promote inflammation through increased expression of chemotactic molecules and caspase-1 induction and suppression of glutathione S-transferase (GST) expression

The CA04 virus induces more severe lesions in the lungs of cynomolgus macaques relative to a seasonal influenza virus isolate (A/Kawasaki/UTK-4/09 [H1N1; referred to KUTK4]) [11]. To identify host functions and pathway activity that might explain this difference, we compared the global transcriptional response of CA04- and KUTK4-infected lung tissues derived from the animals reported in [11]. Microarray analysis of RNA isolated from within gross lesions identified 101 differentially expressed (DE; See Methods) transcripts between the two infections on day 3 p.i. and 854 DE transcripts on day 7 p.i. More than half of the DE transcripts from day 3 were also DE on day 7, and the overlap between time points was significant (Fisher's Exact test, $P < 10^{-16}$). A table summarizing the DE test results is shown in Additional file 1, and total lists of DE transcripts from days 3 and 7 are available in Additional file 2. It is important to note that, while lung lesion severity varied between samples (Additional file 3), principle component analysis did not identify any relationships between the samples based on lesion severity (data not shown).

We first focused on genes commonly associated with inflammation and apoptosis (Figure 1). Enhanced up-regulation of interferon (IFN)-stimulated genes (ISGs) was observed on day 3 and day 7 p.i., indicating that CA04 induced more robust type I IFN signaling than did KUTK4 (Additional file 1). On day 3 p.i., three chemokines (CCL13, CCL2 and CCL3L3) were more up-regulated with CA04 infection; and these together with four others (CCL11, CCL4, CCL8 and CXCL3) were also more up-regulated on day 7. In general, the up-regulated chemokines perform pleiotropic recruitment activities against multiple immune cell subtypes; however, CCL2 and CCL3L3 exhibit potent chemotactic activity toward monocytes [19,20]. On day 7, we also observed enhanced expression of chemokine receptors that are abundant on monocytes (CCR1 and CCR2), interleukins that are primarily expressed and secreted by macrophages (IL10 and IL1A) and molecules that regulate lymphocyte activities (IL21R, IL4R, and IL7). Collectively, these observations are consistent with detailed histopathology analysis of immune cell infiltrates in macaque lungs infected

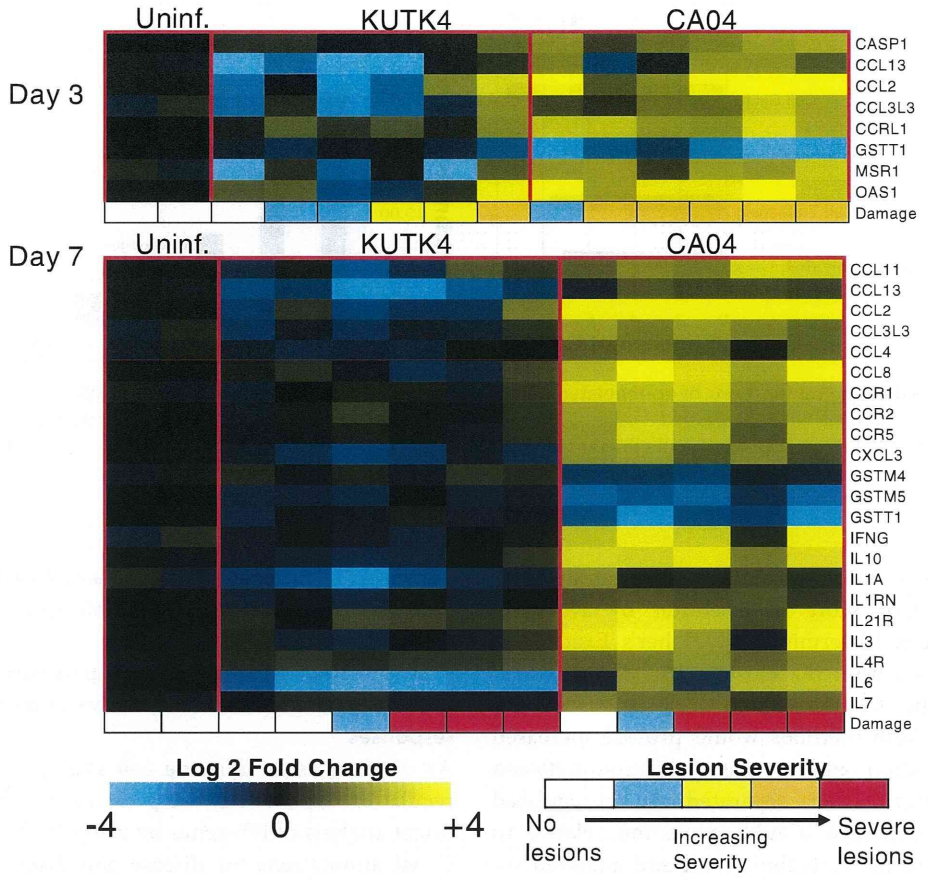

Figure 1 CA04 differentially expressed genes. The transcription profiles of select DE genes on day 3 (top) and day 7 p.i (bottom). Each column represents the gene expression profile of an individual macaque lung tissue sample; gene expression levels for two samples acquired from a uninfected animal are also shown (to the left) for reference. The lesion severity, as described in (11), for each sample is shown at the bottom (Damage) of each heat map. White indicates no lesions; blue represents lung lesions with severe alveolar wall thickening; yellow indicates mild lung lesions with some antigen-positive cells present; orange indicates severe lesions containing a significant number of virus antigen-positive cells; and red indicates severe lung lesions in which the alveolar space was filled with inflammatory cells.

with a similar pH1N1 strain [7], and support the suggested involvement of monocyte and lymphocyte sub-populations in the regulation of CA04 mediated pathology.

Particularly interesting was the observed up-regulation of caspase 1 (CASP1) in CA04-infected lungs. CASP1 is converted to its active form by the inflammasome signaling complex, which is activated by influenza virus infection [21,22]. In turn, CASP1 cleaves latent IL1β and IL18 leading to the secretion of the activated forms of these pro-inflammatory cytokines [23]. Consistent with increased CASP1 expression, IL1β and IL18 protein concentrations were elevated in the lungs of 2 of 3 CA04-infected macaques at day 7 (see Additional files in [11]). Taken together, these observations are indicative of increased inflammasome activity in CA04 infections.

Among the CA04-down-regulated transcripts, we observed several members of the cytosolic glutathione S-transferase (GST) protein family (GSTT1 on day 3; GSTT1, GSTM5, and GSTM4 on day 7). GSTs are antioxidant enzymes that participate in the inactivation of secondary metabolites formed as a result of oxidative stress, and thereby serve a protective role in the cell [24]. Similar suppression of GST transcription was observed in the kidneys of chickens infected with an HPAI virus strain [25], and it has been suggested that oxidative stress may be a key pathway involved in lung injury associated with HPAI virus infection [26]. Thus, the CA04-enhanced down-regulation of GST likely promotes oxidative injury in lung tissue.

CA04-infected macaques exhibit enhanced inflammatory and cell cycle gene expression

IPA and DAVID were next used to identify functional differences between CA04 and KUTK4 infections (Figure 2). IPA uses Fisher's Exact test to determine the enrichment significance for each function annotated in the IPA database. DAVID, on the other hand, uses clusters of related annotations built from several annotation

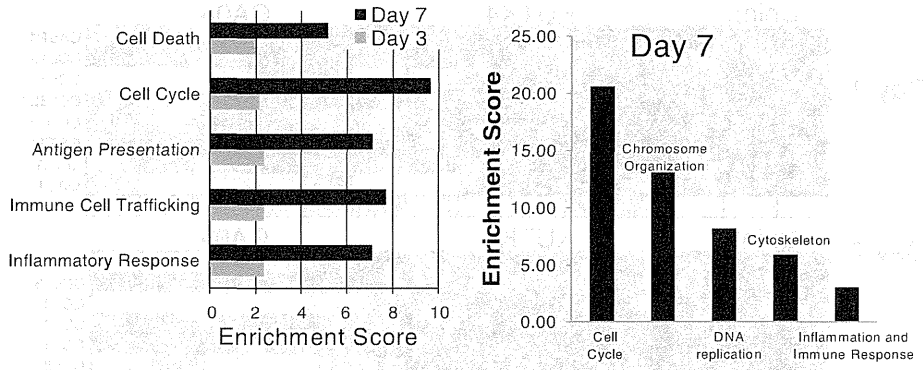

Figure 2 The enriched biological functions of up-regulated genes, as determined by (left) IPA and (right) DAVID. The IPA enrichment score is reported as the -log10 of the FDR-adjusted P-value. DAVID clusters redundant or highly related annotations into clusters and reports the enrichment as the –log10 of the average P-value of the terms in each annotation cluster. DAVID did not identify any significant annotations for genes which were up-regulated on day 3, and only enrichments for day 7 are shown.

databases (e.g., gene ontology, pathway data), and calculates the enrichment score of a cluster by averaging unadjusted *P*-values (determined by Fisher's Exact test) of the annotations within the cluster [15]. We reasoned that analyzing the CA04 pathogenicity-associated DE transcripts using both methods would provide increased functional information and/or cross-validation between methods. DE transcripts were separated into up-regulated or down-regulated groups (CA04 expression relative to KUTK4 expression) for each time point and analyzed separately. All significantly enriched (FDR-adjusted *P*-value < 0.01) IPA functional annotations and DAVID functional clusters are included in Additional files 4 and 5, respectively.

IPA analysis of up-regulated DE transcript identified enrichment for 'Inflammatory Response', 'Immune Cell Trafficking', and 'Antigen Presentation' at day 3 p.i., consistent with the increased immune cell infiltrates observed in the original pathology examination [11]; no significant clusters were identified in this time point using DAVID analysis (Figure 2). Transcripts up-regulated on day 7 were enriched for many of the same IPA categories found on day 3, and both IPA and DAVID identified 'Cell Cycle' as being the most highly enriched, up-regulated process on day 7. More specific categories, such as positive or negative regulation of cell cycle, did not provide conclusive evidence as to the effect of the regulation on cell cycle (see Additional files 5 and 6). Among the downregulated transcripts, we saw enrichment only with IPA at the 7 d p.i. time point, and only in two categories with no obvious connection to the host response to infection (e.g., 'Psychological Disorders' and 'Neurological Diseases') (Additional file 4). These results suggest that functional differences between CA04 and KUTK4 initiate early after infection and are amplified later in the infection, while both analyses identified the cell cycle as the most prominently enriched functional annotation for transcripts that were highly DE with CA04 infection.

CA04 infection enhances gene expression associated with both early/innate and late/adaptive immune cell responses

As differences in immune cell subtypes can also be key regulators of pathology, we performed a cell-specific, functional analysis of DE genes by using IPA's specialized functional annotations for disease and discovery. In addition to providing a general description of function for a category of enriched genes, this approach captures details such as the cell or tissue type in which the gene exerts its function and the orientation of regulation. For down-regulated genes on both days, few categories satisfied our enrichment criteria (FDR-adjusted $P < 0.05$), and no immune cell-specific annotations were enriched. In contrast, the up-regulated gene lists produced more than 500 enriched categories for both days, many of which exhibited functions pertaining to the activation, chemoattraction, infiltration and development of specific immune cell subtypes (a complete list of significant categories for up- and down-regulated genes is shown in Additional file 6). To obtain a clearer representation of how CA04-induced gene expression influences different immune cell subtypes, we categorized enrichment by cell-type and function and employed a heat map to show how the enrichment was distributed in relation to time (summarized in Figure 3; see also Additional files 7 and 8 for a detailed heat map). On day 3, the immune cell subtype most prominently affected by CA04-enhanced gene expression was macrophages, for which categories related to activation, accumulation, and chemotaxis were very highly enriched (Figure 3). T lymphocytes and neutrophils were also broadly affected at this time point, although annotations were not as significant as they were for

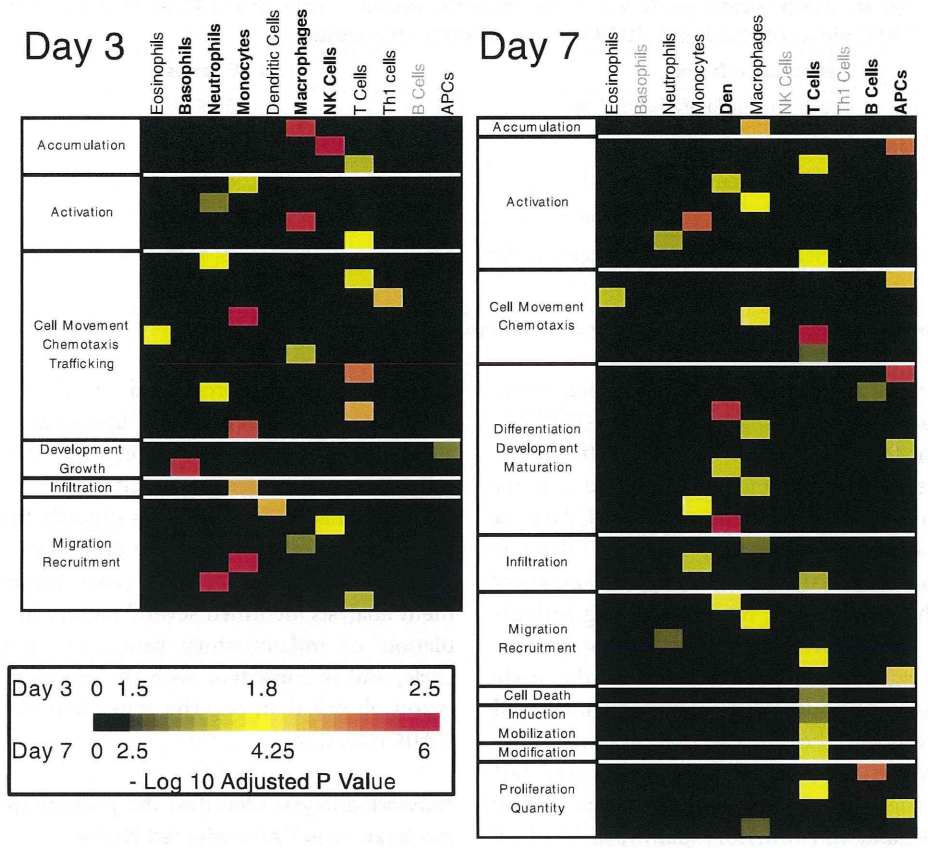

Figure 3 Cell-specific CA04-induced functional enrichment. Cell-type specific enrichment of CA04-induced up-regulated DE transcripts is shown for day 3 (left) and day 7 (right) post-infection. Enriched cell-type-specific IPA annotations (FDR-adjusted *P*-value < 0.05) were sorted according to function (categories shown to the left of each panel) and cell-type (shown at the top of each panel), and the level of enrichment was illustrated by a heat map showing the -$\log_{10}$-adjusted *P*-value for each annotation. Because *P*-values associated with highly related annotations (e.g., 'Cell Movement', 'Chemotaxis' and 'Trafficking') were grouped together, a cell type could be enriched several times within a particular category. Cell types with the highest enriched categories are shown in boldface (top), whereas cell types for which no categories were enriched are shown in grey. All other cell types, exhibiting intermediate levels of enrichment in various categories, are represented in normal black text. A color-key for the heat map is shown at the lower left of the figure, in which day 3 *P*-values are represented by the range above the color bar and day 7 *P*-values are indicated below the bar.

macrophages. In addition, some enrichment was observed for genes affecting the migration of dendritic cells, eosinophils, and natural killer cells. On day 7, no significant terms related to natural killer cells or basophils were observed, and enrichment of the activation and migration of neutrophils was quite minor, demonstrating that several aspects of the innate immune response had tapered. Instead, we found high enrichment for B lymphocyte proliferation and broad enrichment in categories associated with T lymphocyte chemotaxis, infiltration, proliferation, activation, and death. In addition, several functions of antigen presenting cells (APCs) were more up-regulated in the CA04 infection. The original pathology results [11] did not specify the cell types that comprised the enhanced inflammatory infiltrates on day 3 in CA04-infected lung tissue, but they did identify an increase in APCs on day 7 p.i. [11]. These data point to an increased presence and activation of primarily innate immune cell subtypes early after infection, followed by enhanced influx and activity of APCs and adaptive immune cell types in CA04-infected macaque lungs.

Inflammatory and apoptotic transcription factor binding sites are enriched in CA04-specific DE genes

Next, transcription factor activity was determined by identifying transcription factors whose promoter sequences were highly enriched among the genes that were DE between the CA04 and KUTK4 infections. We used the GATHER software tool [27] which matches transcription factors with their experimentally proven binding sites and determines whether there is significant enrichment of a transcription factor binding site within a given gene list. The transcription factors that matched the most highly enriched promoter sequences for each time point after infection are shown in Table 1.

Table 1 Top three transcription factors enriched for genes differentially expressed (DE) on day 3 or day 7 post-infection between CA04 virus-infected and KUTK4 virus-infected lung tissue

	Transcription factor	% DE genes*	-ln (*P* Value)
Day 3	Interferon Regulatory Factor 7 (IRF7)	55.8	3.54
	Forkhead box P3 (FOXP3)	53.8	4.08
	Hepatic nuclear factor 1 (HNF1A)	42.3	7.18
Day 7	MYC-associated zinc finger protein (MAZ)	74.6	6.05
	Nuclear transcription factor Y alpha (NFYA)	24.4	8.31
	Paired box 5 (PAX5)	15.7	5.56

* The percentage of genes DE on each day that were regulated by each transcription, according to GATHER®.

Day 3 promoter sequence enrichment identified interferon regulatory factor 7 (IRF7), forkhead box P3 (FOXP3), and hepatic nuclear factor 1 (HNF1A). IRF7 is activated by toll-like receptors and RIG-I, and its increased activity likely reflects the increased replication of CA04 in macaque lungs observed on day 3 p.i. [11]. FOXP3 is a transcription factor specific to a subset of regulatory T lymphocytes that are known for suppressing inflammatory reactions [28]. FOXP3 promoter site enrichment thus implicates regulatory T lymphocytes in the early phase of CA04 infection, possibly reflecting an increased requirement to confront the CA04-enhanced inflammatory response. HNF1A has a wide range of functions, but recent evidence suggests that it may serve as a link between metabolic and inflammatory pathways involved in atherosclerosis [29].

On day 7, nuclear transcription factor Y alpha (NFYA), paired box 5 (PAX5) and MYC-associated zinc finger protein (MAZ) transcription factor activities were highly significant. NFYA directly induces transcription of many pro-apoptotic genes and is implicated in p53-mediated apoptosis [30,31]. To determine if there was any significant association between the genes regulated by NFYA and the genes annotated with 'Cell Death' in IPA, we used a one-sided Fisher's exact test, but we did not observe significant overlap between these gene sets (only 13.9%, $P = 0.18$, of the DE genes that were annotated for 'Cell Death' were also regulated by NFYA; further limiting the gene set to only upregulated genes resulted in a non-significant overlap of 22.0%, $P = 0.69$). PAX5 is a specific marker for B cells [32], and its enrichment is consistent with the enhanced B cell presence and activity suggested by our cell type-specific functional enrichment analysis for day 7 (Figure 3). The enriched transcription factor affecting the greatest percentage of CA04 DE genes was MAZ, whose regulation is strongly correlated to inflammation and other common inflammatory markers (e.g., IL6, IL1β) [33]. Significant overlap was identified between the genes annotated for 'Inflammation Response' by IPA and genes that contained the MAZ binding sequence, when considering all DE genes on day 7 (72.3% overlap, $P = 0.05$), but the overlap was not significant when only considering genes upregulated on day 7 (63.9% overlap, $P = 0.25$). But within the set of upregulated genes, there was a very significant overlap between the MAZ regulated gene set and the set of genes annotated with "Cell cycle" in IPA (74.5% overlap, $P = 0.003$). This suggests that MAZ is closely associated to cell cycle activity despite its established role in inflammation. Overall, transcription factor enrichment analysis identified several factors involved in the regulation of inflammatory processes, apoptosis and cell cycle, and is consistent with the enriched functional processes observed among the genes that were more DE with CA04 infection.

Network analysis identifies the primary gene expression moderators in CA04-infected tissue

Thus far, we have used enrichment analyses to isolate individual processes differently regulated by the CA04 virus. Next, we applied a network approach to determine the regulatory interactions that may be coordinating gene expression. Within the IPA-curated human protein-protein interaction (PPI) network, we searched for subnetworks that were highly populated with DE genes, and then we used the protein degree (i.e., the "centrality", or the number of interactions a protein has) to identify highly connected hub proteins within each subnetwork. Hub proteins are often more essential than non-hub proteins [34,35] and may play critical roles in regulating the network's overall biological function [36].

An initial assessment allowing for a network size of up to 70 members did not identify any significant subnetworks at 3 d.p.i.; however, several were detected among DE genes on day 7. The most significant of these was highly enriched for 'Cell Growth and Proliferation' and 'Inflammatory Response' annotations (Fisher's exact test P-values $< 10^{-15}$) and was centered on four hub genes: IFNG, MYC, IL6, CDKN1A (Figure 4A, a more detailed illustration in Additional file 9). A more restricted analysis – limiting the number of subnetwork members to 35 genes – recapitulated this result, yielding three smaller subnetworks with the same hub genes, each enriched for unique biological functions (Figure 4B-D). IFNG, an antiviral cytokine with potent macrophage activation activity

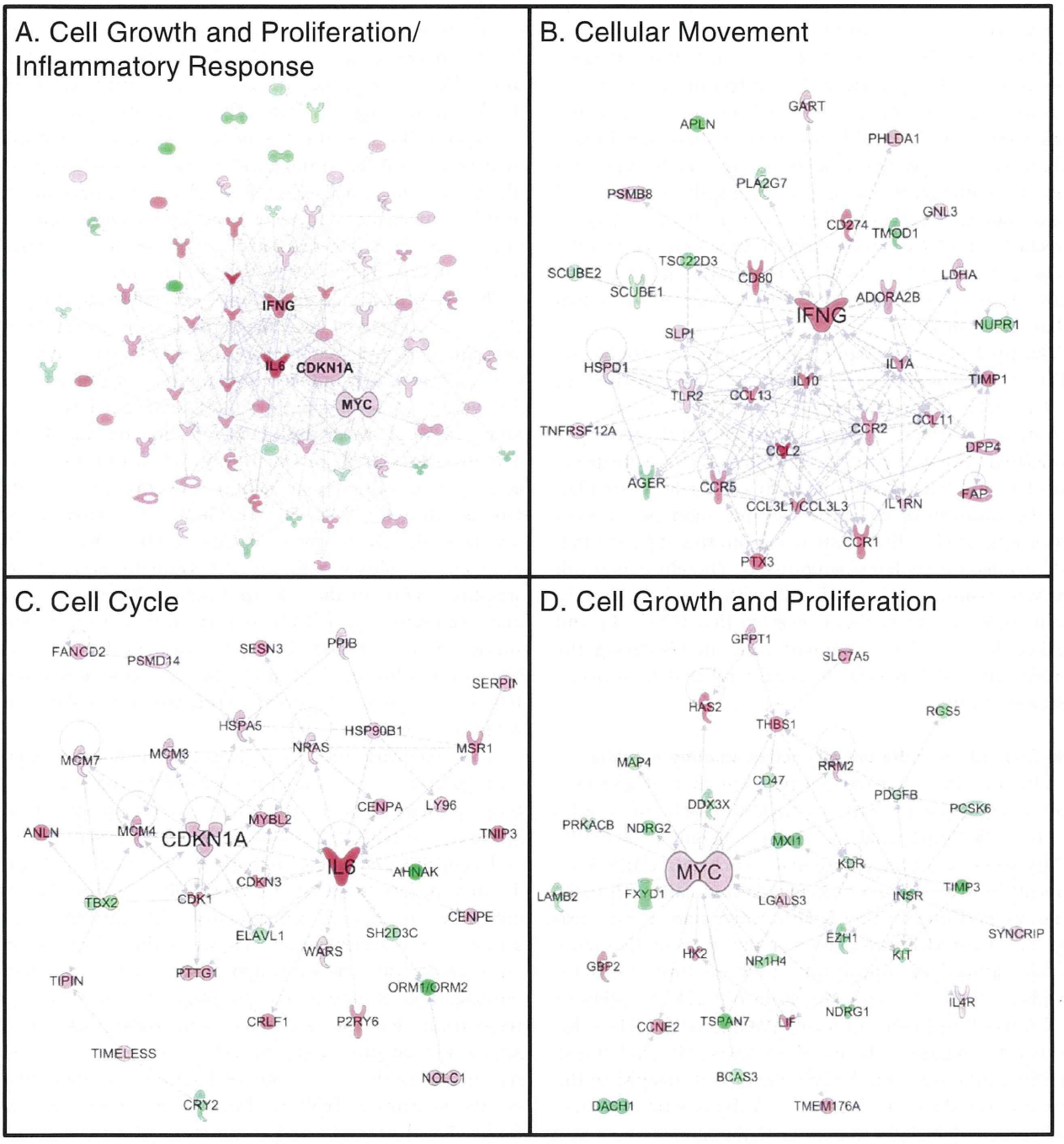

Figure 4 Protein-protein interaction subnetworks differentially regulated in CA04-infected lung tissue. IPA was used to determine subnetworks within the human PPI which were highly populated with genes differentially expressed in CA04-infected tissue on day 7 p.i. We show the top networks (based on their network score) when **(A)** up to 70 proteins are allowed in the subnetwork and **(B-D)** the top 3 networks when the size of the subnetwork is limited to 35 proteins. Each network is labeled according to its most enriched biological function. Proteins that's transcript is up-regulated or down-regulated in CA04-infected lung relative to KUTK4-infected lung are colored red and green, respectively. Additional file 9 provides a larger illustration of panel A with all protein labels shown.

and immunoregulatory functions in adaptive immunity, exhibited high centrality in a largely upregulated subnetwork that was enriched for 'Cellular Movement' (*P*-value = 3.96E-16) (Figure 4B). Notably, 8 of the 35 proteins in this subnetwork were also members of the 'Role of Cytokines in Mediating Communication between Immune Cells' canonical pathway (Additional file 10), thus implicating IFNG as a putative regulator of the increased

late adaptive immune response in CA04 infections. The second subnetwork, consisting primarily of upregulated genes, was highly enriched for 'Cell Cycle' functions (*P*-values = 5.4E-7) (Figure 4C). Interestingly, both the pro-inflammatory cytokine, IL6, and the cyclin-dependent kinase inhibitor, CDKN1A, appeared as hubs within this subnetwork, suggesting a novel connection between the host inflammatory response and the highly enriched, cell cycle-associated processes that occur in the late stage of infection. CDKNIA's involvement further suggests that the broad category of "cell cycle" is likely to be more specifically related to cell cycle arrest, since CDKN1A is a known inhibitor of cell cycle.

The final network (Figure 4D) is centered around the MYC transcription factor, which was up-regulated in CA04 infections, and enriched for 'Cell Growth and Proliferation' (*P*-value = 2.0E-6). Generally, MYC activity is associated with cell proliferation, but MYC overexpression has also been associated with cell cycle arrest [37]. As the majority of the genes in this subnetwork were down-regulated in the CA04 infection, this suggests that cell proliferation is being suppressed. Therefore, network analysis results were consistent with previous enrichment analyses, and further suggested that IFNG, IL6 and CDKN1A may play prominent roles in regulating the severity of CA04-associated disease relative to seasonal influenza virus.

The CA04-KUTK4 differentially active immune network

Finally, we integrated the cell type enrichment and promoter sequence enrichment analysis with the results of the subnetwork analysis into a coherent map of the CA04-induced immune response. To better clarify how critically positioned hub proteins affected canonical pathways, we performed pathway enrichment and selected two significantly enriched pathways – the 'Role of Cytokines in Mediating Communication between Immune Cells' (Additional file 10) and the 'Role of CHK Proteins in Cell Cycle Checkpoint Control' (Additional file 11) – for further examination. These pathways were chosen because (i) they contained several of the genes that mapped to the subnetworks shown in Figure 4, (ii) the function of these pathways matched the function of the subnetworks and (iii) the pathway was highly enriched for genes DE in CA04-infected lung (FDR-adjusted $P < 0.05$, Additional file 12 provided details on the pathway enrichment analysis). Furthermore, these pathways contain cell-specific expression information for many of the genes involved in this study.

Interactions within each pathway were diagrammed, taking into consideration activating and inhibitory relationships (Figure 5). Next, we searched the IPA PPI database for interactions linking the pathways, and we identified the MAZ transcription factor, which, incidentally, was also one of the most enriched transcription factors in the promoter enrichment analysis. Specifically, MAZ activates expression of CDKN1A (a component of the CHK pathway and a 'Cell Cycle' subnetwork hub), thereby controlling cell cycle progression through the G1 checkpoint [38] (see also Figure 5). MAZ is also activated by both IL1 and IL6 through MAP kinase-dependent phosphorylation in human cells [39] and is a member of the MYC complex. Thus, MAZ, in a network context, appears to be a critical intermediary between the identified subnetworks and potentially their biological function.

The newly identified interactions were added to the previously mentioned pathway diagrams, and gene expression and differentially active transcription factors were overlaid to depict the CA04-induced immune response (Figure 5). The viral RNA sensing pathway was added since IRF7 was found to be active in the CA04-infection on day 3 [40]. The precise cell-type in which cell cycle arrest and MAZ activity occurs is an open question; therefore, the specific type of infected cell was not clarified in the network, and was labeled only as 'Infected Cell'. CA04-enhanced DE genes were colored red, and differentially active transcription factors or the cell types their differential activity may represent (e.g., FOX3P is regulatory T lymphocyte-specific) were boxed in red (if enriched on day 3 p.i.) or black (if enriched on day 7 p.i.). The full names and Entrez IDs for all genes depicted in Figure 5 are listed in Additional file 13.

In this network, only four genes (all chemokines) were up-regulated on day 3 p.i. (demarcated by red stars) and their primary role is the chemoattraction of innate immune cell types (NK cells, neutrophils, macrophages, and dendritic cells). The activity of IRF7 and regulatory T lymphocytes (implicated by FOX3P enrichment) was enhanced on day 3, as evidenced by the promoter sequence enrichment (Table 1). Major differences in the immune response network appeared on day 7 p.i., with upregulation of additional chemokines and cytokines responsible for the activation and differentiation of innate and adaptive immune cells (see the non-starred red nodes on the right panel of Figure 5). In particular, interferon gamma (IFNG), which is important for cytotoxic T cell function and elimination of virus-infected cells, was up-regulated (Figure 5, right). Additionally, upregulation of IL6 and IL10 (Figure 1) can lead to B lymphocyte activation, while the presence of B lymphocytes was supported by the enrichment of the PAX5 promoter sequence (Table 1). Increased immune cell presence and continued IL6 (and/or IL1) production would allow the simultaneous upregulation of MAZ activity, thereby impacting inflammatory and cell cycle signaling through CDKN1A transcription (Figure 5, see the MAZ transcription factor node and its incoming and outgoing edges). The cell cycle is further impacted by MAZ's

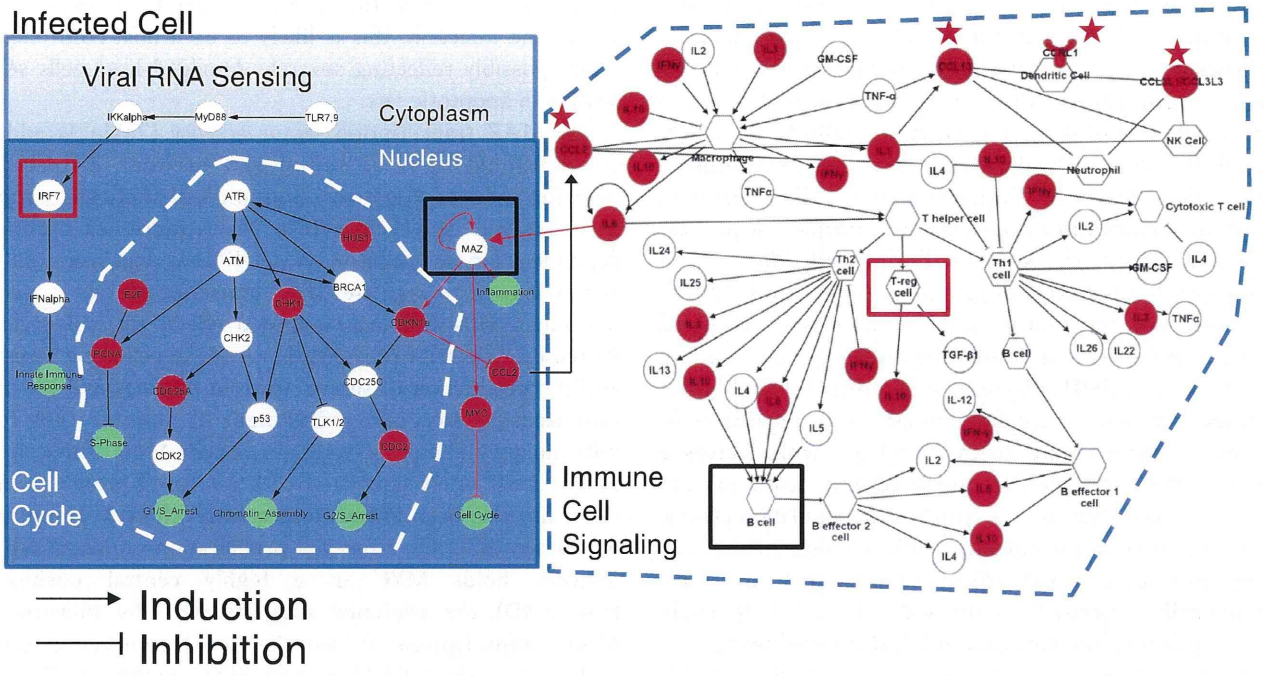

Figure 5 Differential response network between CA04- and KUTK4-infected lung tissue. Highly enriched canonical pathways identified by using IPA (namely, 'Role of CHK Proteins in Cell Cycle Checkpoint Control', enclosed by the white, broken line and 'Role of Cytokines in Mediating Communication between Immune Cells', enclosed by the blue broken line) were integrated by using highly enriched transcription factors and IPA's Expert Curated protein-protein interaction database. Newly identified interactions which are not part of the canonical pathways are illustrated with red arrows, whereas interactions belonging to the canonical pathways are colored black. Proteins that are differentially up-regulated in CA04-infected lung tissue relative to KUTK4-infected lung tissue on day 7 p.i. are colored red. Three chemokines and one chemokine receptor which were up-regulated in the CA04 infection on day 3 as well are demarcated by red stars adjacent the protein. Transcription factors or cell types for which the transcription factors surrogate as cell markers are boxed in red or black representing enrichment on day 3 or day 7 p.i., respectively.

interaction with MYC, which we suspect is involved in cell cycle arrest, since many of MYC's target proteins were downregulated during infection (Figure 4D). Overall, our data identify a novel, transcriptionally active link between influenza-induced inflammation and cell cycle arrest (i.e., the MAZ transcription factor) and suggest that inflammation-induced, MAZ-dependent cell cycle disruption may be responsible, at least in part, for apoptosis and tissue injury related to influenza virus pathogenicity.

Discussion

GO and pathway enrichment studies can detail many key aspects of the host response, but these biological functions and processes must be integrated to create host response models capable of linking the effects of transcription to the molecular and signaling events driving those effects. The differentially regulated host response network presented here represents a novel effort to combine various, independent analyses of DE genes into a coherent protein-protein/protein-cell type interaction architecture. This approach allowed us to link immune cells and their inflammatory activities with cell cycle regulation through the identification of a transcription factor (i.e., MAZ) that acts as an intermediary between these functional correlates of pH1N1 pathogenicity. While many studies have used microarrays to identify global transcriptional changes that characterize the host response to different influenza viruses in different systems, our integrated approach allows for a more specific understanding of the mechanisms of influenza virus-induced pathogenicity.

Prior to developing the differential host response network, we validated our transcriptional data by showing that it was, indeed, indicative of the pathological differences between the two infections (Figures 2 and 3). Initial functional enrichment results confirmed that the biological processes observed during the pathology examination (i.e., enhanced inflammation and increased immune cell infiltrates [11]) were also detectable in the transcriptional differences between CA04- and KUTK4-infected lung tissue. Therefore, it seems very likely that any additional functionality or pathway information derived from these data should have a highly correlative, if not causal, relationship with the enhanced pathology of the CA04 infection.

The enhanced ability of the CA04 virus to replicate in lung tissue does not lead to large differences in the host

response on day 3 of the infection. The enhanced IRF7 transcription factor activity in CA04-infected lung is consistent with the increased toll-like receptor/RIG-I signaling one expects when there are increased levels of viral single-stranded RNA present in a sample. Enhanced chemoattraction of NK cells, dendritic cells, neutrophils, and macrophages is consistent with the enhanced inflammatory infiltrates observed on histopathologic examination. Most interesting is the implied difference in regulatory T lymphocyte populations, evident from FOXP3 promoter site enrichment, and the potentially enhanced IL-1β expression as a result of greater up-regulation of CASP1 in the CA04 infection. Regulatory T lymphocytes manage immune system homeostasis, and imbalances in T cell populations are often associated with increased inflammation and immune-mediated cell death [41,42]. The implicated regulatory T lymphocyte population change may be a factor in the enhanced inflammation and cell damage observed in CA04-infected lung tissue on day 3 p.i. Additionally, increased CASP1 induction of IL-1β could further promote inflammation in CA04-infected tissue.

In a previous study, lung samples infected with a highly pathogenic and mildly pathogenic pH1N1 virus were compared to KUTK4-infected tissue on day 1 p.i., and similar to the work presented here, enhanced inflammation and immune cell infiltration were identified as correlates of increased pH1N1 pathogenicity [7]. This study found NFκB mediated transcription as a potential mechanism of enhanced pathogenesis, but the degree to which the observed gene regulation was independent of viral replication is unclear. Furthermore, of the 101 transcripts DE on day 3 of our study, only 11 were also DE in the previous work. While this does represent a significant overlap (Fisher Exact test; $P < 0.001$), none of the genes identified as DE early in pH1N1-infected lung tissue in both studies are related to the immune response. In all, both studies show that early in the course of the infection, there is no obvious dysregulation of the host response with the potential exception of an imbalance in regulatory T lymphocytes, noted above.

By day 7 p.i., several mechanisms up-regulated in CA04-infected lung tissue can account for the continued enhanced pathology [11]. Lung tissue infected with CA04 on day 7 p.i. showed sustained activation and accumulation of immune cells despite the absence of replicating virus. In addition to the increased immune cell signaling and the activation of the adaptive immune response (evident by the B cell-specific PAX5 promoter site enrichment and GO enrichment analysis shown in Figure 3) we observed increased enrichment for cell cycle arrest. Cell cycle arrest has a complicated relationship with virus replication and the immune response. Arrest during G1 phase promotes greater influenza virus replication [43], but cell cycle arrest also often occurs in cells prior to apoptosis [44]. Given that cell cycle arrest was observed late in the infection, this is likely to be a host-controlled event, possibly reflecting severely damaged host cells selecting an apoptotic fate.

The MAZ transcription factor activity (Table 1) adds an additional layer of complexity between influenza infection-induced apoptosis, immune cell trafficking, and the observed cell cycle arrest. MAZ increases cyclin-dependent kinase inhibitor 1A (CDKN1A) expression [38] and is known to regulate MYC transcription [45] – two molecules with seemingly opposed effects on cell cycle. Increased CDKN1A transcription leads to cell cycle arrest and the production of serum amyloid A (SAA), which in turn leads to increased recruitment of immune cells to inflammatory sites [38,39,46]. Increased MYC transcription is most often associated with increased proliferation but it has also been linked to increased cell cycle arrest in fibroblasts [37]. Since the cell proliferation enriched subnetwork holds MYC in a highly central position (Figure 4D), the evidence suggests that the increased MAZ transcription is simultaneously inducing cell cycle arrest via CDKN1A and MYC pathways. Since lymphocytes typically proliferate in organized lymphoid tissue (e.g., in lymph nodes) and our samples were collected from within infected lung lesions, we suggest that regulation of cell cycle gene expression primarily occurs in infected epithelium or pneumocytes. Lastly, the fact that virus could not be isolated on day 7 for CA04 infected-lung tissue suggests that activation of the MAZ pathway may be in response to an overly aggressive immune response rather than virus replication. Further, while MAZ protein levels are directly correlated with chronic inflammation, the anti-inflammatory suppression of CCL2 transcription by CDKN1A [47] was not observed in our microarray data. Thus, there are multiple interactions involving MAZ which we feel are suitable targets to mitigate inflammation during moderate to highly pathogenic influenza infections. Further study validating the significance of MAZ transcription to local inflammation is warranted.

The differentially regulated network developed here elucidates differences between a low pathogenic and a moderately pathogenic infection, and is likely a suitable model of enhanced pathology in humans, as macaque models of influenza virus infection are considered to be one of the best surrogates of human infection [48]. Several chemokines and interleukins that are up-regulated in the CA04 infection are also up-regulated in the lungs of macaques infected with HPAI H5N1 virus [12]; however, promoter enrichment analysis of avian virus-infected lung tissue may be needed to provide greater clarity on the precise mechanisms active during a highly pathogenic infection. Ultimately, we intend to develop a mathematical model that can quickly identify the correlates of pathogenicity

from microarray experiments to equate transcriptional regulation to infection severity in humans. The network presented here is the first step toward developing such a model.

Conclusions

In summary, CA04-infected macaque lungs showed a prolonged immune response that continued beyond the duration of the local virus infection. The failure of the negative feedback mechanism that exists between MAZ, CDKN1A, and macrophage cell migration (induced by CCL2) could have caused this prolonged inflammation, thereby promoting enhanced cell cycle arrest and apoptosis. The interplay between MAZ induction, immune cell signaling, and inflammation must be finely tuned, and failure to maintain an appropriate, balanced response between these three factors could explain, at least in part, the increased pathogenicity of CA04 and other pH1N1 viruses. Further studies are needed to address the interchange between MAZ, the cell cycle, and the immune response, as well as the effects of this interplay on influenza virus-induced disease pathology. Overall, our strategy of linking functional annotations to the protein-protein interaction networks is suitable for identifying the key mechanisms driving the observed phenotypes.

Methods

Ethics statement

As previously reported in [11], all experiments were performed in accordance to the Guidelines for the Husbandry and Management of Laboratory Animals of the Research Center for Animal Life Science at Shiga University of Medical Science, Shiga, Japan and were approved by the Shiga University of Medical Science's Animal Experiment Committee and Biosafety Committee.

Tissue samples

Lung tissues used for microarray studies were obtained from thirteen female cynomolgus macaques infected with influenza viruses as previously described [11]. Briefly, six animals were inoculated with influenza A/California/04/2009 (H1N1; referred to as CA04), a 2009 pH1N1 virus isolate; six animals were inoculated with influenza A/Kawasaki/UTK-4/2009 (H1N1; referred to as KUTK4), a seasonal isolate; and one uninfected animal served as a negative control. On days 3 and 7 post-infection (p.i.), lung tissues were harvested from the middle and lower lung lobes of three animals in each infection group (N = 26 total samples were collected), and all but three samples were collected from visually apparent virus-induced gross lesions. Two lung tissue samples were obtained from middle and lower lobes of the uninfected animal at the start of the experiment. A more detailed description of tissue sample location and lesion severity is provided in Additional file 3.

RNA extraction

Macaque lung tissues were placed in the RNA stabilization reagent RNAlater (Ambion, CA) and stored at -80°C. All tissues were thawed together and homogenized (2 minutes at 30 Hz) by using a TissueLyser (Qiagen, Hilden, Germany), following the manufacturer's instructions. Total RNA was extracted from homogenized lung tissues with the RNeasy Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's recommendation.

Microarray, normalization, and statistical analysis

Extracted lung RNAs were sent to Takara Bio (Otsu, Shiga, Japan) for microarray analysis. Briefly, sample integrity and quantity were measured with the Agilent 2100 Bioanalyzer (Otsu, Shiga, Japan), which resulted in the exclusion of three samples due to poor RNA quality. In total, two mock-infected samples and at least five infected samples from each infection group at each time point were sent for subsequent microarray analysis. Cy3-labeled cRNA preparations were hybridized with rhesus macaque arrays (Agilent Microarray Design Identification Number 015421) for 17 h at 65°C. Feature Extraction Software version 7 (Agilent Technologies) was used for image analysis and data extraction, and Takara Bio provided whole array quality control metrics.

Per chip probe intensity normalization and differential expression (DE; DE also denotes "differentially expressed", e.g., DE genes) analysis was performed by using GeneSpring GX version 11.0.2 (Agilent Technologies). Individual probe quality control was performed by using the GeneSpring default flag settings, requiring each probe to satisfy the flag conditions for at least 4 of the 25 samples. DE genes were identified between virus infection groups by use of one-way ANOVA complemented with a Tukey Honestly Significant Difference test, followed by a False Discovery Rate (FDR) correction (Benjamini-Hochberg). Criteria for DE were as follows: an absolute fold change > 2 and an FDR-adjusted P-value < 0.01. All microarray data have been deposited Gene Expression Omnibus (series number GSE39018) in accordance with Minimum Information About a Microarray Experiment (MIAME) guidelines.

Ingenuity pathway analysis

Functional and pathway enrichment analyses were performed with Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood, CA, USA), after matching DE macaque genes with their human orthologs by using GenBank Accession identification numbers. For all functional and pathway enrichments, we required the Benjamini-Hochberg corrected P-value to be < 0.05. For

cell-type specific functional enrichment analysis, significant function annotations were separated into their function and respective cell types (e.g., 'recruitment of neutrophils' was split into its function, 'recruitment', and the cell-type, 'neutrophil'). Further, functions that were related were grouped together (e.g., 'Cell Movement', 'Chemotaxis', and 'Trafficking'); thus, one cell-type may be significant more than once in each category. Additionally, leukocytes and mononuclear leukocytes were grouped together. Many of the cellular processes (e.g., *chemotaxis, trafficking* and *cell movement*) are highly related, we grouped them accordingly. Instances in which the same cell type is categorized into two cellular states (e.g., annotations related specifically to memory T cells or the more broad characterization of simply a T cell) were generalized to *T cells*. Both of these factors can make a single cell type appear enriched multiple times in each functional category. For each day, genes that were up-regulated or down-regulated when comparing CA04-infected lung tissue to KUTK4-infected lung tissue were analyzed separately.

DAVID gene ontology analysis

The Genbank accession numbers of genes which were up or down-regulated when comparing CA04-infected lung tissue to KUTK4-infected tissue were analyzed in DAVID [15,49] using DAVID default settings. In addition to the enrichment scores for each annotation cluster produced by DAVID, we also determined the cluster size; i.e., the number of individual annotations which satisfied an FDR-adjusted P-value <0.01.

Transcription factor promoter sequence enrichment analysis

The Genbank accession numbers of DE genes were analyzed in the GATHER [27] website to identify transcription factors whose promoter sites were highly represented within the data. GATHER employs the TRANSFAC 8.2 database, which contains data on transcription factors and their experimentally validated binding sites. The human genome R17 from the UCSC Genome Database was matched to high quality matrices for vertebrate regulator elements by applying the default score thresholds recommended by TRANSFAC. Transcription factor binding sites found 1200 bases upstream and 200 bases downstream of an annotated transcription start site were linked to their RefSeq IDs, which were then mapped to their Entrez Gene IDs based on the cross-references in the Entrez Gene database. GATHER scores the enrichment by using a *P*-value developed from the distribution of Bayes factors developed from randomly sampling 10,000 genes. Significant transcription factors were required to have an adjusted $P < 0.05$. Full details of the algorithm and justification for using the Bayes-based *P*-value is available in the original GATHER publication [27].

Subnetwork construction

All subnetworks were constructed from the IPA PPI network using IPA's internal algorithm. Briefly, the IPA algorithm identifies subnetworks by optimizing the interconnectivity and number of user genes (genes DE on each day) under the constraint of the selected network size. For the studies described here, we limited subnetworks to experimentally validated interactions identified in humans, and we performed iterative analysis of networks restricted to 35 or 70 total members. The degree of enrichment of DE genes in each subnetwork was indicated by the $-\log_{10}$ of the right-tailed Fisher's Exact Test). Networks were constructed for each day separately.

Network integration

The results from the promoter sequence enrichment analysis and pathway enrichment analysis were integrated by using protein-protein interaction data with the IPA interaction database. Stringent conditions were applied to identify binding interactions between the protein MAZ and two highly enriched, canonical pathways, namely "Role of CHK Proteins in Cell Cycle Checkpoint Control" and "Role of Cytokines in Mediating Communication between Immune Cells." We required all added protein-protein interactions between these two networks to be in the Ingenuity Expert Findings and Ingenuity ExpertAssist Findings, and further required that the interaction had been verified in human lungs by using the setting within the IPA software. For completeness, we also included the viral RNA sensing pathway, as described in [40], to illustrate IRF7 activity on day 3 p.i.

Statistical analysis

All tests for significant overlap between two gene lists were done in R using the one-sided Fisher's exact test for enrichment. When determining the overlap between genes annotated with a select IPA term and genes containing a selecting binding sequence (as determined by GATHER), the gene symbols were first converted into unique gene identifiers using the DAVID gene ID converter.

Additional files

Additional file 1: Summary of the number of probes DE and the number of DE probes which are chemokine ligands and receptors (CCL/R), interleukins (IL) or interferon stimulated genes for each day. Numbers in parenthesis show the number of up (left) and down regulated genes (right). We also show the number of genes DE on both days (intersection).

Additional file 2: All genes found differentially expressed on day 3 or 7 post infection.

Additional file 3: Microarray data description file. File describes: from which lobe the RNA was isolated; which virus the animal was infected with; the day the sample was collected; the severity of the lesions from which the sample was isolated, and the amount of virus isolated from the region.

Additional file 4: Work Book Explanation: all genes lists were separated into up- or down-regulated when comparing CA04-infected tissue to KUTK4-infected. DE gene's Genbank Accession IDs were uploaded into Ingenuity and the benjamini hochberg corrected P-value was used to quantify enrichment.

Additional file 5: Work Book Explanation: all genes lists were separated into up- or down-regulated when comparing CA04-infected tissue to KUTK4-infected. DE gene's Genbank Accession IDs were uploaded into DAVID and the benjamini hochberg corrected P-value was used to quantify enrichment.

Additional file 6: Work Book Explanation: cell specific gene ontology enrichment. All genes lists were separated into up- or down-regulated when comparing CA04-infected tissue to KUTK4-infected. DE gene's Genbank Accession ID was uploaded into Ingenuity and the benjamini hochberg corrected P-value was used to quantify enrichment. Here we report all categories with a FDR adjusted P-value < 0.05.

Additional file 7: Cell-specific CA04-induced functional enrichment on day 3 PI. This is an enlarged illustration of Figure 3A which provides information on the specific function of each enriched IPA annotation.

Additional file 8: Cell-specific CA04-induced functional enrichment on day 7 PI. This is an enlarged illustration of Figure 3B which provides information on the specific function of each enriched IPA annotation.

Additional file 9: The subnetwork of the human PPI which contains 70 proteins whose transcripts were significantly expressed in CA04-infected tissue. This network was identified using IPA. Up-regulated genes are colored red while down-regulated genes are colored green.

Additional file 10: The "Role of Cytokines in Mediating Communication between Immune Cells" Pathway from the IPA database. Proteins colored red were identified as upregulated in CA04-infected tissue. Interactions which promote protein production or cell proliferation are illustrates with arrows. Inhibitory interactions are illustrated with ⊥.

Additional file 11: The "Role of CHK Proteins in Cell Cycle" Pathway from the IPA database. Proteins colored red were identified as upregulated in CA04-infected tissue. Interactions which promote protein production or cell proliferation are illustrates with arrows. Inhibitory interactions are illustrated with ⊥. Interactions which promote a particular phenotype (e.g., G2/S arrest) are illustrated with lines ending in a circle.

Additional file 12: Work Book Explanation: enriched canonical biological pathways for genes differentially expressed between CA04 and KUTK4 infections.

Additional file 13: Work Book Explanation: official names, symbols and accession numbers of all proteins shown in Figure 5.

Competing interests
No competing interests to declare.

Authors' contributions
JES performed the enrichment, network and statistical analyses and drafted the manuscript. SF and AJE revised the manuscript and interpreted enrichment results. YM performed the RNA isolation and microarray development. SW and TW reviewed sample quality and participated in the project design. YM revised the manuscript and participated in the network development. HK and YK designed the project and revised the manuscript. All authors have read and approved the final manuscript.

Acknowledgements
The authors would like to thank Jeffrey Chang for useful discussion and additional analysis he provided with the GATHER analytical tool and Susan Watson for editing the manuscript.

This work was supported by grants-in-aid for Specially Promoted Research and for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, by ERATO (Japan Science and Technology Agency), by a Contract Research Fund for the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases and by Public Health Service research grants from the National Institute of Allergy and Infectious Diseases.

Author details
[1]ERATO Infection-Induced Host Responses Project, Saitama 332-0012, Japan. [2]School of Veterinary Medicine, Department of Pathobiological Sciences, Influenza Research Institute, University of Wisconsin-Madison, Madison, WI, USA. [3]Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [4]The Systems Biology Institute, Tokyo, Japan. [5]Division of Systems Biology, Cancer Institute, Tokyo, Japan. [6]Sony Computer Science Laboratories, Inc, Tokyo, Japan. [7]Okinawa Institute of Science and Technology, Okinawa, Japan. [8]Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [9]International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan.

Received: 28 February 2012 Accepted: 18 August 2012
Published: 31 August 2012

References
1. CDC: **Outbreak of swine-origin influenza A (H1N1) virus infection - Mexico, March-April 2009.** *MMWR Morb Mortal Wkly Rep* 2009, **58**(17):467–470.
2. World Health Organization (WHO): *Disease Outbreak News: Pandemic (H1N1) 2009 - update 110.* Geneva, Switzerland: World Health Organization (WHO); 2009 [http://www.who.int/csr/don/2010_07_23a/en/index.html]
3. Taubenberger JK, Morens DM: **The pathology of influenza virus infections.** *Annu Rev Pathol* 2008, **3**:499–522.
4. Rowe T, Leon AJ, Crevar CJ, Carter DM, Xu L, Ran L, Fang Y, Cameron CM, Cameron MJ, Banner D, *et al*: **Modeling host responses in ferrets during A/California/07/2009 influenza infection.** *Virology* 2010, **401**(2):257–265.
5. Trilla A, Trilla G, Daer C: **The 1918 "Spanish flu" in Spain.** *Clin Infect Dis* 2008, **47**(5):668–673.
6. Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, Garten RJ, Gubareva LV, Xu X, Bridges CB, Uyeki TM: **Emergence of a novel swine-origin influenza A (H1N1) virus in humans.** *N Engl J Med* 2009, **360**(25):2605–2615.
7. Safronetz D, Rockx B, Feldmann F, Belisle SE, Palermo RE, Brining D, Gardner D, Proll SC, Marzi A, Tsuda Y, *et al*: **Pandemic swine-origin H1N1 influenza A virus isolates show heterogeneous virulence in macaques.** *J Virol* 2011, **85**(3):1214–1223.
8. Perez-Padilla R, de la Rosa-Zamboni D, Ponce de Leon S, Hernandez M, Quinones-Falconi F, Bautista E, Ramirez-Venegas A, Rojas-Serrano J, Ormsby CE, Corrales A, *et al*: **Pneumonia and respiratory failure from swine-origin influenza A (H1N1) in Mexico.** *N Engl J Med* 2009, **361**(7):680–689.
9. Dharan NJ, Gubareva LV, Meyer JJ, Okomo-Adhiambo M, McClinton RC, Marshall SA, St George K, Epperson S, Brammer L, Klimov AI, *et al*: **Infections with oseltamivir-resistant influenza A(H1N1) virus in the United States.** *JAMA* 2009, **301**(10):1034–1041.
10. Bautista E, Chotpitayasunondh T, Gao Z, Harper SA, Shaw M, Uyeki TM, Zaki SR, Hayden FG, Hui DS, Kettner JD, *et al*: **Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection.** *N Engl J Med* 2010, **362**(18):1708–1719.
11. Itoh Y, Shinya K, Kiso M, Watanabe T, Sakoda Y, Hatta M, Muramoto Y, Tamura D, Sakai-Tagawa Y, Noda T, *et al*: **In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses.** *Nature* 2009, **460**(7258):1021–1025.
12. Baskin CR, Bielefeldt-Ohmann H, Tumpey TM, Sabourin PJ, Long JP, Garcia-Sastre A, Tolnay AE, Albrecht R, Pyles JA, Olson PH, *et al*: **Early and sustained innate immune response defines pathology and death in nonhuman primates infected by highly pathogenic influenza virus.** *Proc Natl Acad Sci U S A* 2009, **106**(9):3455–3460.
13. Kobasa D, Jones SM, Shinya K, Kash JC, Copps J, Ebihara H, Hatta Y, Kim JH, Halfmann P, Hatta M, *et al*: **Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus.** *Nature* 2007, **445**(7125):319–323.

14. Cilloniz C, Shinya K, Peng X, Korth MJ, Proll SC, Aicher LD, Carter VS, Chang JH, Kobasa D, Feldmann F, *et al*: **Lethal influenza virus infection in macaques is associated with early dysregulation of inflammatory related genes.** *PLoS Pathog* 2009, **5**(10):e1000604.
15. da Huang W, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA: **The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists.** *Genome Biol* 2007, **8**(9):R183.
16. Cilloniz C, Pantin-Jackwood MJ, Ni C, Goodman AG, Peng X, Proll SC, Carter VS, Rosenzweig ER, Szretter KJ, Katz JM, *et al*: **Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model of infection.** *J Virol* 2010, **84**(15):7613–7624.
17. Goodman AG, Zeng H, Proll SC, Peng X, Cilloniz C, Carter VS, Korth MJ, Tumpey TM, Katze MG: **The alpha/beta interferon receptor provides protection against influenza virus replication but is dispensable for inflammatory response signaling.** *J Virol* 2010, **84**(4):2027–2037.
18. Viemann D, Schmolke M, Lueken A, Boergeling Y, Friesenhagen J, Wittkowski H, Ludwig S, Roth J: **H5N1 virus activates signaling pathways in human endothelial cells resulting in a specific imbalanced inflammatory response.** *J Immunol* 2011, **186**(1):164–173.
19. Proost P, Menten P, Struyf S, Schutyser E, De Meester I, Van Damme J: **Cleavage by CD26/dipeptidyl peptidase IV converts the chemokine LD78beta into a most efficient monocyte attractant and CCR1 agonist.** *Blood* 2000, **96**(5):1674–1680.
20. Deshmane SL, Kremlev S, Amini S, Sawaya BE: **Monocyte chemoattractant protein-1 (MCP-1): an overview.** *J Interferon Cytokine Res* 2009, **29**(6):313–326.
21. Allen IC, Scull MA, Moore CB, Holl EK, McElvania-TeKippe E, Taxman DJ, Guthrie EH, Pickles RJ, Ting JP: **The NLRP3 inflammasome mediates in vivo innate immunity to influenza A virus through recognition of viral RNA.** *Immunity* 2009, **30**(4):556–565.
22. Thomas PG, Dash P, Aldridge JR Jr, Ellebedy AH, Reynolds C, Funk AJ, Martin WJ, Lamkanfi M, Webby RJ, Boyd KL, *et al*: **The intracellular sensor NLRP3 mediates key innate and healing responses to influenza A virus via the regulation of caspase-1.** *Immunity* 2009, **30**(4):566–575.
23. Pirhonen J, Sareneva T, Kurimoto M, Julkunen I, Matikainen S: **Virus infection activates IL-1 beta and IL-18 production in human macrophages by a caspase-1-dependent pathway.** *J Immunol* 1999, **162**(12):7322–7329.
24. Hayes JD, Flanagan JU, Jowsey IR: **Glutathione transferases.** *Annu Rev Pharmacol Toxicol* 2005, **45**:51–88.
25. Zhang W, Li H, Cheng G, Hu S, Li Z, Bi D: **Avian influenza virus infection induces differential expression of genes in chicken kidney.** *Res Vet Sci* 2008, **84**(3):374–381.
26. Imai Y, Kuba K, Neely GG, Yaghubian-Malhami R, Perkmann T, van Loo G, Ermolaeva M, Veldhuizen R, Leung YH, Wang H, *et al*: **Identification of oxidative stress and Toll-like receptor 4 signaling as a key pathway of acute lung injury.** *Cell* 2008, **133**(2):235–249.
27. Chang JT, Nevins JR: **GATHER: a systems approach to interpreting genomic signatures.** *Bioinformatics* 2006, **22**(23):2926–2933.
28. Zhang L, Zhao Y: **The regulation of Foxp3 expression in regulatory CD4 (+)CD25(+)T cells: multiple pathways on the road.** *J Cell Physiol* 2007, **211**(3):590–597.
29. Armendariz AD, Krauss RM: **Hepatic nuclear factor 1-alpha: inflammation, genetics, and atherosclerosis.** *Curr Opin Lipidol* 2009, **20**(2):106–111.
30. Gatta R, Dolfini D, Mantovani R: **NF-Y joins E2Fs, p53 and other stress transcription factors at the apoptosis table.** *Cell Death Dis* 2011, **2**:e162.
31. Gurtner A, Fuschi P, Martelli F, Manni I, Artuso S, Simonte G, Ambrosino V, Antonini A, Folgiero V, Falcioni R, *et al*: **Transcription factor NF-Y induces apoptosis in cells expressing wild-type p53 through E2F1 upregulation and p53 activation.** *Cancer Res* 2010, **70**(23):9711–9720.
32. Hagman J, Wheat W, Fitzsimmons D, Hodsdon W, Negri J, Dizon F: **Pax-5/BSAP: regulator of specific gene expression and differentiation in B lymphocytes.** *Curr Top Microbiol Immunol* 2000, **245**(1):169–194.
33. Mukhopadhyay S, Mukherjee S, Ray BK, Ray A, Stone WL, Das SK: **Antioxidant liposomes protect against CEES-induced lung injury by decreasing SAF-1/MAZ-mediated inflammation in the guinea pig lung.** *J Biochem Mol Toxicol* 2010, **24**(3):187–194.
34. Jeong H, Mason SP, Barabasi AL, Oltvai ZN: **Lethality and centrality in protein networks.** *Nature* 2001, **411**(6833):41–42.
35. Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, Raval A: **Identifying hubs in protein interaction networks.** *PLoS One* 2009, **4**(4):e5344.
36. Fox AD, Hescott BJ, Blumer AC, Slonim DK: **Connectedness of PPI network neighborhoods identifies regulatory hub proteins.** *Bioinformatics* 2011, **27**(8):1135–1142.
37. Felsher DW, Zetterberg A, Zhu J, Tlsty T, Bishop JM: **Overexpression of MYC causes p53-dependent G2 arrest of normal fibroblasts.** *Proc Natl Acad Sci U S A* 2000, **97**(19):10544–10548.
38. Ray A, Shakya A, Kumar D, Ray BK: **Overexpression of serum amyloid A-activating factor 1 inhibits cell proliferation by the induction of cyclin-dependent protein kinase inhibitor p21WAF-1/Cip-1/Sdi-1 expression.** *J Immunol* 2004, **172**(8):5006–5015.
39. Ray A, Yu GY, Ray BK: **Cytokine-responsive induction of SAF-1 activity is mediated by a mitogen-activated protein kinase signaling pathway.** *Mol Cell Biol* 2002, **22**(4):1027–1035.
40. Sun L, Liu S, Chen ZJ: **SnapShot: pathways of antiviral innate immunity.** *Cell* 2010, **140**(3):436. e432.
41. Johanns TM, Ertelt JM, Rowe JH, Way SS: **Regulatory T cell suppressive potency dictates the balance between bacterial proliferation and clearance during persistent Salmonella infection.** *PLoS Pathog* 2010, **6**(8):e1001043.
42. Nguyen VH: **Balancing act for Treg immunotherapy.** *Blood* 2011, **117**(10):2751–2752.
43. He Y, Xu K, Keiner B, Zhou J, Czudai V, Li T, Chen Z, Liu J, Klenk HD, Shu YL, *et al*: **Influenza A virus replication induces cell cycle arrest in G0/G1 phase.** *J Virol* 2010, **84**(24):12832–12840.
44. Pietenpol JA, Stewart ZA: **Cell cycle checkpoint signaling: cell cycle arrest versus apoptosis.** *Toxicology* 2002, **181–182**:475–481.
45. Ugai H, Li HO, Komatsu M, Tsutsui H, Song J, Shiga T, Fearon E, Murata T, Yokoyama KK: **Interaction of Myc-associated zinc finger protein with DCC, the product of a tumor-suppressor gene, during the neural differentiation of P19 EC cells.** *Biochem Biophys Res Commun* 2001, **286**(5):1087–1097.
46. Ray BK, Murphy R, Ray P, Ray A: **SAF-2, a splice variant of SAF-1, acts as a negative regulator of transcription.** *J Biol Chem* 2002, **277**(48):46822–46830.
47. Nonomura Y, Kohsaka H, Nagasaka K, Miyasaka N: **Gene transfer of a cell cycle modulator exerts anti-inflammatory effects in the treatment of arthritis.** *J Immunol* 2003, **171**(9):4913–4919.
48. Kuiken T, Rimmelzwaan GF, Van Amerongen G, Osterhaus AD: **Pathology of human influenza A (H5N1) virus infection in cynomolgus macaques (Macaca fascicularis).** *Vet Pathol* 2003, **40**(3):304–310.
49. da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44–57.

doi:10.1186/1752-0509-6-117
Cite this article as: Shoemaker *et al.*: Integrated network analysis reveals a novel role for the cell cycle in 2009 pandemic influenza virus-induced inflammation in macaque lungs. *BMC Systems Biology* 2012 **6**:117.

BMC Genomics

SOFTWARE Open Access

CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data

Jason E Shoemaker[1*], Tiago JS Lopes[1], Samik Ghosh[2], Yukiko Matsuoka[1,2], Yoshihiro Kawaoka[1,3,4] and Hiroaki Kitano[1,2,5,6]

Abstract

Background: Interpreting *in vivo* sampled microarray data is often complicated by changes in the cell population demographics. To put gene expression into its proper biological context, it is necessary to distinguish differential gene transcription from artificial gene expression induced by changes in the cellular demographics.

Results: CTen (cell type enrichment) is a web-based analytical tool which uses our highly expressed, cell specific (HECS) gene database to identify enriched cell types in heterogeneous microarray data. The web interface is designed for differential expression and gene clustering studies, and the enrichment results are presented as heatmaps or downloadable text files.

Conclusions: In this work, we use an independent, cell-specific gene expression data set to assess CTen's performance in accurately identifying the appropriate cell type and provide insight into the suggested level of enrichment to optimally minimize the number of false discoveries. We show that CTen, when applied to microarray data developed from infected lung tissue, can correctly identify the cell signatures of key lymphocytes in a highly heterogeneous environment and compare its performance to another popular bioinformatics tool. Furthermore, we discuss the strong implications cell type enrichment has in the design of effective microarray workflow strategies and show that, by combining CTen with gene expression clustering, we may be able to determine the relative changes in the number of key cell types.
CTen is available at http://www.influenza-x.org/~jshoemaker/cten/

Keywords: Cell type enrichment, Microarray data, Deconvolution, Influenza, Systems immunology

Background

Microarray studies quantify genome wide changes in gene expression and have a variety of applications - from tracing allele ancestry as species evolve [1] to the development of genome-based personalized medicine [2]. A major challenge in the microarray analysis of tissue collected *in vivo* is that often the perceived gene regulation is the result of changes in the populations of particular cell types as opposed to an actual change in transcriptional activity (see Figure 1). Particularly in situations which invoke the immune response, as the cell count of various lymphocytes change within the tissue, they bring with them their own unique quantities of RNA [3]. This leads to large changes in the copy number of RNA transcripts and can lead to the false perception of increased transcriptional activity.

Several bioinformatics tools exist to identify the cause and effect of changes in gene regulation, with gene set enrichment analysis (GSEA) [4] and gene ontology (GO) enrichment [5] being the most popular, and there are several other web-based platforms with improvements or variations of these analytical tools [6-8]. GSEA relies on a database of reference gene lists which were previously determined to be regulated under several conditions (e.g., by transcription factors, chemical and genetic perturbations, or between healthy and diseased states).

* Correspondence: jshoe@ims.u-tokyo.ac.jp
[1]JST ERATO KAWAOKA Infection-induced Host Responses Project, Tokyo, Japan
Full list of author information is available at the end of the article

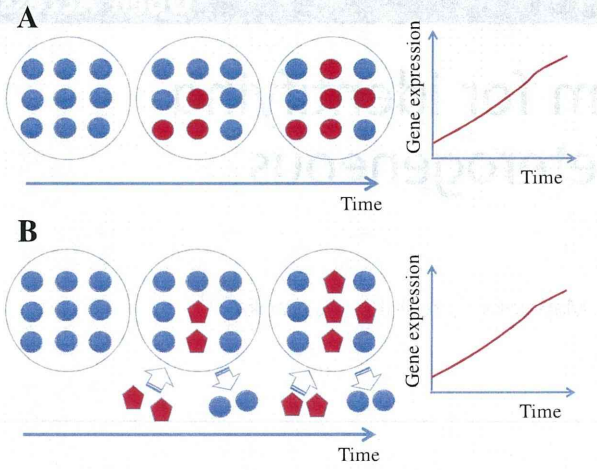

Figure 1 Changes in cell demographics can result in gene expression. Two scenarios which result in similar gene expression changes: (**A**) The cell type(s) within the sample are unchanged, but, over time, inactivated cells (colored blue) become activated and express a marker gene (colored red); (**B**) A second cell type already actively expressing the marker gene (red colored pentagons) migrates into the sample. The change in the marker gene expression is similar in both cases but results from a different reason.

GSEA determines which reference list - if any - has statistically significant, concordant regulation. Although very useful for linking gene expression to specific transcription factors or identifying similarities between diseases, this tool does not include cell specific data at this time. The other popular alternative, GO, relies on a controlled vocabulary to describe the biological role of genes and their products. It is often accurate in predicting the local phenotype from gene expression data (e.g., inflammation annotations are highly enriched in samples from inflamed tissue [9]). However, cell specific GO annotations are often overwhelmed by more ubiquitous terms in the GO annotation hierarchy.

Additionally, some algorithms exist to unmix cellularly heterogeneous gene expression data into expression profiles for each cell type [10,11] but generally either the number of cell types must be known *a priori* or cell counts must be determined. The Gene Expression Barcode [12] and BioGPS [13] web platforms provide tissue specific gene expression data and allow researchers to compare gene expression between different tissues in their databases. However, these tools do not provide a means to relate user-generated sets of differentially expressed genes to specific cell types. Hence, to facilitate the proper interpretation of genomic regulation from *in vivo* microarray data, we developed CTen to determine if the observed gene regulation is the result of changes in the cellular make up of the sample.

Two principles guided the development of our highly expressed cell-specific (HECS) gene database and the CTen website's interface. First, basal gene expression levels strongly differ between cell types [3,14]. By analyzing gene expression across several cell types and tissues, we can select genes with very high expression in a limited number of cell types. In turn, each cell type has a collection of HECS genes to act as a cell-specific signature. Thus for any user generated list of genes, we can determine if the number of HECS genes for a particular cell type is greater than the number expected by chance.

The second principle, which led us to optimize CTen's interface for gene expression clustering studies, is the observation that changes in messenger RNA levels due to cell migration or variances in sample collection techniques result in conserved expression patterns in microarray data. Several clustering strategies, including hierarchical clustering and the weighted gene corregulation network algorithm (WGCNA) [15], have been developed to identify gene expression patterns which are conserved temporally or across experimental groups. By combining clustering with cell type enrichment, CTen can address a major challenge in biology today; namely separating gene expression from cellularly heterogeneous RNA samples into clusters representing differential transcriptional activity and clusters representing changes in gene expression due to cell migration.

Here, we first describe the construction of the HECS database and discuss the workflow behind the CTen website's design. We then validate CTen's ability to correctly identify the appropriate cellular signature and evaluate the benefits of users requiring increasingly strict enrichment scores. We motivate the use of CTen using genes differentially expressed in the lungs of mice infected with influenza virus, and, lastly, provide an illustrative example promoting the use of CTen for detecting changes in the cellular demographics and the critical role this plays in functional enrichment and gene network inference studies.

Implementation

The HECS database construction

We downloaded from BioGPS [13] gene expression data from 96 mouse and 84 human tissues/cell types (Mouse MOE430 Gene Atlas and Human U133A/GNF1H Gene Atlas; a complete list of all cell types used is available in the Additional file 1). The expression values were averaged over the biological replicates (2 per cell type) and, for each cell type, a transcript was identified as a HECS gene if one of its corresponding probes had an expression value (averaged over the 2 replicates) at least 15x or 10x greater than the median expression value of the probe for all cell types in the mouse and human datasets, respectively. Next, probe identifiers were matched to their Entrez Gene IDs and official gene symbols using the Affymetrix Mouse Genome 430 2.0

Array (mouse4302 version 2.5.0) and Affymetrix Human Genome U133 Set (hgu133a version 2.5.0) annotation files available from Bioconductor [16]. The final step was to remove redundant Entrez Gene IDs assigned as HECS genes to the same cell type (due to multiple probes mapping to the same gene). The CTen database is available for download under the "Database Info" tab on the CTen website.

Threshold selection

Importantly, as stated above, preset cutoffs were used in developing the mouse and human HECS databases. These cutoffs (15x and 10x the median expression level for a probe across all cell types) were selected to balance the quantity of genes with the uniqueness of the genes assigned to each cell type. Uniqueness was quantified by determining the percentage of genes identified as a HECS gene for *n* or fewer cell types. As seen in Figure 2A-B, raising the cutoff caused a sharp reduction in the number of genes but significantly improved the uniqueness (Figure 2C-D) of the genes assigned as HECS genes to each cell type. Increasing the cutoff for the mouse data beyond 15x did not significantly improve uniqueness and only served to limit the number of HECS genes per cell type to act as cell signatures. For the cutoffs considered for the human data, a cutoff of 15x slightly improves the uniqueness but the number of HECS genes per cell type became prohibitively small. Thus, the HECS expression threshold requirement was reduced to 10x the median expression value in the human dataset to ensure that all cell types are represented.

At the cutoff values selected (emphasized in Figure 2A-B in blue (mouse) and orange (human)), even when applying a more stringent expression requirement, the number of HECS genes per cell type remains significantly higher in the mouse data (average number of HECS gene per cell = 794 in mouse and only 351 for human derived cell types). In terms of uniqueness of the HECS genes (emphasized in Figure 2C-D in blue (mouse) and orange (human)), we find that 55.8% of human HECS genes are exclusive to 3 or fewer cell types, while 53.3% of mouse HECS genes are limited to 4 or fewer cell types.

We emphasize that for both the mouse and human HECS databases, for values greater than 10x the median gene expression, the number of HECS genes per cell type and the identity of the HECS genes do not change significantly. Thus, cutoff selection within the ranges considered should not strongly bias any results from enrichment analysis. We validated this by showing that CTen's performance was independent of the precise threshold selected. We reconstructed the HECS databases

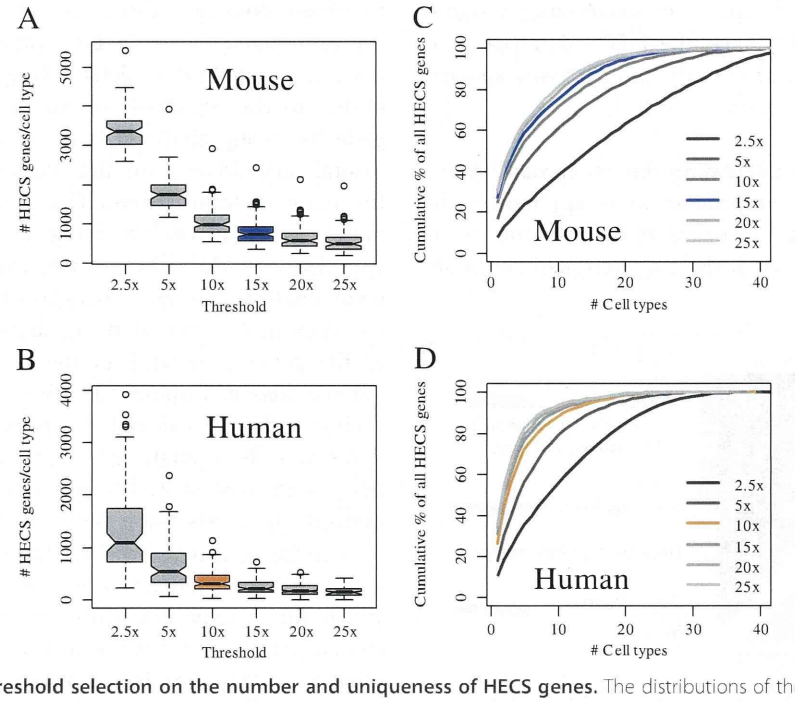

Figure 2 The effect of threshold selection on the number and uniqueness of HECS genes. The distributions of the number of HECS genes per cell type as the threshold criteria used to define a HECS gene is raised from 2x to 25x the median expression value across all cell types for the (**A**) mouse and (**B**) human gene expression data. To quantify uniqueness, we determined the percentage of HECS genes that were mapped to *n* or fewer cell types (i.e., the cumulative %) for the (**C**) mouse and (**D**) human gene expression data for different threshold values. The results corresponding to the threshold values selected in the current implementation of CTen are colored blue and orange for the mouse and human data, respectively.