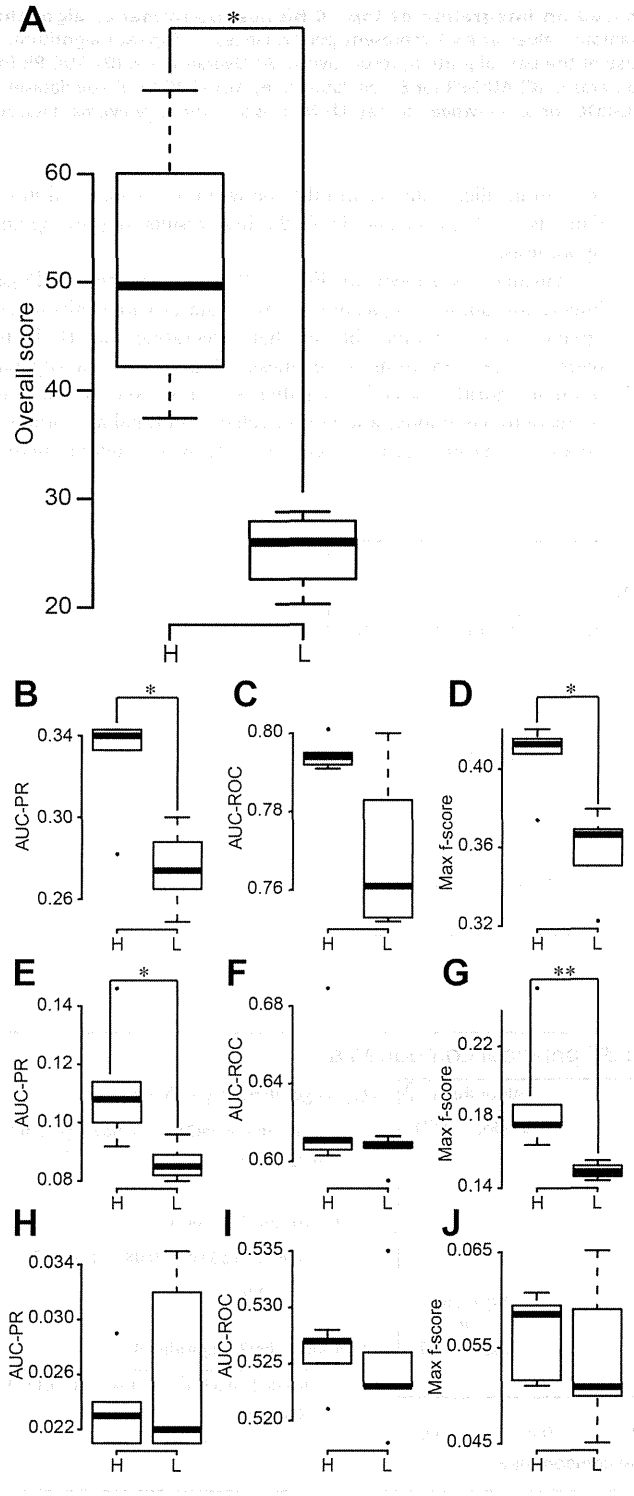


**Figure 5. Performances of Top1Net based on integration of high- or low-diversity algorithm pairs by EUC distance.** H and L represent high-diversity and low-diversity algorithm pairs, respectively. (**A**) Box-plots of overall score. (**B**) Box-plots of AUC-PR for in silico dataset. (**C**) Box-plots of AUC-ROC for in silico dataset. (**D**) Box-plots of Max f-score for in silico dataset. (**E**) Box-plots of AUC-PR for *E. coli* dataset. (**F**) Box-plots of AUC-ROC for *E. coli* dataset. (**G**) Box-plots of Max f-score for *E. coli* dataset. (**H**) Box-plots of AUC-PR for *S. cerevisiae* dataset. (**I**) Box-plots of AUC-ROC for *S. cerevisiae* dataset. (**J**) Box-plots of max f-score for *S. cerevisiae* dataset. * and ** represent $P<0.05$ and $P<0.01$, by the Wilcoxon rank sum test.
doi:10.1371/journal.pcbi.1003361.g005

## Discussion

With an increasing corpus of inference algorithms, leveraging their diverse and sometimes complementary approaches in a community consensus can be a promising strategy for reconstruction of gene regulatory networks from large scale experimental data. A computational platform to systematically analyze, assess and leverage these diverse techniques is essential for the successful application of reverse engineering in biomedical research.

This study presents a reverse engineering framework which can flexibly integrate multiple inference algorithms, based on **Top*k*Net** - a novel technique for building a consensus network based on the algorithms. It is pertinent to note here that the consensus framework based on **Top*k*Net** can be flexibly extended to include various types of network-inference algorithms.

Comparative evaluation on the DREAM5 datasets showed that, although Top*k*Net based on 38-algorithm integration shows lower or at most comparable performance to the best individual algorithms, Top1Net based on integration of top 10 highest performance algorithms significantly outperforms the best individual algorithm as well as community prediction. The results demonstrated that (i) a simple strategy to combine many algorithms does not always lead to performance improvement compared to the cost of consensus and (ii) selection of high-performance algorithms for a given expression dataset and Top1Net based on integration of the selected high-performance algorithms could be a powerful strategy for reliable reverse engineering.

Why does Top1net algorithm integrating 10 optimal algorithms perform quite well and outperform the best individual method? This is because 10 optimal algorithms tend to assign high-confidence scores to true-positive links and Top1net method can recover many true-positive links that are with the highest confidence scores from 10 optimal algorithms. Furthermore, 10 optimal algorithms are based on different techniques (*e.g.*, mutual information, regression, and other statistical techniques) and Top1net can leverage diversity from the optimal algorithms. For example, the optimal algorithms based on mutual-information and regression techniques can accurately recover true positive links in feed-forward loops and linear cascade modules, respectively [35], while Top1net could integrate the algorithms and accurately recover both feed-forward loops and linear cascade module in a GRN. Therefore, Top1net shows higher inference performance than the best individual algorithms.

Why, then, Top1net outperforms community prediction and Topknet with higher $k$? Community prediction and Topknet with larger $k$ recover links with lower confidence scores than Top1net, *i.e.*, community prediction uses mean among confidence scores from 10 optimal algorithms and Topknet uses $k$th highest confidence score from the algorithms. Links with lower confidence scores from optimal algorithms are more likely to be false-positive links and thus Top1net shows higher inference performance than community prediction and Topknet with higher $k$.

A key to reconstruct accurate GRNs is development of a method to determine optimal algorithms for a given expression dataset associated with unknown regulatory network. As mentioned in results, if similarity between expression-data associated with known regulatory network (*i.e.*, DREAM5 datasets) and that with an unknown regulatory network is high, optimal algorithms for data with known regulatory network may be also optimal for reconstruction of the unknown regulatory network.

Based on this observation, we developed a measure to quantify similarity among the expression datasets based on algorithm diversity and demonstrated that, if similarity between the two

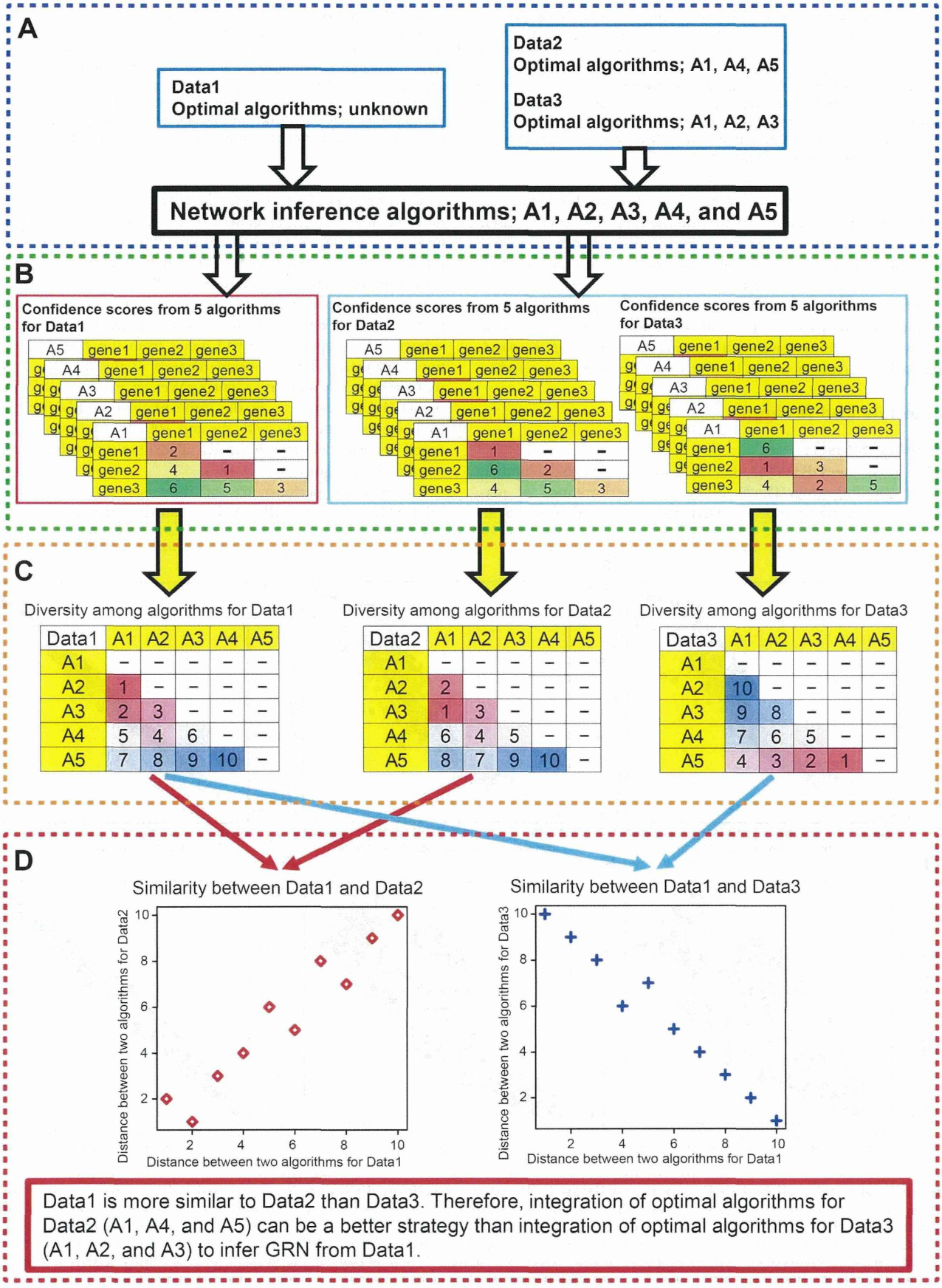
A
Data1
Optimal algorithms; unknown
Data2
Optimal algorithms; A1, A4, A5
Data3
Optimal algorithms; A1, A2, A3
Network inference algorithms; A1, A2, A3, A4, and A5
B
Confidence scores from 5 algorithms for Data1
Confidence scores from 5 algorithms for Data2
Confidence scores from 5 algorithms for Data3
C
Diversity among algorithms for Data1
Diversity among algorithms for Data2
Diversity among algorithms for Data3
D
Similarity between Data1 and Data2
Distance between two algorithms for Data2
Distance between two algorithms for Data1
Similarity between Data1 and Data3
Distance between two algorithms for Data3
Distance between two algorithms for Data1
Data1 is more similar to Data2 than Data3. Therefore, integration of optimal algorithms for Data2 (A1, A4, and A5) can be a better strategy than integration of optimal algorithms for Data3 (A1, A2, and A3) to infer GRN from Data1.

**Figure 6. Overview of a method to calculate similarity between two expression datasets.** (**A**) Datasets. Expression datasets were split into a dataset for which optimal algorithms are unknown (e.g., Data1) and datasets for which optimal algorithms are known (e.g., Data2 and Data3). (**B**) Confidence scores of links between two genes. For each of datasets, confidence scores from each of algorithms (e.g., algorithms, A1, A2, A3, A4, and A5) were calculated. (**C**) Diversity among algorithms. By using confidence scores calculated in (B), diversity among algorithms were calculated for each of three datasets. In this example, we examined five algorithms and thus, for each of the datasets, we have a vector of 10 distances between two algorithms. (**D**) Similarity between two expression datasets. Correlation coefficient between the vector of algorithm distances from Data1 and that from Data2 was calculated. The calculated correlation coefficient is defined as similarity between Data1 and Data2. In the example in this figure, Data1 is more similar to Data2 than Data3. Thus, optimal algorithms for Data2 could perform better than those for Data3 to infer GRN from Data1.
doi:10.1371/journal.pcbi.1003361.g006

expression-datasets is high, integration of algorithms that are optimal for one dataset could perform well on the other dataset. Thus, the similarity measure proposed in this study can be a good clue to identify optimal algorithms for reliable reconstruction of an unknown regulatory network.

The consensus framework outlined in this paper, Top*k*Net, together with analysis of similarity among expression datasets, provide a powerful platform towards harnessing the *wisdom of the crowds* approach in reconstruction of large scale gene regulatory networks.

## Materials and Methods

### DREAM5 datasets

We used the DREAM5 datasets (http://wiki.c2b2.columbia.edu/dream/index.php/D5c4) to evaluate performance of network-inference algorithms. The DREAM5 dataset composed of an in-silico network (1,643 genes), the real transcriptional regulatory network of *E. coli* (4,511 genes), that of *S. celecisiae* (5,950 genes), and corresponding expression dataset (805, 805, and 536 samples for the in-silico, *E. coli*, and *S. celevisiae* networks, respectively). The expression dataset of *E. coli* and that of *S. celevisae* are composed of hundreds of experiments, *i.e.*, genetic, drug, and environmental perturbations. The in-silico network is generated by extracting a subnetwork composed of 1,643 genes from the *E. coli* transcriptional network. The expression datasets of the in-silico network was simulated by software GeneNetWeaver version 2.0 [38]. For the DREAM5 datasets, in the same manner to Marbach et al. [35], we used the links with the top 100,000 highest confidence scores by each network-inference algorithm to evaluate performance of the algorithm.

To evaluate performance of inference algorithms for the DREAM5 datasets, DREAM organizers provide a matlab

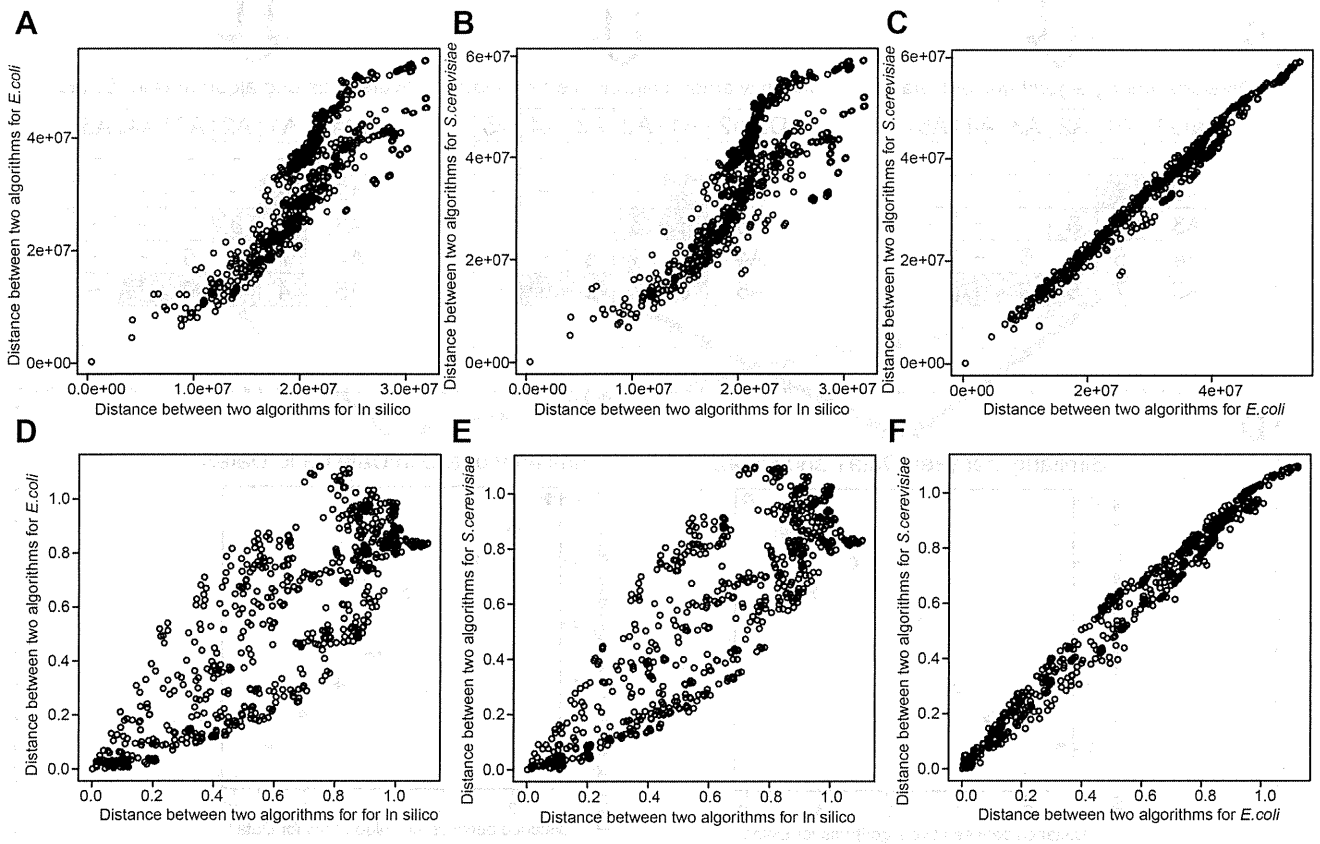


**Figure 7. Similarity among gene-expression datasets based on algorithm diversity.** The scatter plots show correlation of algorithm distance between two gene-expression datasets. Each of points in scatter plots represents each of algorithm pairs. Because we have 703 algorithm pairs among 38 algorithms, 703 points are in each of the figures. Vertical axis represents (EUC or PCA) distance between two algorithms for one gene-expression dataset, while horizontal axis represents that for the other gene-expression dataset. (**A**) Scatter plots of EUC distance for in silico and *E. coli* datasets. (**B**) Scatter plots of EUC distance for in silico and *S. cerevisiae* dataset. (**C**) Scatter plots of EUC distance for *E. coli* and *S. cerevisiae* datasets (**D**) Scatter plots of PCA distance for in silico and *E. coli* datasets. (**E**) Scatter plots of PCA distance for in silico and *S. cerevisiae* datasets. (**F**) Scatter plots of PCA distance for *E. coli* and *S. cerevisiae* datasets.
doi:10.1371/journal.pcbi.1003361.g007

**Table 2.** Correlation coefficient of algorithm distances and that of performance metrics across the DREAM5 gene-expression datasets.

| Dataset 1 | Dataset 2 | EUC distance[1] | PCA distance[2] |
|---|---|---|---|
| In silico[3] | *E.coli*[4] | 0.87 | 0.81 |
| In silico | *S.cerevisiae*[5] | 0.83 | 0.83 |
| *E.coli* | *S.cerevisiae* | 0.99 | 0.99 |

[1]Spearman's correlation coefficient of algorithm distance (EUC distance) between Dataset 1 and Dataset 2.
[2]Spearman's correlation coefficient of algorithm distance (PCA distance) between Dataset 1 and Dataset 2.
[3]In silico Dream 5 dataset.
[4]Dream 5 dataset from *E.coli*.
[5]Dream5 dataset from *S.cerevisiae*.
doi:10.1371/journal.pcbi.1003361.t002

software (http://wiki.c2b2.columbia.edu/dream/index.php/D5c4). The software calculates 4 metrics for each network, *i.e.*, AUC-PR, AUC-ROC, AUC-PR p-value, and AUC-ROC p-value. AUC-PR (AUC-ROC) p-value is the probability that a given or greater AUC-PR (AUC-ROC) is obtained by random scoring of links. Furthermore, the software calculates an overall score that was used to evaluate the overall performance of the algorithms for all three networks (the large synthetic network, large real *E. coli*, and *S. celevisiae* GRNs) of the DREAM5 network inference challenge. The overall score (OS) is defines as $OS = 0.5(p_1+p_2)$, where $p_1$ and $p_2$ are the mean of the log-transformed AUC-PR p-values and that of the log-transformed AUC-ROC p-values taken over the three networks of the DREAM5 challenge, respectively.

## Confidence score of regulatory links from 38 network-inference algorithms

We obtained confidence scores between two genes by 35 algorithms (29 algorithms are from DREAM5 participants and 6 algorithms are commonly used "off-the shelf" algorithms) from supplementary file of Marbach et al. [35]. For c3net, ggm, and mrnet algorithms, we calculated confidence scores of regulatory link by using GeneNet package [39], c3net R package [9,10], and minet R package [40], respectively. Because Marbach et al. used links with top 100,000 highest confidence scores from each of 35 algorithms for analyses [35], we used top 100,000 links from c3net, ggm, and mrnet for analyses in this study.

## Metrics to evaluate performance of inference algorithms

For a given threshold value of confidence level, network-inference algorithms predict whether a pair of genes have regulatory link or not. A pair of genes with a predicted link is considered as a true positive (TP) if the link is present in the underlying synthetic network, while the pair is a false positive (FP)

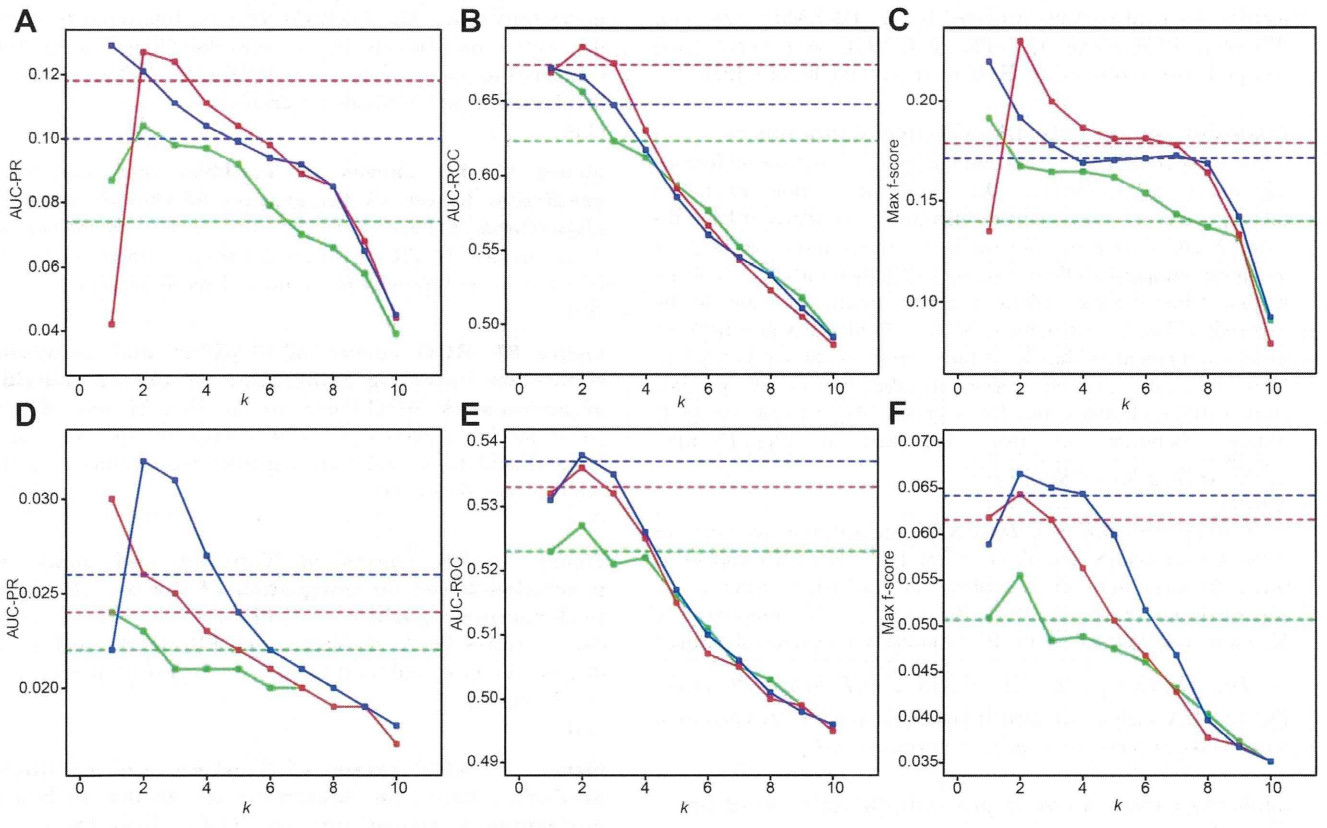


**Figure 8. Optimal algorithm selection based on similarity among expression datasets and its potential to improve network-inference accuracy.** Red lines show performance of Top*k*Net integrating algorithms that are optimal for a dataset with high-similarity, while green lines show that with low-similarity. Blue lines show performance of Top*k*Net integrating top 10 highest-performance algorithms. Dashed lines in red, green, and blue represent performance of community prediction integrating algorithms that are optimal for a dataset with high-similarity, that with low-similarity, and top 10 highest-performance algorithms, respectively. (**A**) AUC-PR for *E. coli* dataset. (**B**) AUC-ROC for *E. coli* dataset. (**C**) Max f-score for *E. coli* dataset. (**D**) AUC-PR for *S. cerevisiae* dataset. (**E**) AUC-ROC for *S. cerevisiae* dataset. (**F**) Max f-score for *S. cerevisiae* dataset.
doi:10.1371/journal.pcbi.1003361.g008

if the synthetic network does not have the link. Similarly, a pair of genes without a predicted link is considered as a true negative (TN) or false negative (FN) depending on whether the link exists or not in the underlying synthetic network, respectively. By using the values of TP, FP, TN, and FN, we can calculate several metrics to evaluate performances of network-inference algorithms.

One representative metric is precision/recall curve where the precision ($p$) and recall ($r$) are defined as $p=\frac{TP}{TP+FP}$ and $r=\frac{TP}{TP+FN}$, respectively. By using many threshold values, we obtained a precision/recall curve that is a graphical plot of the precision vs. the recall and is a straight forward visual representation of performances of network-inference algorithms. The area under the precision/recall curve (AUC-PR) is a summary metric of precision/recall curve and measures the average accuracy of network-inference algorithms. Another representative metric is ROC curve that is a graphical plot of the true-positive rate vs. the false-positive rate. The area under the ROC curve (AUC-ROC) also represents the average inference performance of algorithms. On the other hand, max f-score [41] evaluates optimum performance of network-inference algorithms where f-score is defined as harmonic mean of the precision and recall (f-score$=\frac{2pr}{r+p}$). As predictions of network-inference algorithms become more accurate, the value of AUC-PR, AUC-ROC, and max f-score becomes higher. We used AUC-PR, AUC-ROC, and max f-score for performance evaluation. To obtain these three metrics, we used package provided by the DREAM5 team [35] (PR curve, ROC curve, AUC-PR, AUC-ROC, and overall score) and perl script provided by Küffner et al. (max f-score) [27].

### Distances among network-inference algorithms

By using confidence scores among genes by network-inference algorithms, we calculated, $D_{\mathrm{EUC}}(\mathrm{X},\mathrm{Y})$, the simple Euclidean distance between two network-inference algorithms (EUC distance) X and Y for expression datasets with given number of genes and given sample size. Before giving a definition for $D_{\mathrm{EUC}}(\mathrm{X},\mathrm{Y})$, let us first define some notations. Let $n$ be number of genes in the expression dataset and $CS(i, j, \mathrm{X})$ be confidence value between genes $i$ and $j$ by algorithm X on the expression dataset. $G=\{(1,2), (2,3), \ldots (i,j) \ldots (n\text{-}1,n)\}$ represents the list of all possible combinations of two genes for $n$ genes. We defined the EUC distance between the two algorithms as $D_{\mathrm{EUC}}(\mathrm{X},\mathrm{Y})=\sqrt{\sum_{(i,j)\in G}(CS(i,j,\mathrm{X})-CS(i,j,\mathrm{Y}))^2}$.

Further, we calculated, $D_{\mathrm{PCA}}(\mathrm{X},\mathrm{Y})$, the distance between two network-inference (X and Y) on 2nd and 3rd principal components (PCA distance) from PCA analysis on confidence scores of 38 algorithms. Let $PC_2(\mathrm{X})$ and $PC_3(\mathrm{X})$ be the 2nd and 3rd components of X, respectively. We defined the PCA distance between two algorithms as $D_{\mathrm{PCA}}(\mathrm{X},\mathrm{Y})=\sqrt{(PC_1(\mathrm{X})-PC_1(\mathrm{Y}))^2+(PC_2(\mathrm{X})-PC_2(\mathrm{Y}))^2}$. For the PCA analysis, we used R code, pricomp2.R, obtained from http://aoki2.si.gunma-u.ac.jp/R/src/princomp2.R.

### Similarity between two expression datasets based on algorithm diversity

By using distances among algorithms, we calculated, $S(\mathrm{da1},\mathrm{da2})$, similarity between two expression datasets da1 and da2. Before giving definition of $S(\mathrm{da1},\mathrm{da2})$, let us first define some notation. Let $k$ and $A=\{a_1, a_2, \ldots, a_i, \ldots, a_k\}$ be the number of algorithms and the list of the algorithms, respectively. $AC=\{(a_1,a_2),(a_2,a_3),\ldots, (a_{i-1},a_i),\ldots, (a_{k-1},a_k)\}$ represents all possible combinations of two algorithms among $k$ algorithms ($k(k\text{-}1)/2$ algorithm combinations). For example, in this study, we examined 38 algorithms and have $38*37/2=703$ algorithm combinations. $D(a_i,a_j,\mathrm{da1})$ represents distances between two algorithms $a_1$ and $a_2$ for da1. $\mathbf{D}_{\mathrm{da1}}\{\mathrm{AC}\}=\{D(a_1,a_2,\mathrm{da1}), D(a_2,a_3,\mathrm{da1}), \ldots, D(a_{i-1},a_i,\mathrm{da1}), \ldots D(a_{k-1},a_k,\mathrm{da1})\}$ represents a vector of $k(k\text{-}1)/2$ algorithm distances for da1 (in this study, we have a vector of 703 algorithm distances for each of DREAM5 datasets). We defined $S(\mathrm{da1},\mathrm{da2})$ as Spearman's correlation coefficient between two vectors, $\mathbf{D}_{\mathrm{da1}}\{\mathrm{AC}\}$ and $\mathbf{D}_{\mathrm{da2}}\{\mathrm{AC}\}$.

### Cloud computing infrastructure on Amazon EC2 to infer GRNs from the large-scale DREAM5 expression datasets

To infer GRNs from the large-scale expression data of DREAM5 (expression data of *E.coli* and *S. cerevisiae*), we built a cloud computing infrastructure on Amazon EC2 "High-memory double" instances (34.2 GB memory and 4 virtual cores with 3.25 EC2 Compute Units each) with Redhad linux and R version 2.15.0 [42]. We placed all the input data on the ephemeral storage disk (850 GB) of the Amazon EC2 instances and Top*k*Net output results (*e.g.*, a listing of confidence scores between genes) to files on the storage disk.

## Supporting Information

**Figure S1 The work flow of the experimental framework of this study.** Expression datasets were obtained from the DREAM5 challenge web page (http://wiki.c2b2.columbia.edu/dream/index.php/The_DREAM_Project). Inferred network from the expression datasets by a network-inference algorithm is compared to the networks of the DREAM5 challenge (Step (iii)). See Materials and Methods for details.
(TIF)

**Figure S2 PR curves of Top*k*Net and community prediction based on integration of the 38 individual algorithms.** (**A**) PR curves for in silico datasets. (**B**) PR curves for *E. coli* dataset. (**C**) PR curves for *S. cerevisiae* dataset. Vertical and horizontal axes represent precision and recall, respectively.
(TIF)

**Figure S3 ROC curves of Top*k*Net and community prediction based on integration of the 38 individual algorithms.** (**A**) ROC curves for in silico datasets. (**B**) ROC curves for *E. coli* dataset. (**C**) ROC curves for *S. cerevisiae* dataset. Vertical and horizontal axes represent true-positive and false-positive rate, respectively.
(TIF)

**Figure S4 PR curves of Top*k*Net and community prediction based on integration of the top 10 highest-performance algorithms.** (**A**) PR curves for in silico datasets. (**B**) PR curves for *E. coli* dataset. (**C**) PR curves for *S. cerevisiae* dataset. Vertical and horizontal axes represent precision and recall, respectively.
(TIF)

**Figure S5 ROC curves of Top*k*Net and community prediction based on integration of the top 10 highest-performance algorithms.** (**A**) ROC curves for in silico datasets. (**B**) ROC curves for *E. coli* dataset. (**C**) ROC curves for *S. cerevisiae* dataset. Vertical and horizontal axes represent true-positive and false-positive rate, respectively.
(TIF)

**Figure S6 Performances of community prediction based on integration of high- or low-diversity algorithm**

**pairs by EUC distance.** H and L represent high-diversity and low-diversity algorithm pairs, respectively. (**A**) Box-plots of overall score. (**B**) Box-plots of AUC-PR for in silico dataset. (**C**) Box-plots of AUC-ROC for in silico dataset. (**D**) Box-plots of Max f-score for in silico dataset. (**E**) Box-plots of AUC-PR for *E. coli* dataset. (**F**) Box-plots of AUC-ROC for *E. coli* dataset. (**G**) Box-plots of Max f-score for *E. coli* dataset. (**H**) Box-plots of AUC-PR for *S. cerevisiae* dataset. (**I**) Box-plots of AUC-ROC for *S. cerevisiae* dataset. (**J**) Box-plots of max f-score for *S. cerevisiae* dataset. * and ** represent $P<0.05$ and $P<0.01$, by the Wilcoxon rank sum test.
(TIF)

**Figure S7 Performances of Top1Net based on integration of high- or low- diversity algorithm pairs by PCA distance.** H and L represent high-diversity and low-diversity algorithm pairs, respectively. (**A**) Box-plots of overall score. (**B**) Box-plots of AUC-PR for in silico dataset. (**C**) Box-plots of AUC-ROC for in silico dataset. (**D**) Box-plots of Max f-score for in silico dataset. (**E**) Box-plots of AUC-PR for *E. coli* dataset. (**F**) Box-plots of AUC-ROC for *E. coli* dataset. (**G**) Box-plots of Max f-score for *E. coli* dataset. (**H**) Box-plots of AUC-PR for *S. cerevisiae* dataset. (**I**) Box-plots of AUC-ROC for *S. cerevisiae* dataset. (**J**) Box-plots of max f-score for *S. cerevisiae* dataset. * represents $P<0.05$ by the Wilcoxon rank sum test.
(TIF)

**Figure S8 Performances of community prediction based on integration of high- or low- diversity algorithm pairs by PCA distance.** H and L represent high-diversity and low-diversity algorithm pairs, respectively. (**A**) Box-plots of overall score. (**B**) Box-plots of AUC-PR for in silico dataset. (**C**) Box-plots of AUC-ROC for in silico dataset. (**D**) Box-plots of Max f-score for in silico dataset. (**E**) Box-plots of AUC-PR for *E. coli* dataset. (**F**) Box-plots of AUC-ROC for *E. coli* dataset. (**G**) Box-plots of Max f-score for *E. coli* dataset. (**H**) Box-plots of AUC-PR for *S. cerevisiae* dataset. (**I**) Box-plots of AUC-ROC for *S. cerevisiae* dataset. (**J**) Box-plots of max f-score for *S. cerevisiae* dataset. * and ** represent $P<0.05$ and $P<0.01$, by the Wilcoxon rank sum test.
(TIF)

**Figure S9 Comparison of algorithm performances across gene-expression datasets.** The scatter plots show correlation of algorithm performance between two gene-expression datasets. Vertical axis represents algorithm performance for one gene-expression dataset, while horizontal axis represents that for the other gene-expression dataset. (**A**) Scatter plots of AUC-PR for in silico and E. coli datasets. (**B**) Scatter plots of AUC-PR for in silico and *S. cerevisiae* datasets. (**C**) Scatter plots of AUC-PR for *E. coli* and *S. cerevisiae* datasets. (**D**) Scatter plots of AUC-ROC for in silico and *E. coli* datasets. (**E**) Scatter plots of AUC-ROC for in silico and *S. cerevisiae* datasets. (**F**) Scatter plots of AUC-ROC for *E. coli* dataset and *S. cerevisiae* datasets. (**G**) Scatter plots of max f-score for in silico and *E. coli* datasets. (**H**) Scatter plots of max f-score for in silico and *S. cerevisiae* datasets. (**I**) Scatter plots of max f-score for *E. coli* and *S. cerevisiae* datasets.
(TIF)

**Table S1 Performances of the 38 individual algorithms.** The table shows overall score, AUC-PR, and AUC-ROC of the 38 algorithms.
(XLS)

**Table S2 Correlation coefficient of performance metrics across the DREAM5 gene-expression datasets.** The table shows Spearman's correlation coefficient of performance metrics across the DREAM5 gene expression datasets.
(DOC)

## Acknowledgments

The authors thank Matsuoka Y, Kang H, Fujita K, Lopes T, and Shoemaker J for their useful comments and discussion.

## Author Contributions

Conceived and designed the experiments: TH SG HK. Performed the experiments: TH SG. Analyzed the data: TH SG. Contributed reagents/materials/analysis tools: TH SG RY HK. Wrote the paper: TH SG RY HK.

## References

1. Vidal M, Cusick ME, Barabasi AL (2011) Interactome networks and human disease. Cell 144: 986–998.
2. Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nature Review Genetics 12: 56–68.
3. Bansal M, Belcastro V, Ambesi-Impiombato A, Bernardo DD (2007) How to infer gene networks from expression profiles. Molecular Systems Biology 3: 78.
4. Smet RD, Marchal K (2010) Advantages and limitations of current network inference methods. Nature Reviews Microbiology 8: 717–729.
5. Butte A, Kohane I (2000) Mutual information relevance networks: functional genomic clustering using pair wise entropy measurements. Pacific Symposium on Biocomputing 5: 418–429.
6. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. (2006) ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7: S7.
7. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, et al. (2007) Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biology 5: e8.
8. Meyer PE, Kontos K, Lafitte F, Bontempi G (2007) Information-theoretic inference of large transcriptional regulatory networks. EURASIP Journal on Bioinformatics and Systems Biology 2007: 79879.
9. Altay G, Emmert-Streib F (2010a) Inferring the conservative causal core of gene regulatory networks. BMC Systems Biology 4: 132.
10. Altay G, Emmert-Streib F (2011a) Structural influence of gene networks on their inference: analysis of C3NET. Biology Direct 6: 31.
11. Altay G, Asim M, Markowetz F, Neal DE (2011b) Differential C3NET reveals disease networks of direct physical interactions. BMC Bioinformatics 12: 296.
12. Mani S, Cooper GF (2004) Causal discovery using a bayesian local causal discovery algorithm. In: Proceedings of the World Congress on Medical Informatics 2004; 7–11 September 2004; San Francisco, California, United States. Medinfo 2004. Available: http://ebooks.iospress.nl/publication/21082. Accessed 15 June 2013.
13. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. Journal Computational Biology 7: 601–620.
14. Yu J, Smith VA, Wang PP., Hartemink AJ, Jarvis ED (2004) Advances to bayesian network inference for generating causal networks from observational biological data. Bioinformatics 20: 3594–3603.
15. Tsamardinos I, Aliferis CF, Statnikov A (2003) Time and sample efficient discovery of Markov blankets and direct causal relations. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2003; 24–27 August; Washington, DC, United States. KDD 2003. Available: http://dl.acm.org/citation.cfm?id = 956838. Accessed 15 June 2013.
16. Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsoukos XD (2010) Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: algorithm and empirical evaluation. Journal of Machine Learning Research 11: 171–234.
17. Statnikov A, Aliferis C (2010). Analysis and computational dissection of molecular signature multiplicity. PLoS Computional Biology 6: e1000790.
18. Haury AC, Mordelet F, Vera-Licona P, Vert JP (2012) TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. BMC Systems Biology 6: 145.
19. Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68: 49–67.
20. Lèbre S, Becq J, Devaux F, Stumpf MPH, Lelandais G (2010) Statistical inference of the time-varying structure of gene-regulation networks. BMC Systems Biology 4: 130.
21. Meinshausen N, Bühlmann P (2010) Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72: 417–473.
22. van Someren EP, Vaes BL, Steegenga WT, Sijbers AM, Dechering KJ, et al. (2006) Least absolute regression network analysis of the murine osteoblast differentiation network. Bioinformatics 22: 477–484.

23. Schafer J, Strimmer K (2005) An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics 21: 754–764.
24. Greenfield A, Madar A, Ostrer H, Bonneau R (2010) DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models. PLoS ONE 5: e13397.
25. Watkinson J, Liang K-C, Wang X, Zheng T, Anastassiou D (2009) Inference of regulatory gene interactions from expression data using three-way mutual information. Annals of the New York Academy of Sciences 1158: 302–313.
26. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. PLoS ONE 5: e12776.
27. Küffner R, Petri T, Tavakkolkhah P, Windhager L, Zimmer R (2012) Inferring Gene Regulatory Networks by ANOVA. Bioinformatics 28: 1376–1382.
28. Karlebach G, Shamir R (2012) Constructing logical models of gene regulatory networks by integrating transcription factor-DNA interactions with expression data: an entropy-based approach. Journal of Computational Biology 19: 30–41.
29. Yeung KY, Bumgarner RE, Raftery AE (2005) Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. Bioinformatics 21: 2394–2402.
30. Yip KY, Alexander RP, Yan K-K, Gerstein M (2010) Improved reconstruction of in silico gene regulatory networks by integrating knockout and perturbation data. PLoS ONE 5: e8121.
31. Sîrbu A, Ruskin HJ, Crane M (2011) Stages of gene regulatory network inference: the evolutionary algorithm role. In: Evolutionary Algorithms, Kita E, editor. InTech. 521–546.
32. Song MJ, Lewis CK, Lance ER, Chesler EJ, Yordanova RK, et al. (2009) Reconstructing generalized logical networks of transcriptional regulation in mouse brain from temporal gene expression data. EURASIP Journal on Bioinformatics and Systems Biology 2009: 545176.
33. Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, et al. (2010) Revealing strengths and weakness of methods for gene network inference. Proceedings of National Academy of Science USA 107: 6286–6291.
34. Altay G, Emmert-Streib F (2010b) Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics 26: 1738–1744.
35. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, and et al. (2012) Wisdom of crowds for robust gene network inference. Nature Methods 9: 796–804
36. Breiman L (1996) Bagging Predictors. Machine Learning 24: 123–140.
37. Stolovitzky G, Monroe D, Califano A (2007) Dialogue on reverse-engineering assessment and methods: The dream of high-throughput pathway inference. Annals of the New York Academy of Sciences 1115: 1–22.
38. Schaffter T, Marbach D, Floreano D (2011) GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. Bioinformatics 27, 2263–2270.
39. Schafer J, Opgen-Rhein R (2006) Reverse engineering genetic networks using the GeneNet package. R News 6: 50–53.
40. Meyer PE, Lafitte F, Bontempi G (2008) minet: A R/Bioconductor package for inferring large transcriptional network using mutual information. BMC Bioinformatics 9: 461.
41. Rijsbergen CJV (1979) Information Retrieval. London: Butterworth-Heinemann.
42. R Development Core Team R: A language and environment for statistical computing, (R Foundation for Statistical Computing, Vienna, 2012)

BMC Systems Biology

DATABASE Open Access

# A comprehensive map of the influenza A virus replication cycle

Yukiko Matsuoka[1,2], Hiromi Matsumae[3], Manami Katoh[1], Amie J Eisfeld[4], Gabriele Neumann[4], Takeshi Hase[2], Samik Ghosh[2], Jason E Shoemaker[1], Tiago JS Lopes[1], Tokiko Watanabe[1], Shinji Watanabe[1,5], Satoshi Fukuyama[1], Hiroaki Kitano[1,2,6,7] and Yoshihiro Kawaoka[1,4,8,9,10*]

## Abstract

**Background:** Influenza is a common infectious disease caused by influenza viruses. Annual epidemics cause severe illnesses, deaths, and economic loss around the world. To better defend against influenza viral infection, it is essential to understand its mechanisms and associated host responses. Many studies have been conducted to elucidate these mechanisms, however, the overall picture remains incompletely understood. A systematic understanding of influenza viral infection in host cells is needed to facilitate the identification of influential host response mechanisms and potential drug targets.

**Description:** We constructed a comprehensive map of the influenza A virus ('IAV') life cycle ('FluMap') by undertaking a literature-based, manual curation approach. Based on information obtained from publicly available pathway databases, updated with literature-based information and input from expert virologists and immunologists, FluMap is currently composed of 960 factors (i.e., proteins, mRNAs etc.) and 456 reactions, and is annotated with ~500 papers and curation comments. In addition to detailing the type of molecular interactions, isolate/strain specific data are also available. The FluMap was built with the pathway editor CellDesigner in standard SBML (Systems Biology Markup Language) format and visualized as an SBGN (Systems Biology Graphical Notation) diagram. It is also available as a web service (online map) based on the iPathways+ system to enable community discussion by influenza researchers. We also demonstrate computational network analyses to identify targets using the FluMap.

**Conclusion:** The FluMap is a comprehensive pathway map that can serve as a graphically presented knowledge-base and as a platform to analyze functional interactions between IAV and host factors. Publicly available webtools will allow continuous updating to ensure the most reliable representation of the host-virus interaction network. The FluMap is available at http://www.influenza-x.org/flumap/.

**Keywords:** Drug targets, FluMap, Host factors, Influenza virus, Pathways

## Background

Rapid adaption to new hosts and frequent antigenic alterations make the prevention and treatment of influenza A virus (IAV) infections challenging. To develop better intervention methods, a deeper understanding of the viral infection process and the host response to infection are critical. IAV possesses an RNA genome of ~12 kilobases (kb) that encodes 10–12 proteins. As a consequence of this small coding capacity, IAVs usurp and modify the host cell machinery to replicate. Several studies have now provided extensive datasets on cellular factors that may directly or indirectly affect the viral life cycle [1-6] (works are reviewed in [7,8]). However, it has been challenging to integrate and compare this information with other published data, and to develop a complete picture of the viral life cycle. To this end, a comprehensive illustration and annotation of the current knowledge of the IAV infection process with underlying textual descriptions would greatly assist in elucidating the mechanisms by which influenza

* Correspondence: kawaoka@ims.u-tokyo.ac.jp
[1]JST ERATO Kawaoka infection-induced host responses project, Minato-ku, Tokyo 108-8639, Japan
[4]Department of Pathological Science, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53711, USA
Full list of author information is available at the end of the article

viruses utilize host cell machinery and evade host defence mechanisms.

Interaction networks, such as protein-protein interaction (PPI) networks, are often used to visualize interactions among entities (for example, proteins), but such networks do not capture the directionality of interactions (for example, "who stimulates whom"). In addition, interaction networks typically do not capture interactions between different types of molecules (for example, protein–RNA interactions). For these purposes, pathway visualization approaches, that is, 'pathway maps' – such as those described for Epidermal Growth Factor Receptor (EGFR) [9], Toll-like receptor (TLR) [10,11], retinoblastoma protein/E2F (Rb/E2F) [12], yeast [13], or mammalian target of rapamycin (mTOR) [14] – are better suited. Furthermore, while a graphical representation provides the best overview of biological phenomena, it is also important to represent the model in a machine-readable format that can be rigorously analysed using *in silico* methods.

Several projects have generated open-source, open-access databases of viral genome sequences, structural and interaction data for viral proteins, and host response data (e.g., the Influenza Research Database [15], the Influenza Virus Resource [16], and VirusMINT [17]); or pathway maps of IAV infections (e.g., Reactome [18,19] and KEGG [20]). Among the available pathway maps, the 'Influenza A' KEGG map contains only a limited number of entities and reactions. A greater amount of detail is available in the Reactome 'Influenza Life Cycle' and 'Host Interactions with Influenza Virus Factors' maps; however, these maps have not been updated since their creation in 2006, and the lack of integration between them makes it difficult to obtain insights into how they are interrelated. Both the KEGG and Reactome maps also lack significant additional information about pathway entities (e.g., PubMed IDs, supportive references) and neither is readily amenable to computational analysis approaches unless their pathways are converted to standard file formats that can be imported to analytic tools such as Cytoscape. Therefore, the usefulness of both the KEGG and the Reactome pathways as information- and hypothesis-generating platforms is limited.

To address these shortcomings and improve our understanding of influenza virus infections, we created an integrated, comprehensive and interactive map that includes both viral life cycle and host response processes (i.e., the "FluMap") (Figure 1). Here, we describe the FluMap construction strategy, highlight some of the map's major characteristics, and demonstrate how it can be used as a bioinformatics tool. FluMap will be made available at a website (http://www.influenza-x.org/flumap) and can be used in conjunction with the online curation platform Payao [21] and a pathway browsing platform iPathways+ [22]. Together, these tools enable the scientific community to freely and

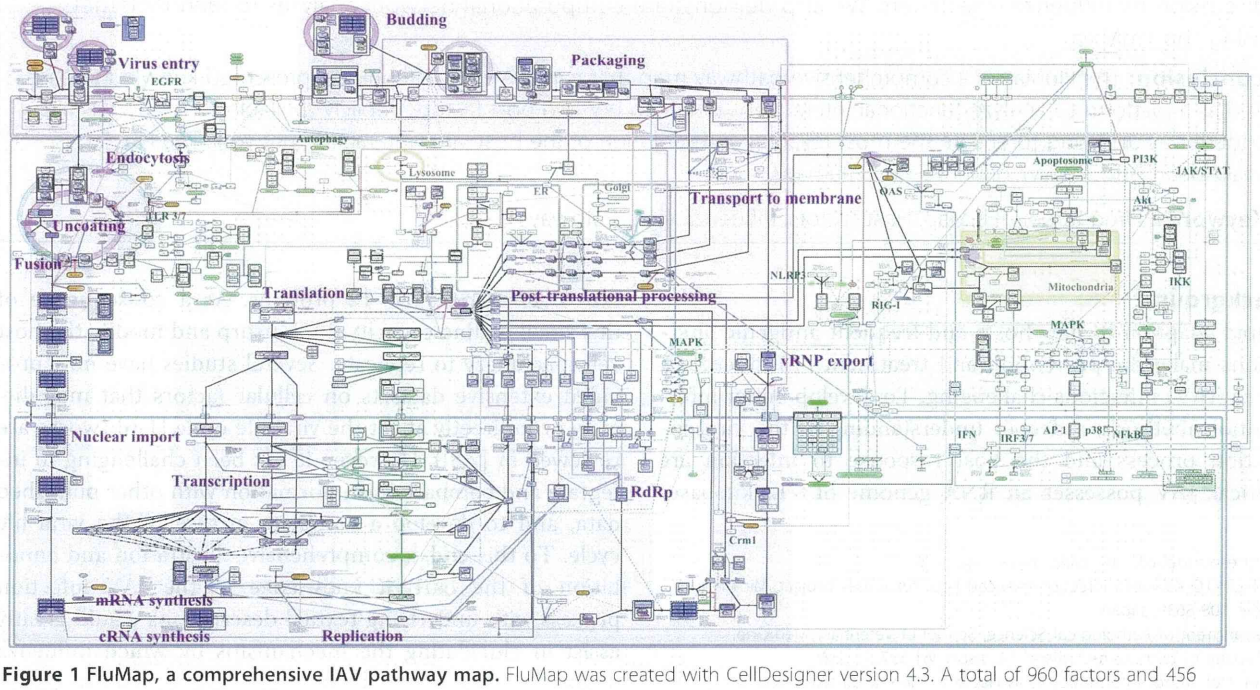


**Figure 1 FluMap, a comprehensive IAV pathway map.** FluMap was created with CellDesigner version 4.3. A total of 960 factors and 456 reactions were included. The SBML and high-resolution image PDF files are available as Additional Data. When FluMap is opened in CellDesigner, all factors, reactions, and cellular compartments included in the map are listed in the SBML file, and symbols used to build the map are illustrated in the legend of Additional file 3. (See also Additional file 2).

simultaneously browse, add, and update FluMap information, thus providing the foundation for a powerful, community-curated knowledge base to further influenza virus research.

## Construction and contents

### Construction of a comprehensive, knowledge-based pathway map of influenza virus infection (FluMap)

The information used to build the FluMap (Figure 1; Additional file 1, Additional file 2, Additional file 3, and Additional file 4) was derived from several different sources. First, we manually reconstructed the Reactome 'Influenza Life Cycle' and 'Host Interactions with Influenza Virus Factors' maps [18,19] into a single map file (the FluMap pathway 'skeleton'). Next, we manually incorporated information about host pathways that are activated in response to influenza virus infection, and - for all validated interaction partners of IAV factors - we included information about downstream signalling and processing events (e.g., phosphorylation cascades). Host factor and pathway data were obtained by using published pathway maps, KEGG [20], PANTHER [23] and/or Reactome [18,19] pathway map databases. Finally, we manually integrated literature-based information regarding the influenza virus replication cycle and virus-host interactions that was absent from the Reactome pathway 'skeleton' (Approximately 13% of the interactions in the map were derived from the "skeleton", and another 10% were collected from the public pathway databases). This information was identified from review articles, extensive searches on PubMed, and text-mining platforms such as iHOP [24].

Although recent siRNA screens [2-4,6], protein-protein interaction studies [5,25-28] and global proteome analyses [29,30] have identified a substantial number of cellular factors with potential roles in the IAV infection process, FluMap includes only those with roles that have been experimentally confirmed. In addition, FluMap focuses on *intracellular* events, and does not include intercellular events (e.g., immune cell interactions). All curated reactions and interactions in the FluMap were categorized into specific parts of the influenza infection process (e.g.,'vRNP export), and for reactions imported from Reactome, we kept the reaction name from this database (e.g.,'Entry of Influenza Virion into Host Cell via Endocytosis'). A similar naming strategy was used for other reactions manually added to the map (Additional file 2 and Additional file 5).

To build the graphical representation of the FluMap (Figure 1; Additional file 2, Additional file 3, and Additional file 4), we used CellDesigner ver.4.3 [31], a modeling software that can be used to depict cellular processes step-by-step, edit annotations, and provide links to reference databases [32]; we also used Payao, a community-based, collaborative web service platform for gene-regulatory and biochemical pathway model curation [21]. The map is stored in the standard Systems Biology Markup Language (SBML) (Additional file 4), a data exchange format based on XML [33]; and it is represented in the CellDesigner's graphical notation [34], which adheres to the Systems Biology Graphical Notation (SBGN) standards [35]. Map graphics were produced using SBGN 'process description' language (Additional file 2), which allows for visualization of state transitions (e.g., stimulation or inhibition events). By using standard formats, we have enabled FluMap to be adaptable to multiple network analysis tools such as Cytoscape or to simulation by employing user-supplied kinetic laws and SBML-compliant simulators.

In addition to a detailed visual representation, we generated comprehensive, text-based annotations, which are stored in the same map file. CellDesigner enables annotation of information in three different ways: (1) in the Notes section; (2) in the MIRIAM (*M*inimum *I*nformation *R*equired *I*n the *A*nnotation of *M*odels) [36] format section; and (3) in an additional layer overlaying the base model. For FluMap, we used all three annotation options to maximise data accessibility (see Additional file 2 for details). Gene IDs, UniProt accession numbers, PubMed (reference) IDs, and Reactome IDs are stored in the Notes and MIRIAM sections. The Notes section also includes information about the intracellular location of specific interactions or reactions (e.g.,'Nucleus' or 'Mitochondria'), the stage of the infection process at which it occurs (e.g., 'Virus Entry' or 'vRNP Export'), the participation of specific viral proteins, and association with multi-protein complexes that regulate host processes (e.g., 'Apoptosome') or signalling pathways (e.g.,'MAPK'). Additional reference information (e.g.,'HA1: Yoshida R *et al.* 2009') is captured in the layer that overlays the base model. CellDesigner provides direct access to the relevant databases mentioned in the Notes section through the CellDesigner database menu, and the weblinks in the MIRIAM section by pressing the access button.

While process description diagrams capture all details of biological processes, it is also useful to have a simplified overview of the system. We, therefore, used the 'reduced notations' option in CellDesigner to illustrate the relationships between entities (positive/negative inferences, modulation, trigger, etc.). This notation depicts positive/negative influence interactions, rather than detailed events, such as phosphorylation or catalysis in the process description notation (see Additional file 2 sections B and C). Finally, we used this notation to manually construct a simplified map (Figure 2; compare to the fully detailed FluMap in Figure 1) that provides a high-level overview of the IAV replication cycle.

The FluMap is posted under http://www.influenza-x.org/flumap, where users can browse its contents using a pathway-browsing platform (iPathways+) and provide updates and improvements using a manual curation platform (Payao).

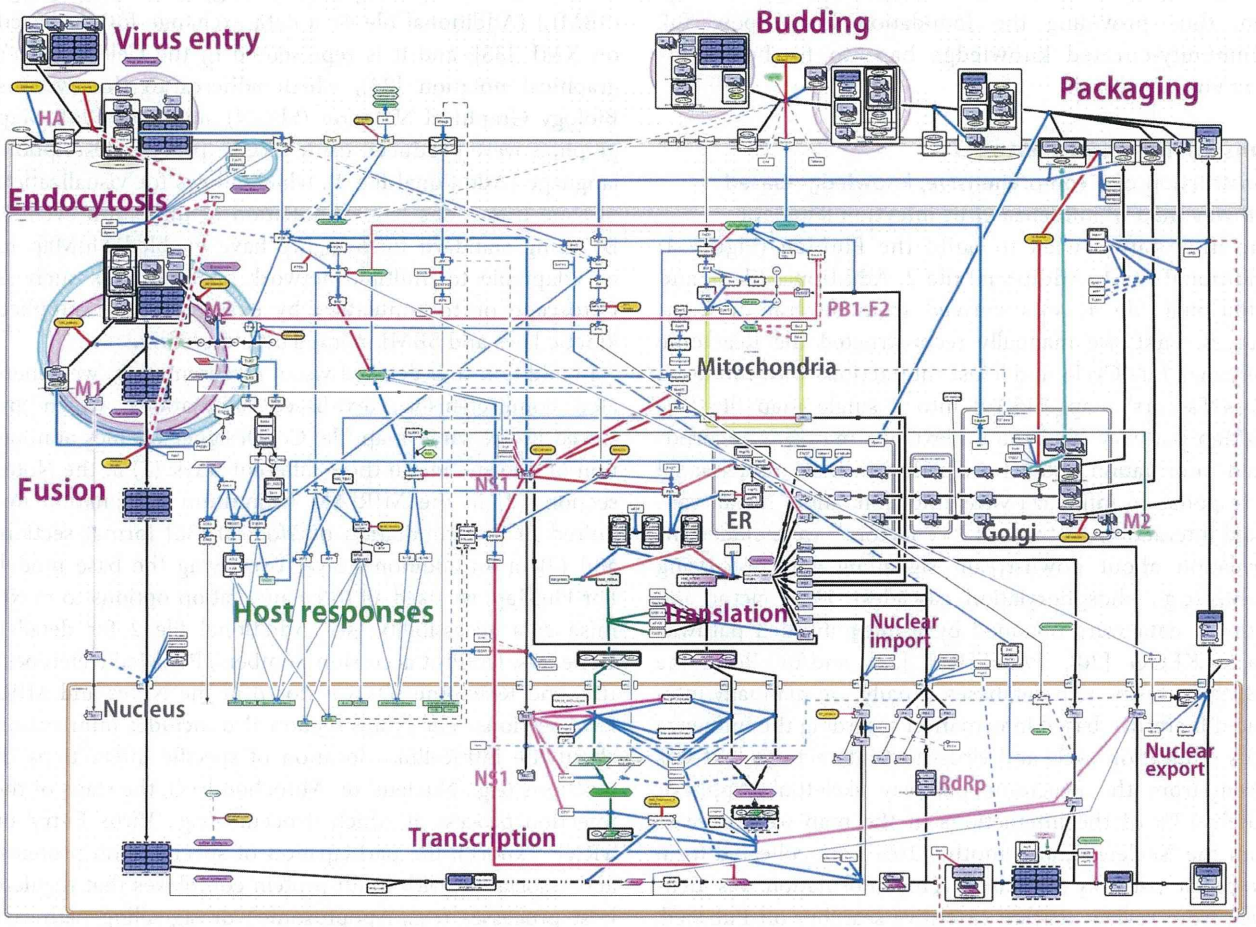


**Figure 2 Simplified version of FluMap.** To generate a simplified version of FluMap for a high-level overview, we extracted central components and reactions from the FluMap (virus factors (purple), host factors (green), antiviral factors (orange)), focusing on the inhibition (red) or activation (blue) of IAV replication by host factors.

### General characteristics of FluMap

The comprehensive FluMap (Figure 1; see Additional file 4 for the original SBML file) contains 960 factors (696 species + 264 factors hidden in complexes) and 456 reactions. Among these, there are 558 viral and cellular proteins, 212 molecular complexes (composed of more than one component), 12 ions, 55 'phenotypes' (representing biological events such as apoptosis or autophagy), and 18 antiviral compounds. As described, all reactions are annotated with PubMed IDs in the Notes section; the entire map is annotated with 476 papers (Additional file 5 and Additional file 6). FluMap thus provides a significant improvement over the Reactome influenza infection pathway, which included 156 species and 58 reactions as of April 2012.

While the FluMap adopts the SBGN's process description graphical notation, the simplified map (Figure 2; Additional file 7) adopts the 'reduced notation' similar to SBGN's activity flow, which better facilitates visualization of the virus-host interplay at different stages of the virus life cycle. To better highlight the virus-host interplay, we manually restructured the simplified FluMap into a linear flowchart that is divided into viral and host response events (Figure 3; Additional file 8). In this representation, it is easier to track the different phases of the viral life cycle (entry, endocytosis, transcription/translation, assembly, and budding).

### Description of the IAV replication cycle

In the following sections, we summarize our current knowledge of the IAV replication process as outlined in the FluMap (Figure 1), focusing on virus-host interactions.

#### *Virus entry*

The first step in the IAV life cycle is virus binding to host cells ('Virus Entry', Figure 1). The viral hemagglutinin (HA) protein is critical for this step since it binds to sialic acids on host cell glycoproteins or glycolipids. The HA proteins of human IAVs preferentially recognize sialic acid linked to galactose by an α2,6-linkage (Siaα2,6Gal) [37-42]

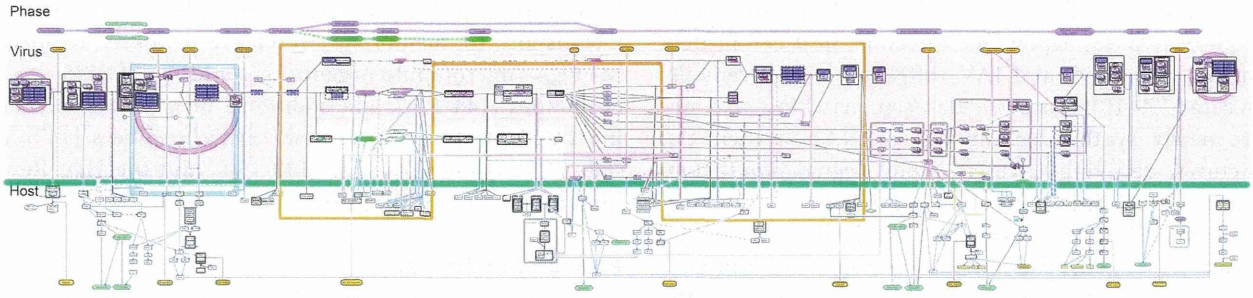


**Figure 3 Flowchart of the IAV life cycle.** The simplified FluMap (Figure 2) was converted into a linear process flow diagram and separated into the different phases of the viral life cycle (top portion), viral processes (middle portion), and host interaction factors (lower portion). Interactions are classified as inhibitory (red), stimulating (blue), and transition (black). Enclosure in the middle (orange line) indicates nucleoplasm.

that is predominant on epithelial cells in the human upper respiratory tract [43-49]. In contrast, avian virus HA proteins preferentially bind to Siaα2,3Gal [37-42], which is predominantly found on epithelial cells of the duck intestine (where avian influenza viruses replicate) [39,50-52]. These differences in HA receptor specificity are a critical determinant of IAV host range (reviewed in [53-55]).

### *Endocytosis*

Following receptor binding, IAVs enter cells through receptor-mediated endocytosis ('Endocytosis' in Figure 1). Clathrin-mediated endocytosis appears to be the primary internalization pathway of IAVs [56]; however, clathrin-independent endocytosis [57,58] and macropinocytosis [59,60] have also been described for IAV internalization. Several host factors including the small GTPases Rab5 and Rab7 [61], and interferon-inducible transmembrane IFITM protein family members (i.e., IFITM1, IFITM2, IFITM3) interfere with IAV internalization [1,62].

### *Fusion*

At the low pH of the late endosome, HA undergoes an irreversible conformational shift which expels the N-terminus of the HA2 subunit (the so-called 'fusion peptide') so that it can insert into the endosomal membrane, resulting in the fusion of the viral and endosomal membranes (reviewed in [63]) ('Fusion' in Figure 1). Through an ion channel formed by the viral M2 protein, proton influx also acidifies the interior of the virus particles, leading to the dissociation of the viral matrix protein (M1) from viral ribonucleoprotein (vRNP) complexes [64]. vRNPs are composed of one of the eight viral RNAs (vRNAs), which are wrapped around the nucleoprotein (NP) and are also associated with the viral polymerase complex (see below). Dissociation from M1 allows vRNP release into the cytoplasm and subsequent nuclear import, which is mediated by the cellular nuclear import factors importin-α (karyopherin-α) and importin-β (karyopherin-β) [65-72] ('Nuclear import' in Figure 1). The M1 protein, after dissociating from vRNP complexes in late endosomes, is imported into the nucleus separately [73].

### *Virus replication and transcription*

The replication and transcription of IAV genomic RNAs takes place in the nucleus and is catalysed by the trimeric viral polymerase complex composed of PB2, PB1, and PA subunits ('Replication', and 'Transcription' in Figure 1). Viral RNA replication starts with the synthesis of a positive-sense copy of the vRNA, termed complementary RNA (cRNA) (reviewed in [74]). This cRNA is then copied to produce large amounts of vRNA (reviewed in [75,76]). Several host factors have been identified that may play a role in viral genome replication (reviewed in [77-79]).

Viral RNA transcription is initiated by the binding of PB2 to the 5′-cap structure of host mRNAs [80-82]. The endonuclease activity of PA [83] then 'snatches' the cap structure and the 10–13 nucleotides included with the cap serve as a primer for viral mRNA synthesis. The synthesis of viral mRNAs is carried out by the polymerase activity of PB1 [84]. The nuclear export of viral mRNAs is reviewed in York and Fodor [79]. Transcription proceeds until the polymerase complex stalls at a polyadenylation signal near the end of the viral RNA [85-88].

Two IAV mRNAs (derived from the two smallest vRNA segments, M and NS) are spliced to yield the M1 and M2, or the interferon antagonist (NS1) and nuclear export (NEP/NS2) proteins. Splicing is carried out by the host cell splicing machinery, but is likely regulated by NS1 [89,90], which binds to several cellular splicing components such as U6 small nuclear RNAs [91,92] and UAP56, a splicing factor involved in spliceosome formation [93,94].

### *Translation*

Influenza viral mRNAs are translated by the host cell translation machinery ('Translation' in Figure 1); thus not surprisingly, several cellular translation factors such

as eIF4A (eukaryotic initiation factor-4A), eIF4E, and eIF4G interact with viral mRNAs and/or polymerase complexes [95-98]. Upon IAV infection, host cell protein synthesis is limited, and IAV mRNAs are preferentially translated [99-101]. In particular, 'cap-snatching' may deplete newly synthesized, nuclear mRNAs of their cap structures, resulting in their rapid degradation before nuclear export and translation. In addition, the interaction of NS1 with the cellular PABII (poly(A)-binding protein II) [95,98] and CPSF (cleavage and polyadenylation specificity factor) proteins [102,103], and the interaction of the viral polymerase complex with the C-terminal domain of the largest subunit of cellular DNA-dependant RNA polymerase II (Pol II) [104,105] may contribute to the inhibition of host mRNA synthesis (reviewed in [106]).

After their synthesis in the cytoplasm, the viral polymerase subunit proteins and NP are imported into the nucleus via their nuclear localization signals [71,74,107-118] to catalyse the replication and transcription of vRNA. In addition, the M1 [64,119], NEP/NS2 [120], and NS1 [121] proteins are imported into the nucleus to execute their roles in vRNP nuclear export (M1 and NEP/NS2) or the processing and export of cellular and viral mRNAs (NS1) (reviewed in [122]).

#### *vRNP export*

The nuclear export of newly synthesized vRNP complexes requires the viral NEP/NS2 [123-126] and M1 [66,127,128] proteins. The latter is thought to form a bridge between vRNPs and NEP/NS2 [129-131], and M1 association with vRNP may require M1 SUMOylation [132]. In the nucleus, vRNPs destined for export are targeted to chromatin where they associate with Rcc1, and export is mediated by the cellular export factor Crm1 ('vRNP export' in Figure 1) [125,127,133] in a manner that is likely regulated by phosphorylation [65,128,134-137]. The cellular Y box binding protein 1 (YB-1) protein also associates with vRNPs in the nucleus, is likely exported from the nucleus in conjunction with vRNPs, and facilitates vRNP association with microtubules for transport to the plasma membrane (see below) [138].

Following their synthesis by the cellular translation machinery, the viral HA, neuraminidase (NA), and M2 proteins enter the endoplasmic reticulum (ER) where they are glycosylated (HA and NA) (reviewed in [139,140]) or palmitoylated (HA and M2). Cleavage of the HA proteins of highly pathogenic avian H5 and H7 viruses (which possess multiple basic amino acids at the HA cleavage site) into the HA1 and HA2 subunits occurs most likely by cellular furin-like proteases [141] in the *trans*-Golgi network; this cleavage event is critical for the virulence of influenza viruses [142,143].

#### *Transport of virus proteins to the cell membrane*

Transport of viral proteins to the plasma membrane ('Transport to membrane' in Figure 1) likely requires MTOCs (microtubule-organizing centers) [144,145], microtubules [144-146], and additional host factors including COPI (coatomer I) protein family members [147], a Rab GTPase (Rab11A) [145,148-150], and the HIV Rev-binding protein (HRB) [151].

#### *Packaging and budding*

At the plasma membrane, HA and NA associate with lipid rafts (membrane regions rich in sphingolipids and cholesterol) that are the site of influenza virus budding [152-160] ('Packaging' and 'Budding' in Figure 1). The assembly and virion incorporation of the eight vRNPs requires segment-specific packaging signals in the viral RNAs [161,162]. The M1 protein may play a role in the assembly process since it interacts with lipid membranes [163-165], vRNPs [130,131,166] (reviewed in [167,168]), and NEP/NS2 [129,169]. In addition, some evidence suggests the possibility that the M2 cytoplasmic tail mediates vRNP incorporation into the assembling virus particle [170].

Influenza virus budding does not require the proteins of the endosomal sorting complexes that are required to transport ESCRT complexes, which are utilized by a number of other viruses for budding. Rather, M2, which is found in the raft periphery [152,157,171], appears to mediate membrane scission and particle release [172]. This process may also require the cellular F1Fo ATPase [25]. The enzymatic activity of the viral NA protein removes sialic acids from host cells and from glycoproteins on virions, allowing virus release and preventing virion aggregation (reviewed in [55,75]).

#### *Post-translational processing*

Several post-translational modifications have been described for IAV proteins, including the glycosylation of HA (reviewed in [75,142]) and NA [173], the palmitoylation (S-acylation) of HA and M2 (reviewed in [174]), and the SUMOylation (i.e., conjugation with the *s*mall *u*biquitin-like *m*odifier) of M1 [132,175], NS1 [176,177], NP [175], PB1 [175], and NEP/NS2 [175] ('Post-translational processing' in Figure 1). Moreover, phosphorylation of M1 [137,178] and NP [107,179-183] may affect vRNP nuclear import and export [66,113,128,134]. Phosphorylation of NS1 [184] and PB1-F2 (a short protein synthesized from the PB1 gene; see below) affects virulence [185], although the mechanisms are not yet fully understood. These phosphorylation events are catalysed by several cellular kinases such as PKC (protein kinase C) which phosphorylates M1 [136], PB1-F2 [185], NS1 [184,186], and PB1 [186], or by CDKs (cyclin-dependent kinases) and ERKs (extracellular signal-regulated kinases), which phosphorylate NS1 [187].

### Host responses

IAV infections trigger multiple host antiviral responses (reviewed in [188,189]). These interactions are summarized in the FluMap (Figure 1) and in the flowchart that depicts the different stages of the viral life cycle (Figure 3).

As a major host defence mechanism, pattern recognition receptors (PRRs) recognize infecting agents and trigger cellular antiviral responses (reviewed in [190]). To date, three major classes of PRRs [Toll-like receptors (TLRs); RIG-I-like receptors; NOD-like receptors (NLRs)] are recognized, all of which play a role in the defence against IAV infections. The activation of PRRs leads to increased production of type I interferon (IFN) and chemokines/cytokines, resulting in the upregulation of antiviral factors.

IAV infections are recognized by TLR3 [191,192], which acts through the adaptor molecule TRIF (TIR-domain-containing adapter-interferon-beta) to stimulate IFN-regulated factor 3 and NFκB (nuclear factor-kappa beta); TLR7 [193,194], which signals through the adaptor protein MYD88 (myeloid differentiation factor 88) and induces IRF7 (interferon regulatory factor 7) and NFκB; and RIG-I [195-198], which signals through MAVS (mitochondrial antiviral signalling), also known as IPS-1, and leads to the stimulation of IRF3, IRF7, and NFκB. Moreover, IAV infection activates the inflammasome [199-203], resulting in the cleavage and activation of pro-caspase-1, interleukin-1 beta (IL-1β), and IL-18.

PRR stimulation leads to the synthesis of IFNα/β, which binds to the ubiquitously expressed IFNα/β (IFNAR) receptor, resulting in the upregulation of the JAK/STAT (janus kinase/signal transducer and activator of transcription) pathway. JAK/STAT signalling induces the formation of a transcription factor complex (composed of STAT1, STAT2, and IRF-9) that upregulates the expression of IFN-stimulated genes (ISGs). A number of ISGs encode proteins with antiviral functions, such as PKR (protein kinase R), OAS (2′-5′-oligoadenylate synthetase), RNaseL (ribonuclease L), Mx, ISG15, IFITM family members, and viperin (see below for details). IAVs have thus evolved mechanisms to counter these host anti-viral defence strategies, primarily through the actions of the NS1 and PB1-F2 proteins.

NS1 is the major viral IFN antagonist ([204]; reviewed in [189,205]). It blocks RIG-I-mediated innate immune responses by targeting RIG-I [195,206] and/or TRIM25 (tripartite motif-containing protein 25) [207], and interferes with caspase-1 activation [208].

NS1 also interferes with the effects of several antiviral host factors. IAV infection activates PKR, resulting in the phosphorylation of the eukaryotic translation initiation factor eIF2α and the subsequent shutdown of protein synthesis. This activation is inhibited by NS1 [209-214]. NS1 also controls the antiviral activity of OAS and RNaseL, a cellular nuclease that degrades viral RNA [215]. ISG15 (interferon-stimulated gene 15) is an IFNα/β-induced, ubiquitin-like protein that conjugates to a wide array of cellular proteins, thus blocking their function. It affects IAV infection by interfering with the function of NS1 [216,217].

IAV infection stimulates the phosphoinsitide-3-kinase PI3K/Akt pathway [218-226], which has pro- and antiviral functions (reviewed in [219]). In particular, this pathway is activated by NS1 binding to the p85 subunit of PI3K [218,221,224,226-228] and by IAV vRNAs via RIG-I [229]. Activation of the PI3K/Akt pathway is critical for efficient IAV replication [219,220], likely by preventing premature apoptosis [222,227,230-232].

The C-terminal four amino acids of most NS1 proteins comprise a PDZ ligand domain motif [233] that affects virulence [234-236] (reviewed in [237]), most likely through interaction with the cellular PDZ domain proteins Scribble, Dlg1 (disks large homolog 1), and membrane-associated guanylate kinase MAGI-1, -2, and −3 [238-240], which play roles in the regulation of apoptosis or tight junction formation.

NS1 also reduces the levels of IFNα/β mRNA by interfering with mRNA splicing [90-92,241] and the polyadenylation and nuclear export of cellular pre-mRNAs [90,91,102,241-246].

PB1-F2 is a short protein of 87–90 amino acids encoded by the +1 reading frame of most, but not all, IAV PB1 genes. It localizes to the mitochondrial membrane [247-249] where it interacts with the mitochondrial membrane proteins ANT3 (adenine nucleotide translocator 3) and VDAC1 (voltage-dependent anion-selective channel 1) [250], resulting in membrane depolarization [251,252] and the induction of apoptosis [247,248,250]. However, a recent study suggested that the induction of apoptosis may not be the major function of PB1-F2 [253]. Rather, PB1-F2 may interfere with the function of MAVS (mitochondrial antiviral-signalling protein) [254], and the resulting inhibition of IFN induction may contribute to PB1-F2-conferred increases in pathogenicity, inflammation, and the frequency and severity of bacterial co-infections [255-259]. In addition, PB1-F2 binding to PB1 affects the intracellular localization of the polymerase protein and reduces polymerase activity, potentially affecting virulence [260].

Other host antiviral factors include the Mx proteins [261-263], which most likely interfere with viral replication [264-266]; members of the IFITM protein family, which interfere with IAV cell entry [1,62,267]; and viperin, which executes its antiviral activity by disrupting lipid rafts that are critical for IAV budding [268].

Other important host responses to IAV infection include the mitogen-activated protein kinase (MAPK) signalling pathways, which regulate multiple cellular events

including cell cycle control, cell differentiation, and apoptosis. All four of the currently recognized MAPK pathways [extracellular signal-regulated kinases 1/2 (ERK1/2); c-jun-N-terminal kinase (JNK); p38; and ERK5] are activated upon IAV infection [135,269-276]. Some of these pathways may have both pro- and antiviral functions [135,274,277-279].

#### *Antiviral compounds*

The FluMap also captures antiviral compounds that are directed against a viral factor or a host target that is critical for efficient viral replication (reviewed in [280-283]). See Additional file 9 for a summary table.

Currently, there are two types of FDA-approved anti-IAV compounds: M2 ion channel inhibitors (amantadine, rimantadine), and NA inhibitors (oseltamivir, zanamivir).

M2 ion channel inhibitors block the ion channel in the viral envelope formed by the viral M2 protein. They prevent the influx of hydrogen ions from the acidic late endosome into the interior of the virion, a process that is necessary for the release of vRNPs into the cytoplasm. However, these inhibitors are no longer recommended for use in humans because most circulating IAVs are resistant to these compounds [284].

The NA inhibitors oseltamivir and zanamivir are the only antivirals currently recommended worldwide for human use. Both compounds block the enzymatic activity of NA that is critical for efficient virus replication [285-288]. Resistance to NA inhibitors has been described but is not widespread among currently circulating IAVs (reviewed in [289]).

Several new antiviral compounds are in different stages of clinical development and/or have been approved for human use in some countries, including two new NA inhibitors, peramivir [290,291] and laninamivir [292], and a viral polymerase inhibitor, T-705 [293-295].

Other strategies include the development of compounds that interfere with virus replication (ribavirin) [296,297], NP function (nucleozin) [298-301], NS1 function (several candidates) [302-304], or HA function [chemical compounds such as arbidol [305] that block HA-mediated membrane fusion, or monoclonal antibodies (MABs) directed against HA]. In particular, the development of monoclonal antibodies that target conserved regions of the HA protein and interfere with HA-mediated receptor-binding or fusion has received increased attention [306-314].

Host factors that are crucial for efficient IAV replication but dispensable for cell viability may be interesting drug targets since they are less likely to acquire resistance to an antiviral compound compared with IAV proteins (reviewed in [281,283]). For example, the sialidase DAS181 (Fludase, NexBio), which cleaves sialic acids on human bronchial tissue and inhibits IAV infection [315-317], is currently in Phase II clinical trials in the U.S. [283]. Several other approaches that are in early stages of development include: (*i*) protease inhibitors that block cellular enzymes required for HA cleavage [318-320]; (*ii*) specific inhibitors of MAPKs, such as U0126 (a MAPK/ERK inhibitor), which blocks the nuclear export of vRNP complexes [135,321]; (*iii*) NFκB inhibitors such as acetylsalicylic acid (ASA; commonly known as aspirin) [322], although aspirin may have adverse effects in IAV-infected individuals [323,324]; and (*iv*) agonists of sphingosine-1-phosphate (S1P) receptors, such as AAL-R, which reduce lung pathology upon IAV infection, likely because of their effect on dendritic cell activation, T-cell responses, and cytokine levels [325,326].

#### *In silico* prioritization of potential drug targets

A critical quest in infectious disease research is to identify and prioritize novel potential therapeutic targets. In our *in silico* analysis of FluMap, we exploited a specific aspect of the network called controllability to identify molecules that, when inhibited, increase the likelihood of deregulating the virus replication cycle. Controllability is the ability to drive a network from any initial state to any desired state in a finite amount of time given a suitable choice of inputs [327]. From a biological network perspective, controllability analyses identify key molecular entities and processes that when perturbed can drive a biological system from a disease state to a healthy state [328].

To begin, we identified the smallest set of driver nodes (molecules, complexes, etc.) needed to attain complete control of all of the other nodes in the network. The size of this smallest set was directly related to how difficult it was to control the network in question. Networks that demand a large set of driver nodes are inherently more difficult to control. Further, as nodes are removed from the network, the identity of the driver nodes may change but, more importantly to our application, the number of driver nodes – and the associated difficulty of controlling the network – may remain fixed or also change. Thus, the second step of the analysis involved identifying 'critical' nodes that when removed from the network, *increased* the number of driver nodes necessary to elicit complete control, that is, increase the difficulty in controlling the network [329]. From a therapeutic perspective, inhibition of critical nodes/links would make it increasingly difficult for the virus to maintain control of the replication process. Further, controllability analysis can also be performed for the network links. Lastly, we investigated whether the critical nodes/links are associated with more commonly used network topology measures (e.g., nodes with a high number of neighbours (degree) or nodes that are bottlenecks in the network (betweeness)).

To facilitate the above analyses, we converted FluMap to a binary network by taking the direction of connections while ignoring the type of reaction (catalysis, inhibition etc.) (Figure 4; Additional file 10 and Additional file 11). Note that controllability analysis does not use the type of reaction (e.g., catalysis, inhibition etc.). Thus, ignoring the type of reaction does not affect the results.

Within the FluMap, we found that 256 (41.2%) of the nodes were driver nodes and 112 (18.0%) were critical nodes. Among the 137 critical links (15.3%), ~15% accounted for interactions among viral factors, whereas ~10% accounted for virus-host interactions. The remaining two-thirds accounted for reactions between host factors. Compared with previous studies [327], the driver nodes ratio of the FluMap is similar to that of metabolic networks (30%–40%), and lower than the gene regulatory networks (>80%).

Topology analysis revealed that critical nodes tended to have a higher degree and higher betweenness than noncritical nodes (two-sided Wilcoxon rank sum test [WRST] of the degree and $\log_{10}$ of the betweenness; $P < 2.2E\text{-}16$ and $P = 3.452e\text{-}06$, respectively, see Additional file 12). By using the node degree to prioritize the critical nodes, we found that the nuclear pore complex (NPC) and the three host proteins, Akt, PKC, and the Ran/GTPase complex (which plays a critical role in the export of proteins from the nucleus to the cytoplasm), are both critical and highly connected within the network. PKR and Y-box binding protein 1 (YB-1) come in the second tier. YB-1 is reported to assist in the transport of influenza virus RNP to microtubules [138]. Perturbation of these complexes/factors would thus be expected to have the greatest impact on the IAV life cycle.

Among the *137 critical interactions* identified, we did not find that critical interactions have a higher or lower edge betweeness than noncritical interactions (P=0.1, WRST of the $\log_{10}$ of the edge betweenness), but we did find that the ISG15-NS1 interaction and several interactions related to pH control involved molecules with high degree. Our controllability analysis identified several current antiviral compounds and targets, such as M2 ion channel inhibitors (which affect the pH inside the virion), the targets of sialidase, and viral polymerase inhibitors (Figure 4).

Our results suggest that the controllability analysis, together with network topology characteristics, can identify important factors for the viral life cycle that may be potential therapeutic targets as well as known drug targets. Given that the current map is constructed by manual curation, many important edges and nodes may be missing, so that the robustness of the controllability analysis cannot be assessed. Nonetheless, we show the potential of identifying and prioritizing critical nodes and edges that may be targeted for antiviral drug development.

## Utility and discussion

Here, we present FluMap, a comprehensive pathway map for IAV infections. This map is the most recent version of the IAV host-virus interaction map and includes

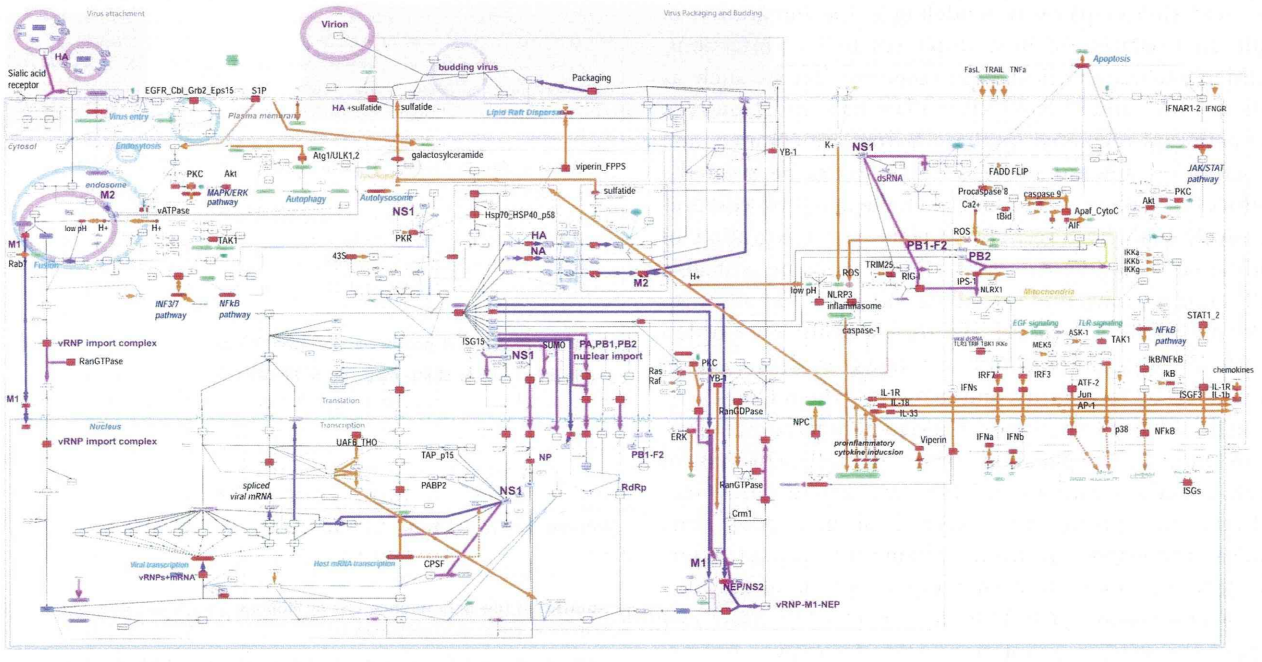


**Figure 4 Controllability analysis.** Critical factors (nodes) identified by the controllability analysis are shown in red. Thick magenta edges indicate critical interactions between viral and host proteins, purple edges indicate virus factor interactions/transitions (for example, transport processes), and orange edges indicate interactions between host factors.

a significantly higher number of factors than previous versions. It is intended to provide a platform for data sharing, community curation, and *in silico* analysis, such as network controllability analysis. We have made FluMap accessible online to allow for pathway and annotation browsing. We have also provided interactive features that will allow the research community to actively participate in improving and updating FluMap.

### FluMap as a data analytic platform

We applied a network controllability analysis to demonstrate that maps like FluMap can be used for *in silico* analysis. Although the controllability analysis we applied here does not take into consideration the nature of the interaction (for example, activating or inhibitory), our analysis identified several events known to be critical for the IAV life cycle, suggesting that the algorithm [327] can be effectively applied to process-descriptive pathway networks such as FluMap to identify and prioritize factors that could be targeted to affect the IAV life cycle. In addition to known targets, our analysis also identified factors that are not currently recognized as critical, such as YB-1; further experimental testing could address the significance of these events in IAV infections.

A comprehensive map such as FluMap can also be used to analyze large-scale data sets (obtained from 'omics' or siRNA inhibition studies) by using the data mapping function of CellDesigner or other visualization tools.

For a deeper insight into IAV virus-host interactions, the next step in pathway modeling is the integration of additional datasets of host responses to IAV infections. FluMap includes critical host response factors such as RIG-I, PKR, and the NLRP3-inflammasome. However, the pathways regulated by these factors are complex and a significant amount of 'cross-talk' occurs between the pathways, making it extremely challenging to comprehensively map host responses. Here, the integration of additional experimental data as they become available will improve our understanding of host responses to IAV infections. Moreover, future versions of FluMap could integrate intercellular reactions, such as events stimulated by interferons and cytokines/chemokines.

Lastly, a key distinction of FluMap compared with previous influenza replication cycles is the inclusions of strain-specific information. There are strong differences between the pathogenic potential of individual virus strains, and highly pathogenic strains may exploit different host machinery to ensure rapid replication and immune suppression [330-333]. Within FluMap, users can exploit the various annotations tools to analyse isolate-specific pathway interactions and attempt to identify critical molecular events associated with highly pathogenic infections. As future studies with H5N1, H7N9, or reconstructed Spanish influenza viruses reveal more information regarding virus-host interactions, the FluMap presented here will provide a basis for rapid consolidation and *in silico* exploration.

## Conclusions

We constructed a publicly available knowledge base called "FluMap" that contains 960 factors and 456 reactions. All reactions are annotated with PubMed IDs in the Notes section and isolate-specific information is available from many interactions; the entire map is annotated with 476 papers. FluMap is a comprehensive Influenza A virus replication life cycle and host response map, and is expected to be a valuable guidance map for those who study influenza infection.

## Availability and requirements

The FluMap is accessible at http://www.influenza-x.org/flumap/.

## Additional files

**Additional file 1: FluMap building and workflow of literature-based pathway modeling.** (a) FluMap was built based on information from the literature and from several pathway databases such as Reactome, KEGG, and PANTHER. The resulting map captures the viral life cycle and host responses. Extensive annotations are provided. We then manually generated a simplified map for high-level overview, and a map in which arrows outline the sequence of events during IAV infection (i.e., binding, internalization, nuclear import, etc.). We conducted controllability and network analyses over the FluMap to identify nodes essential to the replication process. Key interactions and nodes from these analyses are highlighted. (b) Summary of the literature-based pathway modeling process that converts and integrates textual information into a graphical representation. FluMap allows the community to browse, use, and comment on the information provided; this interface with the research community is shown in green.

**Additional file 2: How to browse FluMap.** This document explains how to browse FluMap at the website http://www.influenza-x.org/flumap/, and shows its graphical notation scheme, as well as the annotation policy we adopted for curation of the map. It also describes how to open the map file with CellDesigner for further analysis or modification, and how to curate the map on the Payao system (http://www.payaologue.org).

**Additional file 3: A poster version of FluMap.**

**Additional file 4: SBML map file of FluMap.** The SBML map file FluMap.xml can be browsed using CellDesigner. Please download CellDesigner at http://www.celldesigner.org/, install it, and open the SBML file FluMap.xml to browse FluMap by using CellDesigner. For usage of the software, see the documentation provided at the CellDesigner website: http://www.celldesigner.org/documents.html

**Additional file 5: Entities & Reactions List of FluMap.** This is a list of the entities (such as proteins, genes, etc.) and reactions (interactions between entities) in FluMap.

**Additional file 6: Reference List of FluMap.** This contains all of the references annotated in FluMap.

**Additional file 7: SBML map file of the simple version of FluMap.** The SBML map file of the simplified version of the IAV virus-host interaction map can be browsed by using CellDesigner. Please download CellDesigner at http://www.celldesigner.org/. For detail usage of the

software, see the documentation provided at the CellDesigner website: http://www.celldesigner.org/documents.html

**Additional file 8: SBML map file of the flowchart version of FluMap.** The SBML map file of the IAV virus-host interaction timeline can be browsed by using CellDesigner. Please download CellDesigner at http://www.celldesigner.org/. For detail usage of the software, see the documentation provided at the CellDesigner website: http://www.celldesigner.org/documents.html

**Additional file 9: Antiviral Drug List.** This is a list of the influenza-related antiviral drugs.

**Additional file 10: Controllability Analysis.** This document describes the protocol for the controllability analysis we conducted with FluMap.

**Additional file 11: Controllability Analysis Results.** This file contains the results of the controllability analysis, listing the critical, ordinary, and redundant nodes/links.

**Additional file 12: Topology Analysis Results.** This file contains the results of the topology analysis based on the controllability analysis results to prioritize the target candidates.

### Abbreviations

IAV: Influenza A virus; SBML: Systems biology markup language; SBGN: Systems biology graphical notation.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

YM, HK, and YK conceived the idea of FluMap. YM, HM, and MK developed the map. AJE, TW, SW, SF, GN, and TL reviewed and curated the map. TH, SG, JS, and YM conducted the controllability analysis. YM, HM, and GN wrote the manuscript. All authors approved the manuscript.

### Acknowledgements

We thank the following participants of several 'mapathons' (map-building meetings): Members of the Division of Virology, Department of Microbiology and Immunology, The Institute of Medical Science, The University of Tokyo (Takeo Gorai, Ai Kakumoto, Hirotaka Imai, Ryo Takano, Eiji Takeda, and Ryuta Uraki); members of the Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison (Masato Hatta, Chenjun Li, and Makoto Ozawa); a member of the Systems Biology Institute, Tokyo (Natal'a Polouliakh); and members of the JST ERATO Kawaoka infection-induced host response network project (Hiroko Fujii, Ken Fujii, Eiryo Kawakami, Yukiko Muramoto, Tadasuke Iizumi, Saori Sakabe, Yuko Shoya-Imai, and Yuriko Tomita). We also thank the software/platform development teams of CellDesigner, Payao, and iPathways+ at the Systems Biology Institute, Keio University, and Okinawa Institute of Science and Technology Graduate University. The authors would like to particularly thank to the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. This research was funded by the Exploratory Research for Advanced Technology (ERATO) program (Japan Science and Technology Agency).

### Author details

[1]JST ERATO Kawaoka infection-induced host responses project, Minato-ku, Tokyo 108-8639, Japan. [2]The Systems Biology Institute, Minato-ku, Tokyo 108-0071, Japan. [3]Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo 113-8540, Japan. [4]Department of Pathological Science, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53711, USA. [5]Laboratory of Veterinary Microbiology, Department of Veterinary Sciences, University of Miyazaki, Miyazaki 889-2192, Japan. [6]Sony Computer Science Laboratories, Inc., Shinagawa, Tokyo 141-0022, Japan. [7]Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa 904-0495, Japan. [8]Division of Virology, Department of Microbiology and Immunology, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. [9]Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. [10]Laboratory of Bioresponses Regulation, Department of Biological Responses, Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan.

Received: 1 July 2013 Accepted: 24 September 2013
Published: 2 October 2013

### References

1. Brass AL, Huang IC, Benita Y, John SP, Krishnan MN, Feeley EM, Ryan BJ, Weyer JL, van der Weyden L, Fikrig E, *et al*: **The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus.** *Cell* 2009, **139**(7):1243–1254.
2. Hao L, Sakurai A, Watanabe T, Sorensen E, Nidom CA, Newton MA, Ahlquist P, Kawaoka Y: **Drosophila RNAi screen identifies host genes important for influenza virus replication.** *Nature* 2008, **454**(7206):890–893.
3. Karlas A, Machuy N, Shin Y, Pleissner KP, Artarini A, Heuer D, Becker D, Khalil H, Ogilvie LA, Hess S, *et al*: **Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication.** *Nature* 2010, **463**(7282):818–822.
4. Konig R, Stertz S, Zhou Y, Inoue A, Hoffmann HH, Bhattacharyya S, Alamares JG, Tscherne DM, Ortigoza MB, Liang Y, *et al*: **Human host factors required for influenza virus replication.** *Nature* 2010, **463**(7282):813–817.
5. Shapira SD, Gat-Viks I, Shum BO, Dricot A, de Grace MM, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, *et al*: **A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection.** *Cell* 2009, **139**(7):1255–1267.
6. Sui B, Bamba D, Weng K, Ung H, Chang S, Van Dyke J, Goldblatt M, Duan R, Kinch MS, Li WB: **The use of random homozygous gene perturbation to identify novel host-oriented targets for influenza.** *Virology* 2009, **387**(2):473–481.
7. Watanabe T, Watanabe S, Kawaoka Y: **Cellular networks involved in the influenza virus life cycle.** *Cell Host Microbe* 2010, **7**(6):427–439.
8. Stertz S, Shaw ML: **Uncovering the global host cell requirements for influenza virus replication via RNAi screening.** *Microbes Infect* 2011, **13**(5):516–525.
9. Oda K, Matsuoka Y, Funahashi A, Kitano H: **A comprehensive pathway map of epidermal growth factor receptor signaling.** *Mol Syst Biol* 2005, **1**:2005-0010.
10. Oda K, Kitano H: **A comprehensive map of the toll-like receptor signaling network.** *Mol Syst Biol* 2006, **2**:2006-0015.
11. Li F, Thiele I, Jamshidi N, Palsson BØ: **Identification of potential pathway mediation targets in Toll-like receptor signaling.** *PLoS Comput Biol* 2009, **5**(2):e1000292–e1000292.
12. Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E: **A comprehensive modular map of molecular interactions in RB/E2F pathway.** *Mol Syst Biol* 2008, **4**:173–173.
13. Kaizu K, Ghosh S, Matsuoka Y, Moriya H, Shimizu-Yoshida Y, Kitano H: **A comprehensive molecular interaction map of the budding yeast cell cycle.** *Mol Syst Biol* 2010, **6**:415–415.
14. Caron E, Ghosh S, Matsuoka Y, Ashton-Beaucage D, Therrien M, Lemieux S, Perreault C, Roux PP, Kitano H: **A comprehensive map of the mTOR signaling network.** *Mol Syst Biol* 2010, **6**:453–453.
15. Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N, Suarez D, Pickett BE, Zhang Y, Larsen CN, *et al*: **Influenza research database: an integrated bioinformatics resource for influenza research and surveillance.** *Influenza Other Respi Viruses* 2012, **6**(6):404–416.
16. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D: **The influenza virus resource at the National Center for Biotechnology Information.** *J Virol* 2008, **82**(2):596–601.
17. Chatr-aryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, Sacco F, Tinti M, Smolyar A, Castagnoli L, Vidal M, *et al*: **VirusMINT: a viral protein interaction database.** *Nucleic Acids Res* 2009, **37**(Database issue):D669–D673.
18. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, *et al*: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res* 2005, **33**(Database issue):D428–D432.
19. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, *et al*: **Reactome knowledgebase of human biological pathways and processes.** *Nucleic Acids Res* 2009, **37**(Database issue):D619–D622.
20. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**(1):27–30.

21. Matsuoka Y, Ghosh S, Kikuchi N, Kitano H: **Payao: a community platform for SBML pathway model curation.** *Bioinformatics* 2010, **26**(10):1381–1383.
22. *iPathways+*. [http://www.ipathways.org/plus/]
23. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, *et al*: **The PANTHER database of protein families, subfamilies, functions and pathways.** *Nucleic Acids Res* 2005, **33**(suppl 1):D284–D288.
24. Hoffmann R, Valencia A: **A gene network for navigating the literature.** *Nat Genet* 2004, **36**(7):664–664.
25. Gorai T, Goto H, Noda T, Watanabe T, Kozuka-Hata H, Oyama M, Takano R, Neumann G, Watanabe S, Kawaoka Y: **F1Fo-ATPase, F-type proton-translocating ATPase, at the plasma membrane is critical for efficient influenza virus budding.** *Proc Natl Acad Sci USA* 2012, **109**(12):4615–4620.
26. Jorba N, Juarez S, Torreira E, Gastaminza P, Zamarreno N, Albar JP, Ortin J: **Analysis of the interaction of influenza virus polymerase complex with human cell factors.** *Proteomics* 2008, **8**(10):2077–2088.
27. Guan ZH, Zhang ML, Hou PL, Duan M, Cui YM, Wang XR: **Identification of cellular proteins interacting with influenza A virus PB1-F2 protein.** *Acta Virol* 2012, **56**(3):199–207.
28. Mayer D, Molawi K, Martinez-Sobrido L, Ghanem A, Thomas S, Baginsky S, Grossmann J, Garcia-Sastre A, Schwemmle M: **Identification of cellular interaction partners of the influenza virus ribonucleoprotein complex and polymerase complex using proteomic-based approaches.** *J Proteome Res* 2007, **6**(2):672–682.
29. Coombs KM, Berard A, Xu W, Krokhin O, Meng X, Cortens JP, Kobasa D, Wilkins J, Brown EG: **Quantitative proteomic analyses of influenza virus-infected cultured human lung cells.** *J Virol* 2010, **84**(20):10888–10906.
30. Shaw ML, Stone KL, Colangelo CM, Gulcicek EE, Palese P: **Cellular proteins in influenza virus particles.** *PLoS Pathog* 2008, **4**(6):e1000085.
31. CellDesigner: [http://celldesigner.org]
32. Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H: **Cell designer 3.5: a versatile modeling tool for biochemical networks.** *Proc IEEE* 2008, **96**(8):1254–1265.
33. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, *et al*: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.** *Bioinformatics* 2003, **19**(4):524–531.
34. Kitano H, Funahashi A, Matsuoka Y, Oda K: **Using process diagrams for the graphical representation of biological networks.** *Nat Biotechnol* 2005, **23**(8):961–966.
35. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, *et al*: **The systems biology graphical notation.** *Nat Biotechnol* 2009, **27**(8):735–741.
36. Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, *et al*: **Minimum information requested in the annotation of biochemical models (MIRIAM).** *Nat Biotechnol* 2005, **23**(12):1509–1515.
37. Connor RJ, Kawaoka Y, Webster RG, Paulson JC: **Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates.** *Virology* 1994, **205**(1):17–23.
38. Gambaryan AS, Tuzikov AB, Piskarev VE, Yamnikova SS, Lvov DK, Robertson JS, Bovin NV, Matrosovich MN: **Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-acetyllactosamine).** *Virology* 1997, **232**(2):345–350.
39. Ito T, Kawaoka Y: **Host-range barrier of influenza A viruses.** *Vet Microbiol* 2000, **74**(1–2):71–75.
40. Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, Donatelli I, Kawaoka Y: **Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals.** *J Virol* 2000, **74**(18):8502–8512.
41. Rogers GN, Pritchett TJ, Lane JL, Paulson JC: **Differential sensitivity of human, avian, and equine influenza A viruses to a glycoprotein inhibitor of infection: selection of receptor specific variants.** *Virolcgy* 1983, **131**(2):394–408.
42. Rogers GN, Paulson JC: **Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin.** *Virology* 1983, **127**(2):361–373.
43. Baum LG, Paulson JC: **Sialyloligosaccharides of the respiratory epithelium in the selection of human influenza virus receptor specificity.** *Acta Histochem Suppl* 1990, **40**:35–38.
44. Couceiro JN, Paulson JC, Baum LG: **Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity.** *Virus Res* 1993, **29**(2):155–165.
45. Matrosovich MN, Matrosovich TY, Gray T, Roberts NA, Klenk HD: **Human and avian influenza viruses target different cell types in cultures of human airway epithelium.** *Proc Natl Acad Sci USA* 2004, **101**(13):4620–4624.
46. Nicholls JM, Chan MC, Chan WY, Wong HK, Cheung CY, Kwong DL, Wong MP, Chui WH, Poon LL, Tsao SW, *et al*: **Tropism of avian influenza A (H5N1) in the upper and lower respiratory tract.** *Nat Med* 2007, **13**(2):147–149.
47. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y: **Avian flu: influenza virus receptors in the human airway.** *Nature* 2006, **440**(7083):435–436.
48. van Riel D, Munster VJ, de Wit E, Rimmelzwaan GF, Fouchier RA, Osterhaus AD, Kuiken T: **H5N1 virus attachment to lower respiratory tract.** *Science* 2006, **312**(5772):399.
49. Yao L, Korteweg C, Hsueh W, Gu J: **Avian influenza receptor expression in H5N1-infected and noninfected human tissues.** *FASEB J* 2008, **22**(3):733–740.
50. Pillai SP, Lee CW: **Species and age related differences in the type and distribution of influenza virus receptors in different tissues of chickens, ducks and turkeys.** *Virol J* 2010, **7**:5.
51. Nicholls JM, Chan RW, Russell RJ, Air GM, Peiris JS: **Evolving complexities of influenza virus and its receptors.** *Trends Microbiol* 2008, **16**(4):149–157.
52. Ito T, Suzuki Y, Suzuki T, Takada A, Horimoto T, Wells K, Kida H, Otsuki K, Kiso M, Ishida H, *et al*: **Recognition of N-glycolylneuraminic acid linked to galactose by the alpha2,3 linkage is associated with intestinal replication of influenza A virus in ducks.** *J Virol* 2000, **74**(19):9300–9305.
53. Wilks S, de Graaf M, Smith DJ, Burke DF: **A review of influenza haemagglutinin receptor binding as it relates to pandemic properties.** *Vaccine* 2012, **30**(29):4369–4376.
54. Viswanathan K, Chandrasekaran A, Srinivasan A, Raman R, Sasisekharan V, Sasisekharan R: **Glycans as receptors for influenza pathogenesis.** *Glycoconj J* 2010, **27**(6):561–570.
55. Wright PF, Neumann G, Kawaoka Y: **Orthomyxoviruses**. In *Fields Virology, vol. 2*. 5th edition. Edited by Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE. Philadelphia, Baltimore, New York, London, Buenos Aires, Hong Kong, Sydney, Tokyo: Wolters Kluwer; Lippincott Williams & Wilkins; 2007:1691–1740.
56. Matlin KS, Reggio H, Helenius A, Simons K: **Infectious entry pathway of influenza virus in a canine kidney cell line.** *J Cell Biol* 1981, **91**(3 Pt 1):601–613.
57. Nunes-Correia I, Eulalio A, Nir S, Pedroso de Lima MC: **Caveolae as an additional route for influenza virus endocytosis in MDCK cells.** *Cell Mol Biol Lett* 2004, **9**(1):47–60.
58. Sieczkarski SB, Whittaker GR: **Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis.** *J Virol* 2002, **76**(20):10455–10464.
59. de Vries E, Tscherne DM, Wienholts MJ, Cobos-Jimenez V, Scholte F, Garcia-Sastre A, Rottier PJ, de Haan CA: **Dissection of the influenza A virus endocytic routes reveals macropinocytosis as an alternative entry pathway.** *PLoS Pathog* 2011, **7**(3):e1001329.
60. Rossman JS, Leser GP, Lamb RA: **Filamentous influenza virus enters cells via macropinocytosis.** *J Virol* 2012, **86**(20):10950–10960.
61. Sieczkarski SB, Whittaker GR: **Differential requirements of Rab5 and Rab7 for endocytosis of influenza and other enveloped viruses.** *Traffic* 2003, **4**(5):333–343.
62. Feeley EM, Sims JS, John SP, Chin CR, Pertel T, Chen LM, Gaiha GD, Ryan BJ, Donis RO, Elledge SJ, *et al*: **IFITM3 inhibits influenza A virus infection by preventing cytosolic entry.** *PLoS Pathog* 2011, **7**(10):e1002337.
63. Skehel JJ, Wiley DC: **Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin.** *Annu Rev Biochem* 2000, **69**:531–569.
64. Martin K, Helenius A: **Transport of incoming influenza virus nucleocapsids into the nucleus.** *J Virol* 1991, **65**(1):232–244.
65. Bui M, Whittaker G, Helenius A: **Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins.** *J Virol* 1996, **70**(12):8391–8401.
66. Martin K, Helenius A: **Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import.** *Cell* 1991, **67**(1):117–130.
67. Whittaker G, Bui M, Helenius A: **Nuclear trafficking of influenza virus ribonuleoproteins in heterokaryons.** *J Virol* 1996, **70**(5):2743–2756.
68. Wu WW, Sun YH, Pante N: **Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein.** *Virol J* 2007, **4**:49.

69. O'Neill RE, Jaskunas R, Blobel G, Palese P, Moroianu J: **Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import.** *J Biol Chem* 1995, **270**(39):22701–22704.
70. O'Neill RE, Palese P: **NPI-1, the human homolog of SRP-1, interacts with influenza virus nucleoprotein.** *Virology* 1995, **206**(1):116–125.
71. Wang P, Palese P, O'Neill RE: **The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal.** *J Virol* 1997, **71**(3):1850–1856.
72. Cros JF, Garcia-Sastre A, Palese P: **An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein.** *Traffic* 2005, **6**(3):205–213.
73. Helenius A: **Unpacking the incoming influenza virus.** *Cell* 1992, **69**(4):577–578.
74. Resa-Infante P, Jorba N, Zamarreno N, Fernandez Y, Juarez S, Ortin J: **The host-dependent interaction of alpha-importins with influenza PB2 polymerase subunit is required for virus RNA replication.** *PLoS One* 2008, **3**(12):e3904.
75. Palese P, Shaw ML: ***Orthomyxoviridae*: The Viruses and Their Replication.** In *Fields Virology, vol. 2*. 5th edition. Edited by Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE. Philadelphia: Wolters Kluwer; Lippincott Williams & Wilkins; 2007:1647–1689.
76. Neumann G, Brownlee GG, Fodor E, Kawaoka Y: **Orthomyxovirus replication, transcription, and polyadenylation.** *Curr Top Microbiol Immunol* 2004, **283**:121–143.
77. Boivin S, Cusack S, Ruigrok RW, Hart DJ: **Influenza A virus polymerase: structural insights into replication and host adaptation mechanisms.** *J Biol Chem* 2010, **285**(37):28411–28417.
78. Nagata K, Kawaguchi A, Naito T: **Host factors for replication and transcription of the influenza virus genome.** *Rev Med Virol* 2008, **18**(4):247–260.
79. York A, Fodor E: **Biogenesis, assembly and export of viral messenger ribonucleoproteins in the influenza A virus infected cell.** *RNA biology* 2013, **10**(8):1274–1282.
80. Blaas D, Patzelt E, Kuechler E: **Identification of the cap binding protein of influenza virus.** *Nucleic Acids Res* 1982, **10**(15):4803–4812.
81. Blaas D, Patzelt E, Kuechler E: **Cap-recognizing protein of influenza virus.** *Virology* 1982, **116**(1):339–348.
82. Ulmanen I, Broni BA, Krug RM: **Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m7GpppNm) on RNAs and in initiating viral RNA transcription.** *Proc Natl Acad Sci USA* 1981, **78**(12):7355–7359.
83. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW: **The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit.** *Nature* 2009, **458**(7240):914–918.
84. Braam J, Ulmanen I, Krug RM: **Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription.** *Cell* 1983, **34**(2):609–618.
85. Li X, Palese P: **Characterization of the polyadenylation signal of influenza virus RNA.** *J Virol* 1994, **68**(2):1245–1249.
86. Luo GX, Luytjes W, Enami M, Palese P: **The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure.** *J Virol* 1991, **65**(6):2861–2867.
87. Robertson JS, Schubert M, Lazzarini RA: **Polyadenylation sites for influenza virus mRNA.** *J Virol* 1981, **38**(1):157–163.
88. Poon LL, Pritlove DC, Fodor E, Brownlee GG: **Direct evidence that the poly (A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template.** *J Virol* 1999, **73**(4):3473–3476.
89. Park YW, Katze MG: **Translational control by influenza virus. Identification of cis-acting sequences and trans-acting factors which may regulate selective viral mRNA translation.** *J Biol Chem* 1995, **270**(47):28433–28439.
90. Fortes P, Beloso A, Ortin J: **Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport.** *EMBO J* 1994, **13**(3):704–712.
91. Lu Y, Qian XY, Krug RM: **The influenza virus NS1 protein: a novel inhibitor of pre-mRNA splicing.** *Genes Dev* 1994, **8**(15):1817–1828.
92. Wang W, Krug RM: **U6atac snRNA, the highly divergent counterpart of U6 snRNA, is the specific target that mediates inhibition of AT-AC splicing by the influenza virus NS1 protein.** *Rna* 1998, **4**(1):55–64.
93. Momose F, Basler CF, O'Neill RE, Iwamatsu A, Palese P, Nagata K: **Cellular splicing factor RAF-2p48/NPI-5/BAT1/UAP56 interacts with the influenza virus nucleoprotein and enhances viral RNA synthesis.** *J Virol* 2001, **75**(4):1899–1908.
94. Momose F, Handa H, Nagata K: **Identification of host factors that regulate the influenza virus RNA polymerase activity.** *Biochimie* 1996, **78**(11–12):1103–1108.
95. Burgui I, Aragon T, Ortin J, Nieto A: **PABP1 and eIF4GI associate with influenza virus NS1 protein in viral mRNA translation initiation complexes.** *J Gen Virol* 2003, **84**(Pt 12):3263–3274.
96. Yanguez E, Castello A, Welnowska E, Carrasco L, Goodfellow I, Nieto A: **Functional impairment of eIF4A and eIF4G factors correlates with inhibition of influenza virus mRNA translation.** *Virology* 2011, **413**(1):93–102.
97. Yanguez E, Rodriguez P, Goodfellow I, Nieto A: **Influenza virus polymerase confers independence of the cellular cap-binding factor eIF4E for viral mRNA translation.** *Virology* 2011, **422**(2):297–307.
98. Bier K, York A, Fodor E: **Cellular cap-binding proteins associate with influenza virus mRNAs.** *J Gen Virol* 2011, **92**(Pt 7):1627–1634.
99. Katze MG, Chen YT, Krug RM: **Nuclear-cytoplasmic transport and VAI RNA-independent translation of influenza viral messenger RNAs in late adenovirus-infected cells.** *Cell* 1984, **37**(2):483–490.
100. Katze MG, DeCorato D, Krug RM: **Cellular mRNA translation is blocked at both initiation and elongation after infection by influenza virus or adenovirus.** *J Virol* 1986, **60**(3):1027–1039.
101. Katze MG, Detjen BM, Safer B, Krug RM: **Translational control by influenza virus: suppression of the kinase that phosphorylates the alpha subunit of initiation factor eIF-2 and selective translation of influenza viral mRNAs.** *Mol Cell Biol* 1986, **6**(5):1741–1750.
102. Chen Z, Li Y, Krug RM: **Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery.** *EMBO J* 1999, **18**(8):2273–2283.
103. Chen Z, Krug RM: **Selective nuclear export of viral mRNAs in influenza-virus-infected cells.** *Trends Microbiol* 2000, **8**(8):376–383.
104. Engelhardt OG, Fodor E: **Functional association between viral and cellular transcription during influenza virus infection.** *Rev Med Virol* 2006, **16**(5):329–345.
105. Engelhardt OG, Smith M, Fodor E: **Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II.** *J Virol* 2005, **79**(9):5812–5818.
106. Hutchinson EC, Fodor E: **Nuclear import of the influenza A virus transcriptional machinery.** *Vaccine* 2012, **30**(51):7353–7358.
107. Bullido R, Gómez-Puertas P, Albo C, Portela A: **Several protein regions contribute to determine the nuclear and cytoplasmic localization of the influenza A virus nucleoprotein.** *J Gen Virol* 2000, **81**(Pt 1):135–142.
108. Deng T, Sharps J, Fodor E, Brownlee GG: **In vitro assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex.** *J Virol* 2005, **79**(13):8669–8674.
109. Fodor E, Smith M: **The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex.** *J Virol* 2004, **78**(17):9144–9153.
110. Jones IM, Reay PA, Philpott KL: **Nuclear location of all three influenza polymerase proteins and a nuclear signal in polymerase PB2.** *EMBO J* 1986, **5**(9):2371–2376.
111. Mukaigawa J, Nayak DP: **Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2.** *J Virol* 1991, **65**(1):245–253.
112. Nath ST, Nayak DP: **Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1).** *Mol Cell Biol* 1990, **10**(8):4139–4145.
113. Neumann G, Castrucci MR, Kawaoka Y: **Nuclear import and export of influenza virus nucleoprotein.** *J Virol* 1997, **71**(12):9690–9700.
114. Nieto A, de la Luna S, Bárcena J, Portela A, Ortín J: **Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit.** *J Gen Virol* 1994, **75**(Pt 1):29–36.
115. Smith GL, Levin JZ, Palese P, Moss B: **Synthesis and cellular location of the ten influenza polypeptides individually expressed by recombinant vaccinia viruses.** *Virology* 1987, **160**(2):336–345.
116. Weber F, Kochs G, Gruber S, Haller O: **A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins.** *Virology* 1998, **250**(1):9–18.
117. Deng T, Engelhardt OG, Thomas B, Akoulitchev AV, Brownlee GG, Fodor E: **Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex.** *J Virol* 2006, **80**(24):11911–11919.