

RESEARCH

Open Access

Detecting virus-specific effects on post-infection temporal gene expression



Quan Chen^{1,2} and Jun Zhu^{1,2,3*}

From The 17th Asia Pacific Bioinformatics Conference (APBC 2019)
Wuhan, China. 14-16 January 2019

Abstract

Background: Different types of viruses have different envelope proteins, and may have their shared or distinctive host-virus interactions which result in various post-infection effects in humans and animals. These effects often do not appear at once but take time to unfold. To characterize the virus-specific effects, we applied a Multivariate Polynomial Time-dependent Genetic Association (MPTGA) method, previously proposed for detecting differences in temporal gene expression traits, to test for the differences in mouse lung transcriptome response to infection of different subtypes of influenza A viruses.

Results: We compared two methods: the Multivariate Polynomial Time-dependent Genetic Association (MPTGA) method, and the conventional modified t-test, to study the virus-specific effects on mouse lung gene expression. Both methods found H3N2 to be the most different virus among the three viruses tested, with the largest number of genes with H3N2-specific effects. However, the MPTGA method demonstrated much higher power of detection, and the detected genes with virus-specific effects showed better biological relevance.

Conclusions: Transcriptome response to virus infection is dynamic. MPTGA which leverages temporal gene expression traits showed increased power in detecting biologically relevant virus-specific effects comparing with conventional t-test method.

Keywords: Temporal association, Dynamic response, Flu virus

Background

Influenza viruses have various types and subtypes that differ in their morbidity, virulence and pathogenicity. Influenza A virus is the most virulent type and can be further divided into several subtypes [1]. These subtypes also include different strains which may evolve over time and vary in their manifestations: zoonotic, pandemic and seasonal [2]. The highly pathogenic avian influenza H5N1 virus, for example, is a zoonotic influenza virus that is transmitted from poultry to humans. The H5N1 virus strain A/Vietnam/1203/04 was the most pathogenic strain from the 2004 outbreaks of H5N1

influenza viruses [3]. The pandemic H1N1/09 virus strain, on the other hand, has led to the swine-origin H1N1 flu pandemic in 2009. The seasonal H3N2 vaccine seed strain A/Wyoming/03/2003 represents H3N2 viruses isolated during the last three seasons [4].

Previous studies [5] have shown in time course experiments that seasonal H3N2, pandemic H1N1, and zoonotic H5N1 viruses differ strongly in the spatial and temporal dynamics of infection and associated lesions in the ferret respiratory tract. Therefore, to characterize and to differentiate the infection processes with the diverse categories and subtypes of influenza viruses, it is important to acknowledge its highly dynamic nature.

In this paper, we investigated temporal mouse lung gene expression traits in response to infection with influenza A/California/04/09 (H1N1), A/Wyoming/03/03 (H3N2), and A/Vietnam/1203/04 (H5N1) HA₀ virus (GSE98527) and aimed to identify the virus-specific effects. To

*Correspondence: jun.zhu@mssm.edu

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 10029 New York, NY, USA

²Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, 10029 New York, NY, USA

Full list of author information is available at the end of the article



compare temporal gene expression traits, we proposed a multivariate polynomial temporal genetic association (MPTGA) method [6] previously in the context of genetic association with temporal gene expression. We further extended this method to a more general setting that compares the temporal gene expression trait under different conditions, which applies to the specific aim in this study—to detect the differences in the way each virus affects temporal gene expression traits over time. We also compared the MPTGA method with the conventional modified t-test to detect differential gene expression post infection of different viruses, and demonstrated significant increase in the power and specificity of our temporal method.

Results

Mouse lung dataset

To characterize the cellular transcriptome response of mouse lung to infection of H1N1, H3N2, and H5N1 influenza virus, six to eight week-old female C57BL/6 mice were infected with influenza A/California/04/09 (H1N1), A/Wyoming/03/03 (H3N2), and A/Vietnam/1203/04 (H5N1) HALo virus in 3 experimental groups respectively. The Influenza A/Vietnam/1203/04 (H5N1) HALo mutant virus is an attenuated H5N1 virus generated from wild-type Influenza A/Vietnam/1203/04 (H5N1) virus [7]. Lungs were collected from each experimental group with 3 mice infected with the same virus at 12h, 1d, 2d, 3d and 4d post infection. The cellular transcriptome at each time point was profiled by mRNA-Seq technology (GSE98527).

Detecting virus-specific effects on post-infection temporal expression

We applied MPTGA to all 24,421 genes profiled in the mouse lung dataset, testing for differential temporal post-infection gene expression between infection with each virus and the other two. In addition, we applied the moderated t-test using an empirical Bayes method implemented in the R package *Limma* [8] to test for differential expression based on the last time point.

Genes have different temporal trajectories in response to different subtypes of viruses. For example, Fig. 1 shows expression of *Mill2* (MHC I like leukocyte 2), an mouse ortholog for the *MICB* gene in human. *Mill2* is highly expressed in mouse lung according to Mouse ENCODE transcriptome data [9]. It involves in immunoregulatory interactions between a lymphoid and a non-lymphoid cell and innate immune system. Its related GO processes include immune response and immune response-activating cell surface receptor signaling pathway. Figures 1 and 2 illustrate the performance of different methods in detecting virus-specific effects on post-infection temporal expression of *Mill2*. Previous

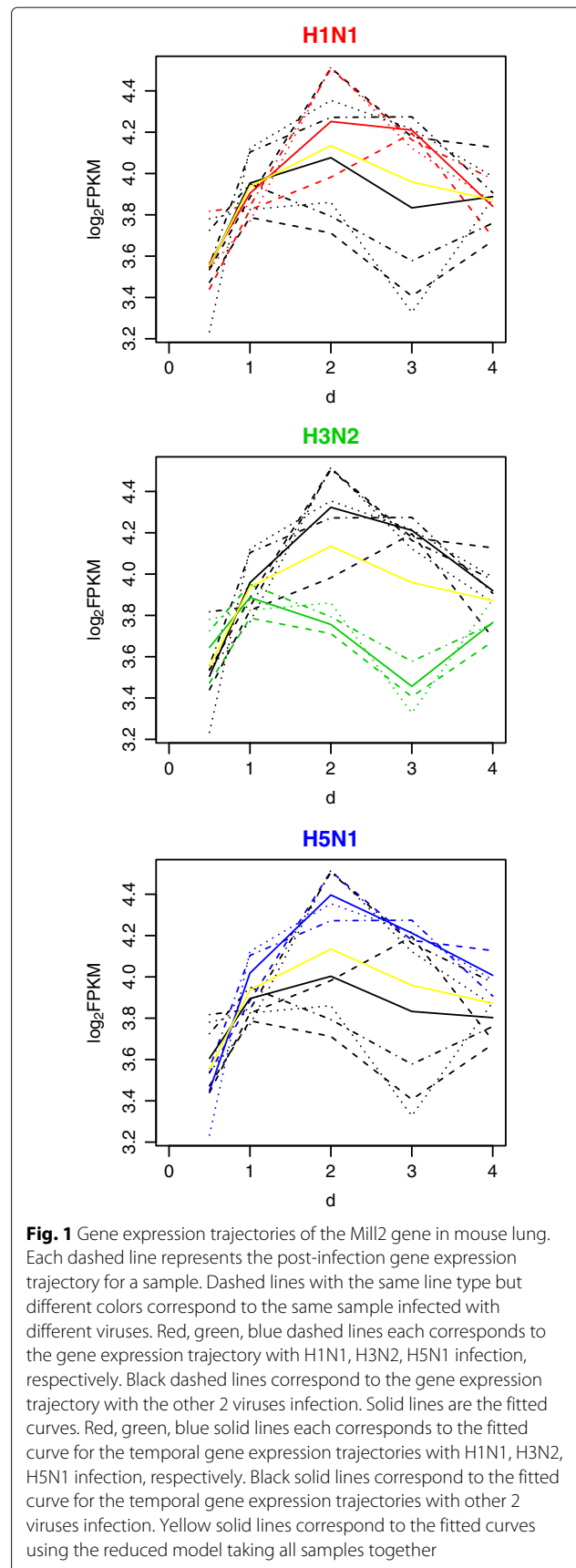
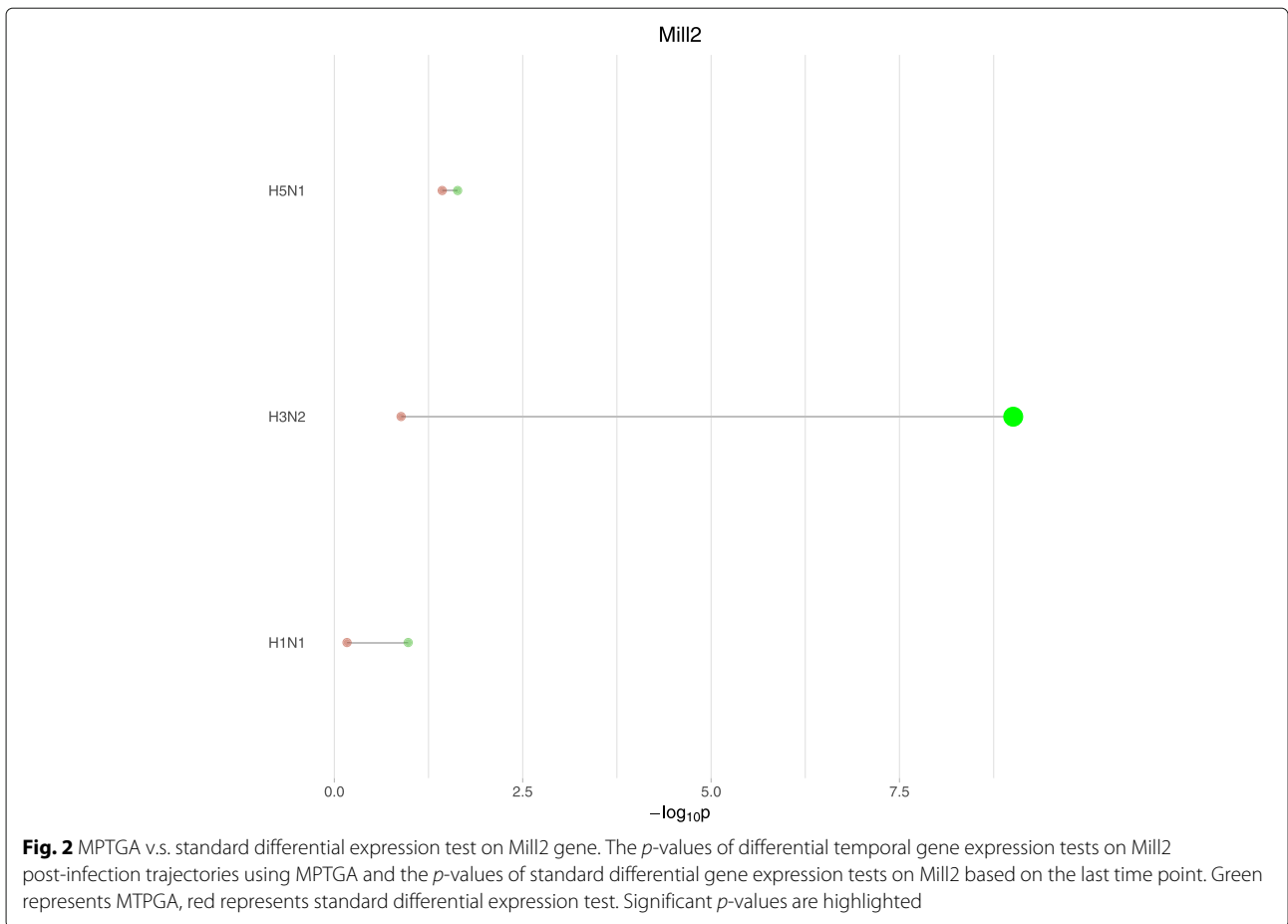
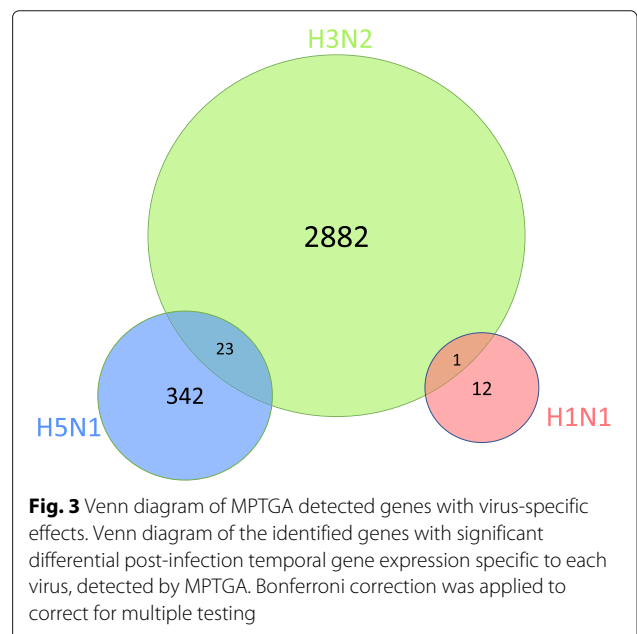


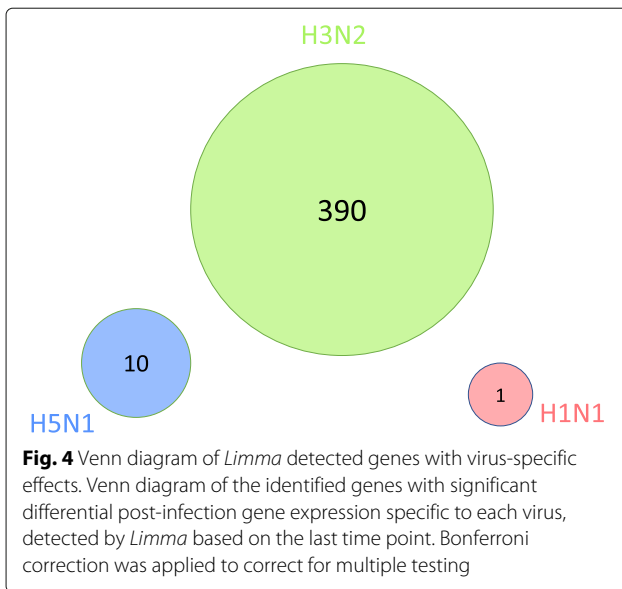
Fig. 1 Gene expression trajectories of the *Mill2* gene in mouse lung. Each dashed line represents the post-infection gene expression trajectory for a sample. Dashed lines with the same line type but different colors correspond to the same sample infected with different viruses. Red, green, blue dashed lines each corresponds to the gene expression trajectory with H1N1, H3N2, H5N1 infection, respectively. Black dashed lines correspond to the gene expression trajectory with the other 2 viruses infection. Solid lines are the fitted curves. Red, green, blue solid lines each corresponds to the fitted curve for the temporal gene expression trajectories with H1N1, H3N2, H5N1 infection, respectively. Black solid lines correspond to the fitted curve for the temporal gene expression trajectories with other 2 viruses infection. Yellow solid lines correspond to the fitted curves using the reduced model taking all samples together



studies [10–14] have investigated the immune-related functions of Mill2 (or its orthologs) and have characterized the underlying mechanisms. MICB activates the cytolytic response of natural killer (NK) cells and CD8 T-cells. As is shown in Fig. 1 with H3N2 infection, Mill2 expression trajectories demonstrate an evidently distinct pattern from the trajectories post-infection with the other two viruses, consistent with previous observations that H3N2 is less effective in induction of immune response in lung [15]. However, at the last time point, all trajectories return to similar levels of gene expression values, which suggests that the standard differential expression test based on the last time point was not able to detect the difference, whereas the MPTGA method was able to detect the subtype specific effect of post-infection temporal expression of Mill2, as shown in Fig. 2.

Figure 3 provides a summary of the genes identified by MPTGA with differential post-infection temporal gene expression specific to each virus. Figure 4 provides a summary of the genes identified by the conventional differential gene expression test implemented by the R package *Limma* [8] as a moderated t-test based on the





last time point. MPTGA identified more temporal differential expression genes than the conventional test, with 3260 genes showing differential temporal gene expression specific to at least one virus, while only 401 genes were identified by the conventional method. In addition, the conventional test results in three disjoint sets of genes with differential gene expression specific to each virus, suggesting that the test is only able to distinguish the post-infection effect on the gene expression between at most one of the viruses and the others. On the contrary, MPTGA identified 24 genes with differential temporal expression specific to multiple viruses. These results demonstrate significant superiority of the MPTGA method over the conventional method in both the power of detection and the specificity of virus-specific effects.

In spite of the differences, both MPTGA and the conventional test resulted in the largest number of differentially expressed genes specific to H3N2. The same observation also holds for the genes uniquely specific to each virus, with H3N2 having the largest number of genes that are only found to take on a post-infection effect specific to H3N2 but not to H5N1 nor H1N1. On the contrary, H1N1 appears to be the least specific virus, with the smallest number of a temporal effect specific to H1N1.

Functional enrichment

We looked into the functional enrichment in the identified differentially expressed genes specific to H3N2 only (without being specific to H1N1 or H5N1), as H3N2 was shown to be the virus with the most specific post-infection effects by both the MPTGA and the conventional method. The MPTGA method identified 2882 genes of only H3N2-specific effects, whereas the conventional method identified 390 genes of only H3N2-specific effects. Among

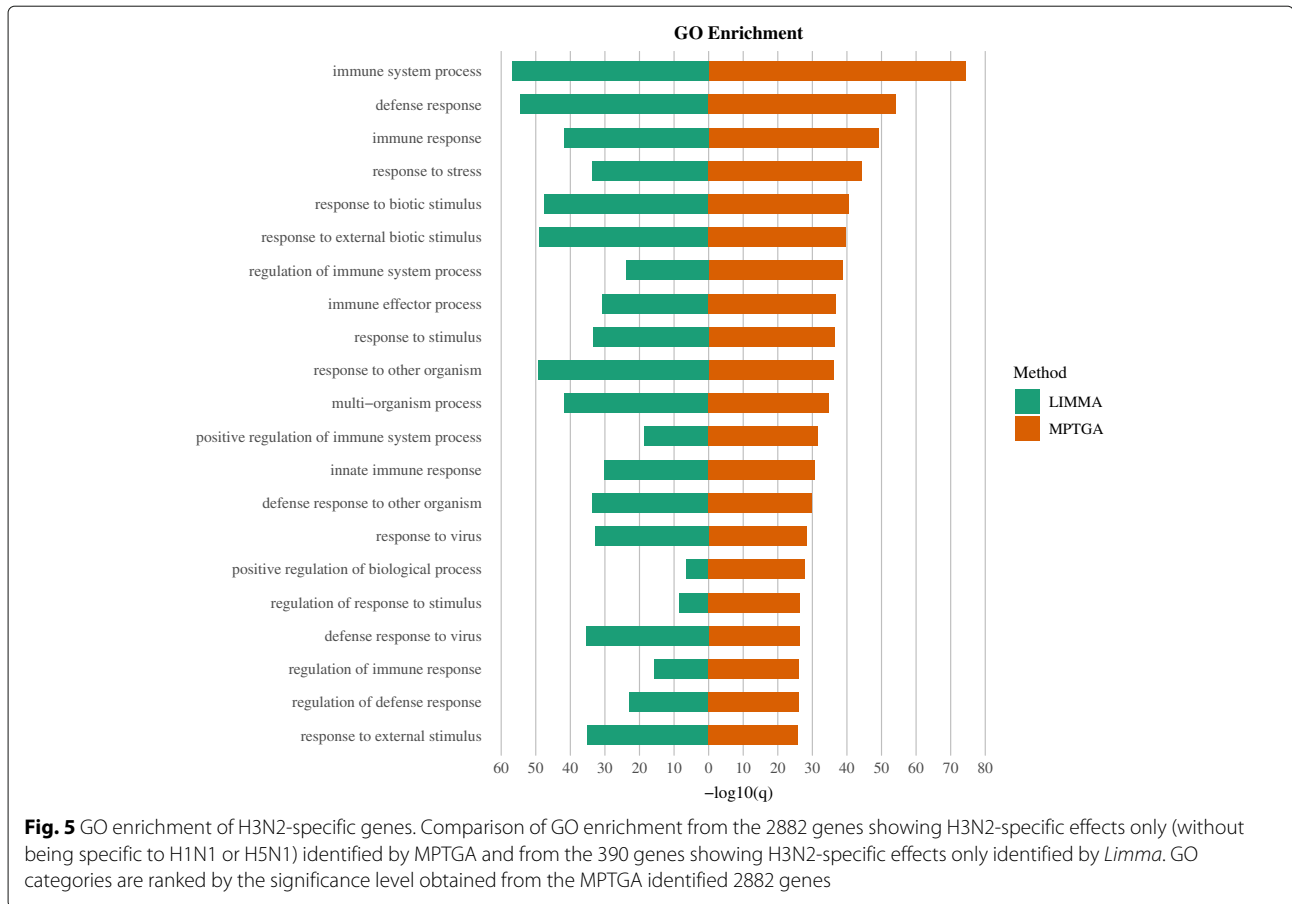
them, 381 (98%) genes were in the common between the two. We compare the significance level of the top enriched GO categories for the 2882 genes identified by MPTGA with the significance level of these GO categories for the 390 genes identified by the conventional method in Fig. 5. The majority of the top enriched GO terms from the MPTGA identified 2882 genes of only H3N2-specific effects were related to immune or defense mechanisms. Comparing with *Limma*, the MPTGA method shows a consistently stronger enrichment in the immune related categories. We also look into the 9 genes identified by *Limma* but not by MPTGA and find no function enrichment in any immune or defense related GO categories. Next, we show the GO enrichment of the 2501 genes among the set of 2882 genes identified by MPTGA but not by *Limma* in Fig. 6. Most of the top enriched GO terms are related to immune/defense mechanisms or cell cycle regulation.

Previous studies [16–22] have compared the pathogenesis of different viruses in human cases and animal models. The H5N1 and H1N1 subtype viruses are more pathogenic and have a larger effect on lower respiratory tract comparing the H3N2 subtype virus. The pandemic 2009 H1N1 virus is shown to have similar pathogenesis to other influenza A viruses with high virulence such as the H5N1 viruses, with extension of its inflammation process of the larger airways into the alveoli causing diffuse alveolar damage, while the less virulent seasonal H3N2 virus is found to have less extension of its inflammatory process to alveoli. The difference in the morbidity and pathogenesis observed in these viruses is consistent with what we find in comparing the effect of each of the viruses on the temporal gene expression after infection, and finding the most number of genes showing specific effects unique to H3N2. Our results showed that H3N2 specific effects in lung were related to immune response and cell cycle regulation, indicating that the MPTGA method did not only detect more genes with virus-specific effects, but also revealed significant biological pathways underlying virus-specific effects.

It has been shown that innate lymphoid cells (ILCs) accumulated in lung after virus infection to promote lung tissue homeostasis [23]. Monticelli et al. [23] reports a ILC enriched signature consisting of 121 genes. We compared the signature with our virus specific genes and identified 41, 1, and 0 genes in the overlap with H3N2, H1N1, and H5N1 specific genes, respectively. The significant overlap with H3N2 specific genes (Fisher's Exact test p -value = $5.0e-10$) suggests a potential mechanism explaining why H3N2 is least pathogenic to lung tissues.

Discussion

Virus infection in humans and animals is a dynamic process, with wildly different virulence and pathology specific

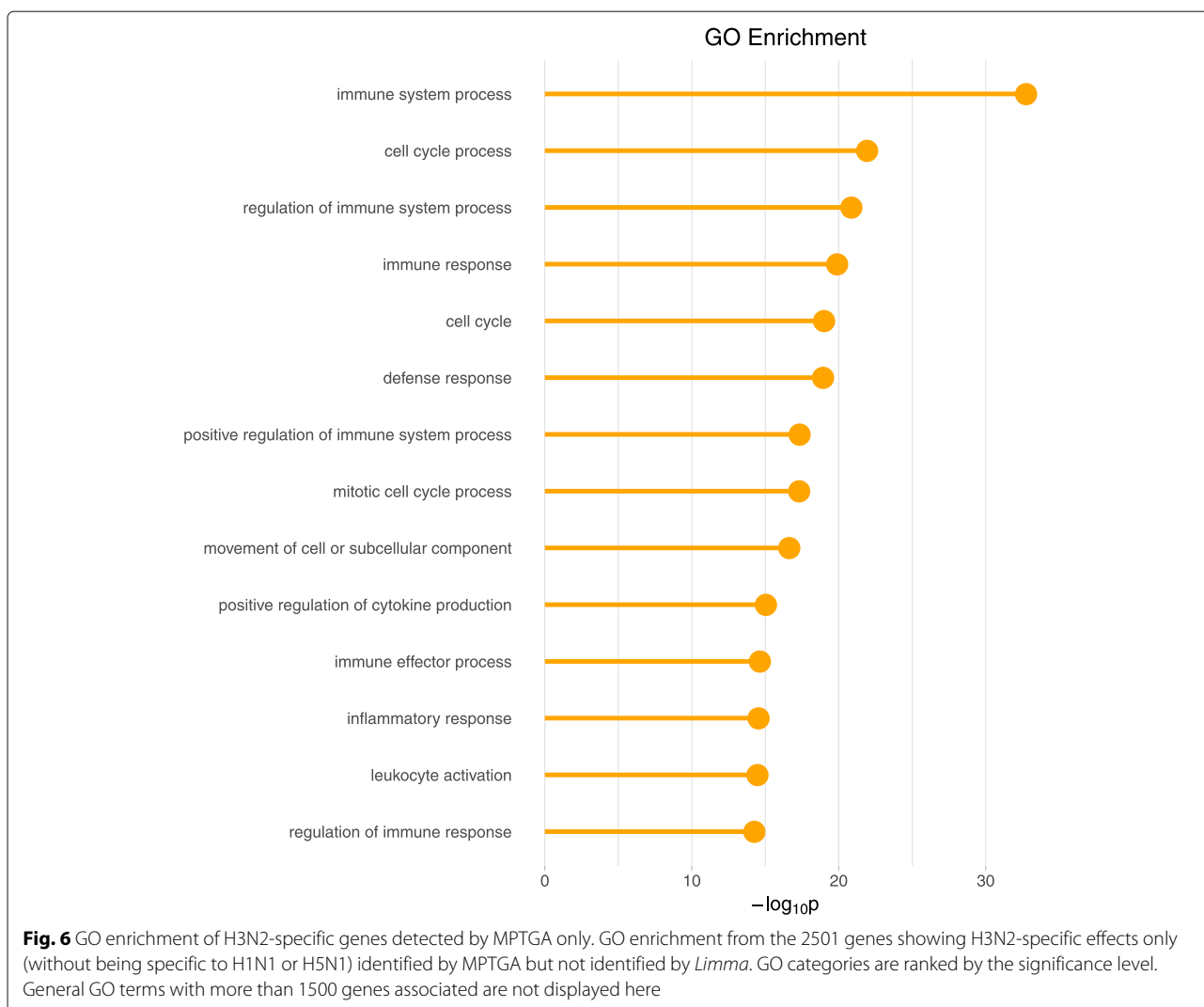


to types or subtypes of viruses. For health-care professionals, it is essential to understand the virus-specific effects over a temporal dimension in order to accurately identify the virus and to proceed with appropriate treatment strategies.

In this paper, we studied the temporal gene expression response in mouse lung to infection with a seasonal H3N2 virus strain, a pandemic H1N1 virus strain, and a zoonotic H5N1 virus strain. We aimed to identify the virus-specific effects on the post-infection gene expression. For this purpose, we compared two methods, a Multivariate Polynomial Time-dependent Genetic Association (MPTGA) method proposed for temporal gene expression traits, and the conventional modified t-test for differential gene expression detection. Both methods showed that the seasonal H3N2 virus was the most different among the three tested, showing distinct response in the expression of the largest number of genes. The MPTGA method identified significantly more genes with virus-specific effects for each of the three viruses comparing with the modified t-test. We looked into specific examples and observed that the effects of the viruses infection varied over time, which explained why the modified t-test failed to capture the

post-infection effects for many of the genes based on a single static time point. Moreover, by comparing the genes with virus-specific effects identified by the two methods, we showed that the MPTGA method identified more genes involved in immune response and cell cycle regulation pathways, consistent with previous studies comparing patho-physiological effects of these virus subtypes. Nevertheless, what virus-host interactions drive the virus specific effects can not be derived from this study. Further studies are needed to elucidate causal regulations leading to the virus specific effects.

We also generalized the MPTGA model based on two genotypes/groups to three or more genotypes/groups (the MPTGA models based on 2 or 3 genotypes/groups are noted as MPTGA2 and MPTGA3, respectively). We applied MPTGA3 to the mouse lung data set and compared results based on MPTGA2 and MPTGA3 (Fig. 7). MPTGA3 identified more genes differentially regulated among 3 subtypes of viruses. There were also genes identified by MPTGA2 but not by MPTGA3, suggesting the model selection should be based on underlying data distributions. For results based on MPTGA2, it is easier to interpret common of viruses and unique to



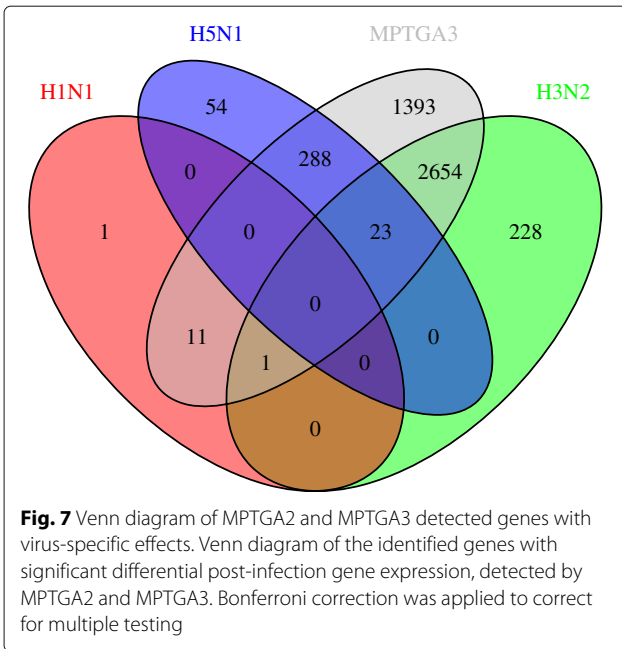
each specific virus. However, it is not easy to biologically interpret the results based on MPTGA3. Thus, we focused on the results based on MPTGA2 in the “Results” section.

In the MPTGA model, we assumed expression levels of a gene followed a group mean trajectory with variance at each time following a normal distribution. To assess the performance of MPTGA in the cases where variances deviate from a normal distribution, we simulated data sets based on the empirical patterns identified in the mouse lung data and added residuals sampling from different distributions (normal, uniform, exponential distributions, and Student’s t-distribution with 3, 5, and 7 degrees). Then, we applied MPTGA to the simulated data sets. The accuracy were greater than 99% for T3, T5, and exponential distributions and 100% for other distributions tested. The results are summarized in Fig. 8. In general, MPTGA was robust to residuals deviated from a normal distribution.

It is worth to note that there were lots of assumptions made in the MPTGA model, such as cubic polynomial function for mean trajectories and covariances related in first order auto correlation. The choices were made mainly due to limited number of time points in a time series data set. If a long time series is available, high degree polynomial functions and higher order correlation structure should be explored.

Conclusions

Transcriptome response to virus infection may vary between viruses and over time. The Multivariate Polynomial Time-dependent Genetic Association (MPTGA) method can be applied to detect virus-specific effects on temporal gene expression traits. It is shown to enhance the power and specificity of detection subtype specific effects, which could result a more accurate target gene set for understanding the virus pathology.



Methods

MPTGA and extension

Lin et al. [6] proposed a Multivariate Polynomial Time-dependent Genetic Association (MPTGA) method to detect genetic effects in temporal gene expression trajectories. The MPTGA method assumes that for each genotype j , the temporal gene expression trait \mathbf{y} across m time points follows a multivariate normal distribution $\mathcal{N}(\mathbf{g}_j, \Sigma)$, with density function

$$f_j(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp \left[-(\mathbf{y} - \mathbf{g}_j) \Sigma^{-1} (\mathbf{y} - \mathbf{g}_j)^T / 2 \right].$$

The mean vector \mathbf{g}_j for genotype j is modeled with a polynomial curve

$$\mathbf{g}_j = [\mathbf{g}_j(t)]_{1 \times m} = \left[\sum_{k=0}^K \beta_{kj} t^k \right]_{1 \times m},$$

where K is the degree of the polynomial function. In this study, K was set to be 3.

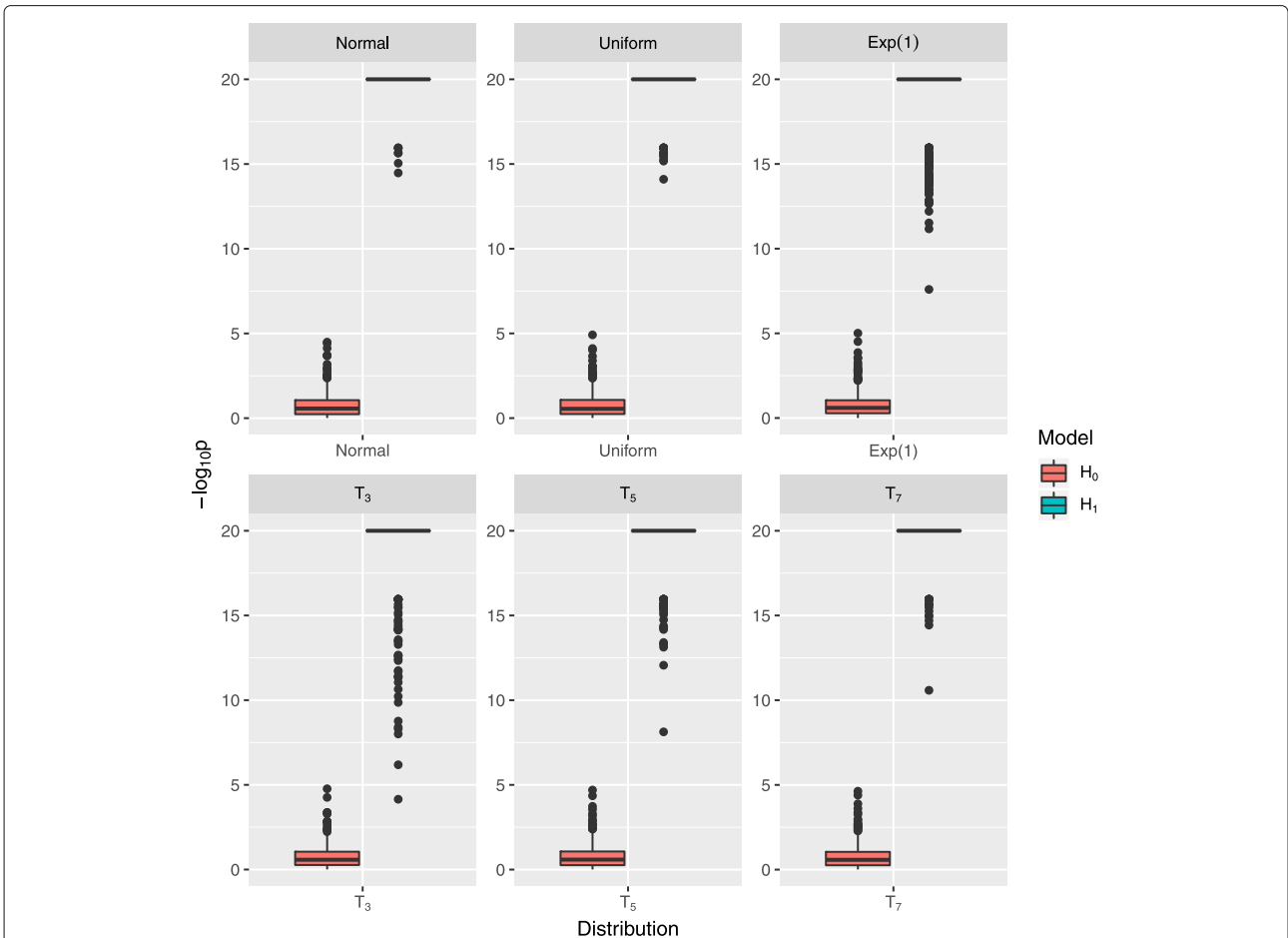


Fig. 8 Evaluation of MPTGA performance robustness to deviations from a normal distribution. Summary of the MPTGA test performance when the underlying assumption of normal distribution is met or when the expression traits follow different distributions (normal, uniform, exponential, Student's t with 3,5, and 7 degrees)

In particular, MPTGA assumes a first order auto-regression model to take into account the auto-correlation between different time points, with the covariance matrix specified as

$$\Sigma = \sigma_e^2 \begin{bmatrix} 1 & \rho & \dots & \rho^{m-1} \\ \rho & 1 & \dots & \rho^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \dots & 1 \end{bmatrix}.$$

For each sample i , its gene expression trajectory $\mathbf{y}_i(t)$ across m time points can be written as

$$\mathbf{y}_i(t) = \delta_{i0} \sum_{k=0}^K \beta_{k0} t^k + \delta_{i1} \sum_{k=0}^K \beta_{k1} t^k + \epsilon_i,$$

where δ_{i0} and δ_{i1} are the indicator variables of sample i taking genotype 0 or 1.

Given the joint likelihood

$$\mathcal{L}(\Theta) = \prod_{i=1}^N [\delta_{i0} f_0(\mathbf{y}_i) + \delta_{i1} f_1(\mathbf{y}_i)]$$

The maximum likelihood estimates of the parameter set $\Theta = (\beta_{kj}, \rho, \sigma_e^2)$ can be derived as described in the Lin et al. [6] paper.

Lin et al. [6] compared longitudinal trajectories between two different conditions (e.g. two possible genotypes in a haploid system). The method can be extended to accommodate three or more conditions (e.g. effects of three possible genotypes in a diploid system; e.g. effects of different viruses infection). This comparison can be done using done using a pairwise test between conditions. Alternatively, we can construct a full model to detect the (genetic) effects:

$$y_i(t) = \delta_{i0} \sum_{k=0}^K \beta_{k0} t^k + \delta_{i1} \sum_{k=0}^K \beta_{k1} t^k + \delta_{i2} \sum_{k=0}^K \beta_{k2} t^k + \epsilon_i$$

for a given trait Y , then the reduced model H_0 (single gene expression trait curve)

$$H_0 : \beta_{k0} = \beta_{k1} = \beta_{k2} \quad \text{for all } k$$

can be compared against the full model H_1 (different gene expression trait curve for different conditions (genotypes/viruses infection)):

$$H_1 : \text{at least one of the equalities does not hold}$$

to test the hypothesis of the existence of eQTL at a locus or difference between effects of viruses by estimating these parameters with a MLE procedure and performing a likelihood ratio test. The full model approach differs from the pairwise test in that it imposes an extra condition of a shared covariance structure among all three conditions, which is estimated in the full model taking all three conditions into account; whereas the pairwise test only requires

the covariance structure to be shared between each pair, which is estimated separately for each pair.

Similar to the MPTGA model with two genotypes, we assume that for each genotype $j = 0, 1, 2$, the mean vector \mathbf{g}_j for genotype j is modeled with a polynomial curve

$$\mathbf{g}_j = [\mathbf{g}_j(t)]_{1 \times m} = \left[\sum_{k=0}^K \beta_{kj} t^k \right]_{1 \times m} \quad j = 0, 1, 2.$$

Given the joint log-likelihood

$$\begin{aligned} \log \mathcal{L}(\Theta) &= \sum_{i=1}^N \delta_{i0} \log f_0(\mathbf{y}_i) + \sum_{i=1}^N \delta_{i1} \log f_1(\mathbf{y}_i) \\ &+ \sum_{i=1}^N \delta_{i2} \log f_2(\mathbf{y}_i), \end{aligned}$$

the maximum likelihood estimates of the parameter set $\Theta = (\beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}, \rho, \sigma_e^2), j = 0, 1, 2$ can be derived by first looking for the critical point of $\log \mathcal{L}(\Theta)$ by taking its derivative with respect to β 's and σ_e^2 , finding that both β 's, σ_e^2 can be expressed as functions of ρ at the critical point, therefore $\log \mathcal{L}(\Theta)$ can be expressed as functions of ρ at the critical point as well; then the MLE of ρ can be derived by taking derivative of $\log \mathcal{L}(\Theta)$ with respect to ρ , and thus β 's and σ_e^2 can be obtained accordingly.

The detailed derivation is as follows. Define $\mathbf{T}_j = \sum_{i=1}^N \delta_{ij} \mathbf{y}_i, j = 0, 1, 2; \mathbf{I}_0 = [1 \dots 1]_{1 \times m}, \mathbf{I}_1 = [1 \dots m], \mathbf{I}_2 = \mathbf{I}_1^2 = [1 \dots m^2], \mathbf{I}_3 = \mathbf{I}_1^3 = [1 \dots m^3]; Q(\rho, \mathbf{U}, \mathbf{V}) = \frac{1}{1-\rho^2} (\mathbf{U}_1 \mathbf{V}_1 + \mathbf{U}_m \mathbf{V}_m) - \frac{\rho}{1-\rho^2} \left[\sum_{i=1}^{m-1} \mathbf{U}_i \mathbf{V}_{i+1} + \mathbf{U}_{i+1} \mathbf{V}_i \right] + \frac{1+\rho^2}{1-\rho^2} \sum_{i=2}^{m-1} \mathbf{U}_i \mathbf{V}_i$, where $\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_m]$ and $\mathbf{V} = [\mathbf{V}_1, \dots, \mathbf{V}_m]$.

Taking derivative of $\log \mathcal{L}(\Theta)$ with respect to β_0 's gives the following linear system:

$$\begin{bmatrix} \alpha_{11} \beta_{00} + \alpha_{12} \beta_{10} + \alpha_{13} \beta_{20} + \alpha_{14} \beta_{30} = b_1 \\ \alpha_{21} \beta_{00} + \alpha_{22} \beta_{10} + \alpha_{23} \beta_{20} + \alpha_{24} \beta_{30} = b_2 \\ \alpha_{31} \beta_{00} + \alpha_{32} \beta_{10} + \alpha_{33} \beta_{20} + \alpha_{34} \beta_{30} = b_3 \\ \alpha_{41} \beta_{00} + \alpha_{42} \beta_{10} + \alpha_{43} \beta_{20} + \alpha_{44} \beta_{30} = b_4 \end{bmatrix},$$

where $\alpha_{ij} = n_0 Q(\rho, \mathbf{I}_{i-1}, \mathbf{I}_{j-1}), b_i = Q(\rho, \mathbf{T}_0, \mathbf{I}_{i-1}), n_0 = \sum_{i=1}^N \delta_{i0}$. Then β_0 's can be estimated by

$$\begin{bmatrix} \beta_{00} \\ \beta_{10} \\ \beta_{20} \\ \beta_{30} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix}^{-1} * \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}.$$

Similarly, define $n_1 = \sum_{i=1}^N \delta_{i1}, \alpha_{ij}^{(1)} = n_1 Q(\rho, \mathbf{I}_{i-1}, \mathbf{I}_{j-1}), b_i^{(1)} = Q(\rho, \mathbf{T}_1, \mathbf{I}_{i-1}); n_2 = \sum_{i=1}^N \delta_{i2}, \alpha_{ij}^{(2)} = n_2 Q(\rho, \mathbf{I}_{i-1}, \mathbf{I}_{j-1}), b_i^{(2)} = Q(\rho, \mathbf{T}_2, \mathbf{I}_{i-1});$ then β 's can be derived accordingly.

Next, taking derivative with respect to σ_e^2 gives

$$\sigma_e^2 = \left\{ \sum_{i=1}^N Q(\rho, \mathbf{y}_i, \mathbf{y}_i) + \sum_{i=0}^2 [Q(\rho, \mathbf{T}_i, \mathbf{T}_i) - \sum_{j=0}^3 2\beta_{ji} Q(\rho, \mathbf{T}_i, \mathbf{I}_j) + \sum_{j=0}^3 \sum_{k=0}^3 n_i \beta_{ji} \beta_{ki} Q(\rho, \mathbf{I}_j, \mathbf{I}_k)] \right\} / (mN)$$

With the previously derived estimators of β 's and σ_e^2 also as functions of ρ , the log-likelihood can now be written as

$$\log \mathcal{L}(\Theta) = -\frac{mN}{2} \log 2\pi - \frac{N}{2} [(m-1) \log(1-\rho^2) + m \log \sigma_e^2] - \frac{mN}{2}$$

Based on the MLE of the unknown parameters, a LRT ($-2 \log \Lambda \sim \chi_8^2$) can be conducted comparing the aforementioned full model H_1 with the null model H_0 to test for virus-specific effects.

Detecting virus-specific effects on post-infection temporal expression

The MPTGA method can be applied to a broader setting that compares the temporal gene expression traits between two different conditions, whereas the original setting proposed by Lin et al. [6] is a special case that compares two possible genotypes in a haploid system. In this paper, we are interested in detecting the virus-specific effects on the post-infection temporal gene expression trajectories. For a given virus ν , we define $j = 0, 1$ as the indicator variable of a sample infected by virus ν or other viruses. The null hypothesis H_0 assumes that the temporal gene expression trait \mathbf{y}_j share the same trajectory pattern after infected by virus ν and other viruses:

$$H_0 : \beta_{k0} = \beta_{k1} \quad \text{for all } k$$

which is tested against the full model H_1 , assuming that the temporal post-infection gene expression trait \mathbf{y}_j with virus ν has a unique trajectory pattern distinct from the post-infection trajectories with other viruses:

$$H_1 : \text{at least one of the equalities does not hold}$$

The hypothesis can be tested using a likelihood ratio test.

Similarly, MPTGA can be applied to cases with three or more groups as outlined above.

Abbreviations

ENCODE: Encyclopedia of DNA elements; GEO: Gene expression omnibus; GO: Gene ontology; ILC: Innate lymphoid cell; LRT: Likelihood ratio test; MLE: Maximum likelihood estimation; MPTGA: Multivariate polynomial time-dependent genetic association; NK: Natural killer

Acknowledgements

We thank members of Zhu's lab and Dr. García-Sastre for productive discussion.

Funding

This work was partially supported by the U19 AI118610 and U01 HG008451 grants. Publication of this article was supported by NIH-U19 AI118610.

Availability of data and materials

The data used in study is from the public database GEO with accession number GSE98527.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 3, 2019: Selected articles from the 17th Asia Pacific Bioinformatics Conference (APBC 2019): bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-3>.

Authors' contributions

QC and JZ designed the study, QC performed data analysis. QC and JZ wrote the manuscript. Both authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 10029 New York, NY, USA. ²Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, 10029 New York, NY, USA. ³Sema4, a Mount Sinai venture, 06902 Stamford, CT, USA.

Published: 29 March 2019

References

- Hay AJ, Gregory V, Douglas AR, Lin YP. The evolution of human influenza viruses. *Philos Trans R Soc Lond Ser B*. 2001;356(1416):1861–70. <https://doi.org/10.1098/rstb.2001.0999>.
- Reperant LA, Moesker FM, Osterhaus ADME. Influenza: from zoonosis to pandemic. *ERJ Open Research*. 2016;2(1). <https://doi.org/10.1183/23120541.00013-2016>.
- Govorkova EA, Rehg JE, Krauss S, Yen H-L, Guan Y, Peiris M, Nguyen TD, Hanh TH, Puthavathana P, Long HT, Buranathai C, Lim W, Webster RG, Hoffmann E. Lethality to Ferrets of H5n1 Influenza Viruses Isolated from Humans and Poultry in 2004. *J Virol*. 2005;79(4):2191–8. <https://doi.org/10.1128/JVI.79.4.2191-2198.2005>.
- Schweiger B, Bruns L, Meixnerberger K. Reassortment between human A(H3N2) viruses is an important evolutionary mechanism. *Vaccine*. 2006;24(44-46):6683–90. <https://doi.org/10.1016/j.vaccine.2006.05.105>.
- Brand JMAvd, Stittelaar KJ, Amerongen Gv, Reperant L, Waal Ld, Osterhaus ADME, Kuiken T. Comparison of Temporal and Spatial Dynamics of Seasonal H3N2, Pandemic H1N1 and Highly Pathogenic Avian Influenza H5N1 Virus Infections in Ferrets. *PLoS ONE*. 2012;7(8):42343. <https://doi.org/10.1371/journal.pone.0042343>.
- Lin L, Chen Q, Hirsch JP, Yoo S, Yeung K, Bumgarner RE, Tu Z, Schadt EE, Zhu J. Temporal genetic association and temporal genetic causality methods for dissecting complex networks. *Nat Commun*. 2018;9(1):3980. <https://doi.org/10.1038/s41467-018-06203-3>.
- Steel J, Lowen AC, Pena L, Angel M, Solórzano A, Albrecht R, Perez DR, García-Sastre A, Palese P. Live attenuated influenza viruses containing NS1 truncations as vaccine candidates against H5n1 highly pathogenic avian influenza. *J Virol*. 2009;83(4):1742–53. <https://doi.org/10.1128/JVI.01920-08>.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and

- microarray studies. *Nucleic Acids Res.* 2015;43(7):47. <https://doi.org/10.1093/nar/gkv007>.
9. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, Shen Y, Pevouchine DD, Djebali S, Thurman RE, Kaul R, Rynes E, Kirilusha A, Marinov GK, Williams BA, Trout D, Amrhein H, Fisher-Aylor K, Antoshechkin I, DeSalvo G, See L-H, Fastuca M, Drenkow J, Zaleski C, Dobin A, Prieto P, Lagarde J, Bussotti G, Tanzer A, Denas O, Li K, Bender MA, Zhang M, Byron R, Groudine MT, McCleary D, Pham L, Ye Z, Kuan S, Edsall L, Wu Y-C, Rasmussen MD, Bansal MS, Kellis M, Keller CA, Morrissey CS, Mishra T, Jain D, Dogan N, Harris RS, Cayting P, Kawli T, Boyle AP, Euskirchen G, Kundaje A, Lin S, Lin Y, Jansen C, Malladi VS, Cline MS, Erickson DT, Kirkup VM, Learned K, Sloan CA, Rosenbloom KR, Lacerda de Sousa B, Beal K, Pignatelli M, Flicek P, Lian J, Kahveci T, Lee D, Kent WJ, Ramalho Santos M, Herrero J, Notredame C, Johnson A, Vong S, Lee K, Bates D, Neri F, Diegel M, Canfield T, Sabo PJ, Wilken MS, Reh TA, Giste E, Shafer A, Kutayavin T, Haugen E, Dunn D, Reynolds AP, Neph S, Humbert R, Hansen RS, De Bruijn M, Selleri L, Rudensky A, Josefowicz S, Samstein R, Eichler EE, Orkin SH, Levasseur D, Papayannopoulou T, Chang K-H, Skoultschi A, Gosh S, Disteche C, Treuting P, Wang Y, Weiss MJ, Blobel GA, Cao X, Zhong S, Wang T, Good PJ, Lowdon RF, Adams LB, Zhou X-Q, Pazin MJ, Feingold EA, Wold B, Taylor J, Mortazavi A, Weissman SM, Stamatoyannopoulos JA, Snyder MP, Guigo R, Gingeras TR, Gilbert DM, Hardison RC, Beer MA, Ren B, Mouse ENCODE Consortium. A comparative encyclopedia of DNA elements in the mouse genome. *Nature.* 2014;515(7527):355–64. <https://doi.org/10.1038/nature13992>.
 10. Martínez-Naves E, Lafuente EM, Reche PA. Recognition of the ligand-type specificity of classical and non-classical MHC I proteins. *FEBS Lett.* 2011;585(21):3478–84. <https://doi.org/10.1016/j.febslet.2011.10.007>.
 11. Kajikawa M, Baba T, Tomaru U, Watanabe Y, Koganei S, Tsuji-Kawahara S, Matsumoto N, Yamamoto K, Miyazawa M, Maenaka K, Ishizu A, Kasahara M. MHC class I-like MILL molecules are beta2-microglobulin-associated, GPI-anchored glycoproteins that do not require TAP for cell surface expression. *J Immunol (Baltimore, Md.: 1950).* 2006;177(5):3108–15.
 12. Zhang B, Wei H, Zheng X, Zhang J, Sun R, Tian Z. The inhibitory effects of synthetic short peptides, mimicking MICA and targeting at NKG2d receptors, on function of NK cells. *Peptides.* 2005;26(3):405–12. <https://doi.org/10.1016/j.peptides.2004.10.008>.
 13. Ritter C, Fan K, Paulson KG, Nghiem P, Schrama D, Becker JC. Reversal of epigenetic silencing of MHC class I chain-related protein A and B improves immune recognition of Merkel cell carcinoma. *Scientific Reports.* 2016;6:21678. <https://doi.org/10.1038/srep21678>.
 14. Rabinovich BA, Ketchum RR, Wolfson M, Goldstein L, Skelly M, Cosman D. A role for the MHC class I-like Mill molecules in nutrient metabolism and wound healing. *Immunol Cell Biol.* 2008;86(6):489–96. <https://doi.org/10.1038/icc.2008.41>.
 15. Wu W, Booth JL, Duggan ES, Wu S, Patel KB, Coggeshall KM, Metcalf JP. Innate immune response to H3N2 and H1N1 influenza virus infection in a human lung organ culture model. *Virology.* 2010;396(2):178–88. <https://doi.org/10.1016/j.virol.2009.10.016>.
 16. Guarner J, Falcón-Escobedo R. Comparison of the Pathology Caused by H1N1, H5N1, and H3N2 Influenza Viruses. *Arch Med Res.* 2009;40(8):655–61. <https://doi.org/10.1016/j.arcmed.2009.10.001>.
 17. Gu J, Xie Z, Gao Z, Liu J, Korteweg C, Ye J, Lau LT, Lu J, Gao Z, Zhang B, McNutt MA, Lu M, Anderson VM, Gong E, Yu ACH, Lipkin WI. H5N1 infection of the respiratory tract and beyond: a molecular pathology study. *Lancet.* 2007;370(9593):1137–45. [https://doi.org/10.1016/S0140-6736\(07\)61515-3](https://doi.org/10.1016/S0140-6736(07)61515-3).
 18. Sreta D, Kedkovid R, Tuamsang S, Kitikoon P, Thanawongnuwech R. Pathogenesis of swine influenza virus (Thai isolates) in weanling pigs: an experimental trial. *Viol J.* 2009;6(1):34. <https://doi.org/10.1186/1743-422X-6-34>.
 19. Fislová T, Gocník M, Sládková T, Ďurmanová V, Rajčáni J, Varečková E, Mucha V, Kostolanský F. Multiorgan distribution of human influenza A virus strains observed in a mouse model. *Arch Virol.* 2009;154(3):409–19. <https://doi.org/10.1007/s00705-009-0318-8>.
 20. Tang X, Chong KT. Histopathology and growth kinetics of influenza viruses (H1N1 and H3N2) in the upper and lower airways of guinea pigs. *J Gen Virol.* 2009;90(2):386–91. <https://doi.org/10.1099/vir.0.007054-0>.
 21. Munster VJ, de Wit E, van den Brand JMA, Herfst S, Schrauwen EJA, Bestebroer TM, van de Vijver D, Boucher CA, Koopmans M, Rimmelzwaan GF, Kuiken T, Osterhaus ADME, Fouchier RAM. Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. *Science.* 2009;325(5939):481–3. <https://doi.org/10.1126/science.1177127>.
 22. van den Brand JMA, Stittelaar KJ, van Amerongen G, Rimmelzwaan GF, Simon J, de Wit E, Munster V, Bestebroer T, Fouchier RAM, Kuiken T, Osterhaus ADME. Severity of pneumonia due to new H1N1 influenza virus in ferrets is intermediate between that due to seasonal H1N1 virus and highly pathogenic avian influenza H5N1 virus. *J Infect Dis.* 2010;201(7):993–9. <https://doi.org/10.1086/651132>.
 23. Monticelli LA, Sonnenberg GF, Abt MC, Alenghat T, Ziegler CGK, Doering TA, Angelosanto JM, Laidlaw BJ, Yang CY, Sathaliyawala T, Kubota M, Turner D, Diamond JM, Goldrath AW, Farber DL, Collman RG, Wherry EJ, Artis D. Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nat Immunol.* 2011;12(11):1045. <https://doi.org/10.1038/ni.2131>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

