



OPEN

Network analysis of the immune state of mice

Elohim Fonseca dos Reis¹, Mark Viney² & Naoki Masuda^{1,3,4}✉

The mammalian immune system protects individuals from infection and disease. It is a complex system of interacting cells and molecules, which has been studied extensively to investigate its detailed function, principally using laboratory mice. Despite the complexity of the immune system, it is often analysed using a restricted set of immunological parameters. Here we have sought to generate a system-wide view of the murine immune response, which we have done by undertaking a network analysis of 120 immune measures. To date, there has only been limited network analyses of the immune system. Our network analysis identified a relatively low number of communities of immune measure nodes. Some of these communities recapitulate the well-known T helper 1 vs. T helper 2 cytokine polarisation (where ordination analyses failed to do so), which validates the utility of our approach. Other communities we detected show apparently novel juxtapositions of immune nodes. We suggest that the structure of these other communities might represent functional immunological units, which may require further empirical investigation. These results show the utility of network analysis in understanding the functioning of the mammalian immune system.

The vertebrate immune system defends animals against infection and disease. There has been extensive study of how the immune system functions, not only to understand its basic biology, but also to be able to manipulate it for therapeutic benefit. The immune system is a complex system with many cell populations, sub-populations and soluble molecules (such as cytokines) that act and interact to generate an immune response and to regulate that response. At its heart, the purpose of the immune system is to recognize foreign antigen molecules, to then remove, destroy or nullify them and their source, while usually retaining a molecular memory of the antigen in question. The immune system is dynamic, so that different components of the immune system respond depending on the type of antigen and its location in the individual. For example, immune responses to viruses infecting lung cells are qualitatively very different from the immune responses directed against macroscopic worms living in the gut¹. Effector cells and molecules of the immune system act against invading pathogens, but these effects can also cause harm to the individual, so-called immunopathology². Therefore, a tightly controlled, well-regulated immune system is needed to protect an individual from pathogen-induced harm, without causing direct harm to the individual in the process.

Despite the complex, multi-faceted nature of the immune system it is commonly analysed and quantified using relatively few immunological parameters; these parameters may be those that can feasibly be measured and / or those that are thought to be of particular importance with respect to what is being studied. This often low-dimension analysis contrasts with the complex, high-dimensionality of the immune response. To date there has been rather little large scale, immune system-wide analysis of the immune response. Therefore, within the study, analysis and understanding of the immune system there is a paradox: that immunology text books and review articles are filled with complex network diagrams (not unlike electrical circuits) showing how cells and molecules of the immune system are thought to act and interact³⁻⁶, but there has actually been limited immune system-level analyses of immunological data, which should actually underpin such system-level views of the immune system.

In the present work, we seek to address this by performing a network analysis of immunological data that we previously collected^{7,8}. Networks (also called graphs) have proved to be a useful mathematical language to describe complex biological systems including interactions among cells, molecules and genes^{9,10}. However, as far as we are aware, there are only a limited number of published studies that have used network analysis for investigating immunological data. Specifically, correlational networks have been constructed, similar to the present study, but limited to cytokine data only^{11,12}, or a combination of data on cytokines, other soluble molecules and

¹Department of Mathematics, State University of New York at Buffalo, Buffalo 14260, USA. ²Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool L69 7ZB, UK. ³Computational and Data-Enabled Science and Engineering Program, State University of New York at Buffalo, Buffalo 14260, USA. ⁴Faculty of Science and Engineering, Waseda University, Tokyo 169-8555, Japan. ✉email: naokimas@buffalo.edu

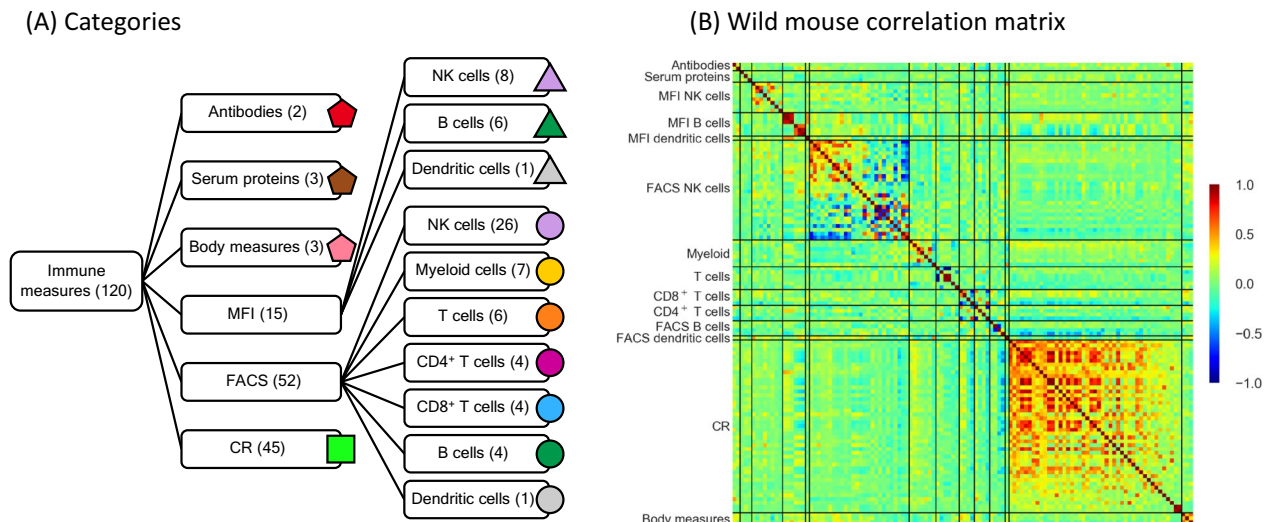


Figure 1. (A) Categories of immune measures. The numbers in the parentheses are the number of immune measures within that category. The symbol and colour associated with each category are used in the subsequent figures. The Mean Fluorescence Index (MFI) and Fluorescence Activated Cell Sorted (FACS) categories are divided into three and seven sub-categories, respectively. (B) Correlation matrix (coloured on a -1 to $+1$ scale) ordered according to the categories of immune measures for wild mice as shown in (A). The solid black lines separate different categories of immune measures.

some cell populations, but from a small number of groups of subjects (i.e., a group of at most 11 participants)¹³. Other studies have undertaken a network analysis of components of the Nuclear Factor κ B system, which is involved in mediating inflammation¹⁴, or networks of cytokines, chemokines and other relevant proteins in patients with pulmonary arterial hypertension¹⁵. A different approach that has been used is to construct networks based on protein abundance data, where the presence of those proteins in different immune cell populations was then analysed¹⁶. Network analysis has also been undertaken on proteomic data of haemopoietic cell types, to investigate interactions between those cell types¹⁷. Other analyses have constructed networks from prior immunological knowledge, and these networks have then been analysed, including by the addition of data^{18,19}.

In contrast to these studies, here we use a large sample of real-world immunological data from hundreds of mice (*Mus musculus domesticus*), which has quantified a full range of serum proteins, cytokines, and cellular populations and sub-populations, totalling 120 immune measures. We build and analyse correlational networks from these data. Benefiting from the relatively large size of our network, we analyse the community structure of the immune networks aiming to reveal groups of immune responses and how they are interconnected²⁰, which we show is not attainable by ordination-based analysis such as principal component analysis. In particular, we analyse signatures of T helper type 1 (Th1) and type 2 (Th2) responses in these networks. Crucially, our network analysis is purely data-driven and does not use a priori grouping of data based on biological knowledge. In this manner, we substantially advance network analyses of the immune system showing that the network recovers some known features of the immune response, thus showing the potential utility of such analyses in future immunological research.

Materials and methods

Data. We used a previously published data set of the immune state of 460 wild mice (*Mus musculus domesticus*) and of 102 laboratory mice⁷. The wild mice were live trapped from 12 sites in the southern UK, and their mass, age, infection status and physiological state were measured, and these data are presented in our previous paper⁷. From the original 126 immune measures we pre-processed the data and excluded six immune measures (see Supplementary Text S1), leaving data for 120 immune measures, which are classified into six categories and sub-categories, as shown in Fig. 1A.

Concerning the six categories of immune measures (Fig. 1A) these are: (i) antibodies, which are the serum concentration of immunoglobulins G (IgG) and E (IgE); (ii) serum proteins, which are the concentration of the acute phase proteins serum amyloid P (SAP), Haptoglobin, and alpha-1 antitrypsin (AAT); (iii) spleen mass, number of spleen cells and body mass; (iv) the ex vivo activation state measured as mean fluorescence index (MFI) for three cell types, which has a total of 15 measures that are detailed in Supplementary Table S1; (v) fluorescence activated cell sorted (FACS)-derived percentages of seven cell types, which has a total of 52 measures, detailed in Supplementary Table S1, and (vi) cytokine responses (CR), which is the concentration of a cytokine produced in vitro after each of four types of stimulation, and one control, for nine cytokines, yielding 45 CR measures, detailed in Supplementary Table S1.

Many of the 120 immune measures had missing values, which is to be expected when working with wild populations. In particular the CR response data were missing for 223 wild mice (48%) and 75 laboratory mice (74%), and because of the central importance of cytokines in immune responses, we excluded these mice entirely from our analysis. For the remaining mice that had some missing values we imputed these, which we did by calculating

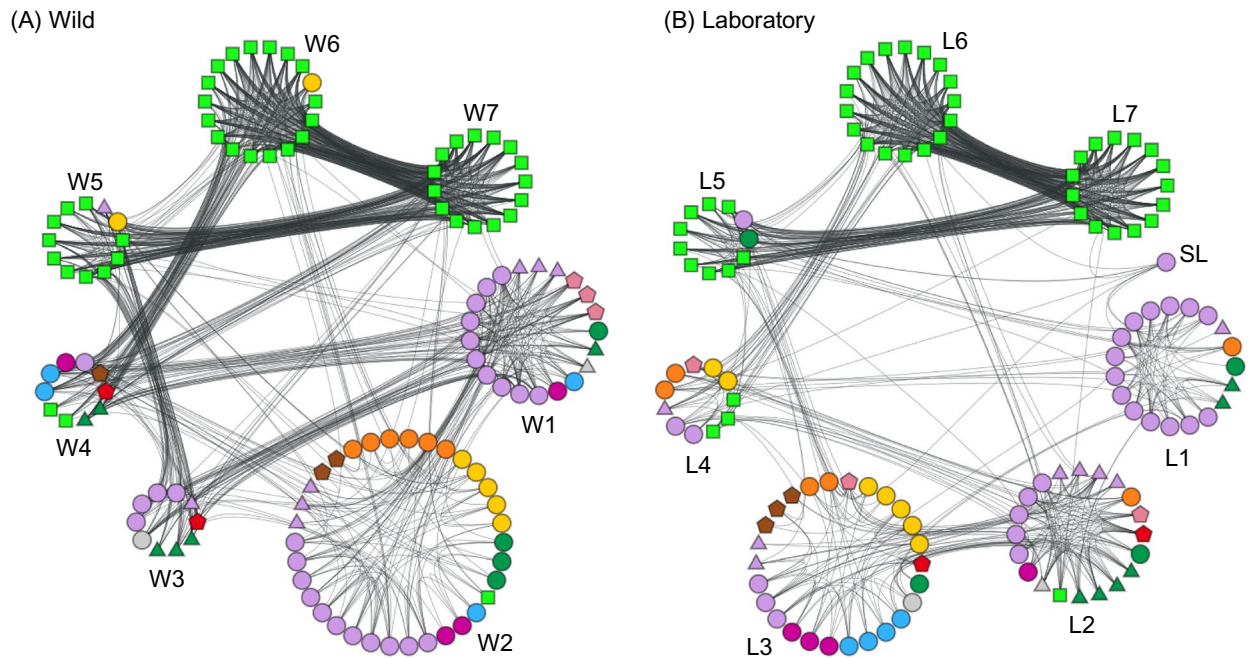


Figure 2. Community structure of the (A) wild and (B) laboratory mouse network, with the seven communities within each (W1–W7 and L1–L7), and the additional single (SL) node in the laboratory mouse network. Nodes in both networks are coded as Fig. 1A, and fully labelled networks are shown in Supplementary Fig. S2.

the average value from the available observations, and then replaced the missing values by this calculated average. We carried out this imputation procedure separately for the wild and for the laboratory mouse data sets.

Correlation of immune measures and comparative analyses. From these data, separately for the wild and for the laboratory mice, we calculated the Pearson correlation coefficient between each pair of immune measures, with each mouse being a sample. We used two methods to analyse these correlation matrices. Firstly, we built positive correlation networks using the 120 immune measures as network nodes. We placed an edge (which joins nodes) if, and only if, the correlation value exceeded a prescribed threshold; we explored the results at different correlation threshold values. Although using such a threshold discards the information contained in the value of the correlation coefficient, it does allow the application of various network analysis tools to such data. In fact, using thresholds with correlation networks have been shown to be useful for analysing various types of data such as neuroimaging^{21,22}, genomic^{23,24}, psychological²⁵, climatological²⁶ and financial²⁷ data.

To build a connected network we used a threshold that was the largest possible correlation value at which the network remained connected and consisted of a single connected component. In these networks, the edge density is defined as $2e/N(N-1)$, i.e. the number of edges e present in the network, divided by the maximum number of possible edges $N(N-1)/2$ in the network, with N nodes. To detect communities of nodes in these networks we combined a stochastic block model (SBM)^{28–31} and a consensus clustering approach, uncovering the mesoscopic block structure of the correlation networks, which is described in Supplementary Text S2. Secondly, we used Principal Component Analyses (PCA), which is described in Supplementary Text S3.

Results and discussion

Pairwise correlations among wild mouse immune measures. The pairwise correlation matrix for the wild mice is shown in Fig. 1B, where the immune measures are grouped according to the categories of immune measures, and within each category individual measures are arranged in descending order of correlation coefficient. This shows that there is a concentration of large positive correlations among many members of the CR category. There are also strong positive and negative correlations within many of the other categories of immune measures (i.e., diagonal blocks in Fig. 1B). Within the FACS NK cells category there is a large proportion of immune measure pairs that are strongly, negatively correlated, while others are strongly, positively correlated. A frequency distribution of these pairwise correlation coefficients shows this mix of positive and negative correlations, and that there is a skew to positive correlation coefficients (Supplementary Fig. S1). While negative regulation is common in biological systems, in our network analyses, below, we focus on the positive correlations.

Wild mouse immune network. We constructed a wild mouse network with a threshold of 0.2, which was the highest correlation coefficient threshold that generated a connected network, consisting of seven communities and an edge density of 0.16. (Fig. 2A; Supplementary Fig. S2). Networks constructed with lower thresholds had a broadly similar structure of relatively few communities (specifically consisting of 10 or 12 communities,

with a number of nodes not belonging to any community) and that the CR nodes were overwhelmingly concentrated in just 3 or 4 communities (Supplementary Fig. S3).

There are several notable features of the wild mouse network. Firstly, that the different categories of immune measures are distributed among the seven communities (Fig. 2A; Supplementary Fig. S2). Secondly, that three of the communities are almost exclusively composed of CR measures. Thirdly, that all communities are connected to each other (though to differing degrees), except for W3 and W6 that are not connected at all. Fourthly, that most communities have multiple within-community links as well as among-community links, though W3 and W4 have comparatively few within-community links.

These results show that the wild mouse immune network does not overtly resemble standard diagrams that summarise what is understood about the functioning of the immune system. This is notable because the network we present is wholly generated from analysis of empirical data. We suggest that the network and community structure revealed by this analysis represent functional aspects of the immune system. A priori we suggest that the different communities represent integrated, functional immunological units.

Cytokines in the wild mouse network. To investigate whether the different communities in the network do represent functional immunological units we examined the network for evidence of known immunological features. Specifically, we sought evidence of cytokine communities consistent with the well-established Th1 vs. Th2 immune system polarization. Specifically, Interferon-gamma (IFN- γ) responses are indicative of Th1 responses, whereas Interleukin-4 (IL-4) and IL-13 responses are indicative of Th2 responses¹. (IL-5 is similarly indicative of Th2 responses, but these data are not available in this study.) In the wild mouse network, the Th2-marker cytokine responses are present in communities W4 and W6, and the Th1-marker cytokine responses in W5 and W7. We tested whether the concentration of Th1/Th2-marker nodes into just two communities was statistically significant, by comparing their observed distribution with that for random assignments of these nodes into communities (Supplementary Text S4). In doing this we collapsed the five IL-13 nodes into one node, because each of these nodes was highly correlated with each other (Supplementary Text S4). The results show that the observed concentration of the Th2-marker nodes into W4 and W6 was significantly different from a random assignment of these nodes. The observed concentration of the Th1-marker nodes into W5 and W7 was marginally significantly different from a random assignment of these nodes. There are further features of the wild mouse network that also support W6 being a Th2 community, based on the strong links between W4 and W6 (Supplementary Figure S4). These are that W4 contains two of the five IL-4 cytokine responses, and that W4 contains two of the four CD8⁺ T cell populations, themselves indicative of Th2 responses (Supplementary Fig. S4). Overall, these results support the idea that there is polarization of Th1 vs. Th2 immune measures among different communities in the wild mouse network.

These results therefore show that this network analysis is able to recover evidence of known features of the mouse immune response. Therefore, the network approach may be a promising tool for analysing immunological data to reveal functional immunological groups. That our results have found other communities suggests that these may represent other, hitherto unknown, functional features of the immune response. Further observational and experimental work would now need to be undertaken to investigate this.

It is notable that the immunological phenotype of knock-out mice can be different from the predicted phenotype, which is in part due to the complex, inter-connected nature of the immune system, such that if a single feature is knocked-out, then there are multiple consequences throughout the immune system³². In an analogous fashion, the communities detected in this network analysis may represent immune-system wide immunological units, where their function or role (which, of course, may be multi-faceted) is not intuitively obvious. Further empirical investigation of these communities would be warranted.

Wild mouse PCA analysis. To understand how network analysis compared with another method commonly used to analyse biological data, we used a principal component analysis (PCA) with the immune data. The first two principal components explained 99.6% and 0.4% of the variance in the data, respectively (Fig. 3A; Supplementary Text S3). The high concentration of the variance in the first component appears to be due to one outlier measure, the number of spleen cells. We therefore removed this measure and repeated the PCA with the remaining 119 measures. This showed that the first and second components now explained 67% and 19% of the variance, respectively (Fig. 3B). Notwithstanding, the PCA results do not show any clustering of the immune measures, except for MFI B cells. Similar results were obtained for the laboratory network (Supplementary Text S3).

Overall, this analysis did not reveal any strong patterns within the data, and nor did it recapitulate the different categories of immune measures present within our data set. This analysis also did not reveal the Th1 vs. Th2 cytokine communities, which were evident in the network analysis. We therefore conclude that network analysis of the present immunological data is superior to this PCA-based analysis in identifying functional immunological units.

Comparison of wild and laboratory mouse immune networks. Our data set consisted of data obtained from wild mice and from laboratory mice. There are clear differences among mice living in these two environments, and immunological differences between these groups of mice have already been identified^{7,8,33}. We therefore used network analysis to further compare the immune state of these two groups of mice, comparing single immune networks for both the wild (above) and laboratory mice. For the laboratory mouse network, the threshold was 0.42 and the edge density 0.13, and the network had seven communities and a single node that did not belong to any community (Fig. 2B; Supplementary Fig. S5). Laboratory mouse networks constructed with lower thresholds had a broadly similar structure of relatively few communities (specifically, consisting of 10

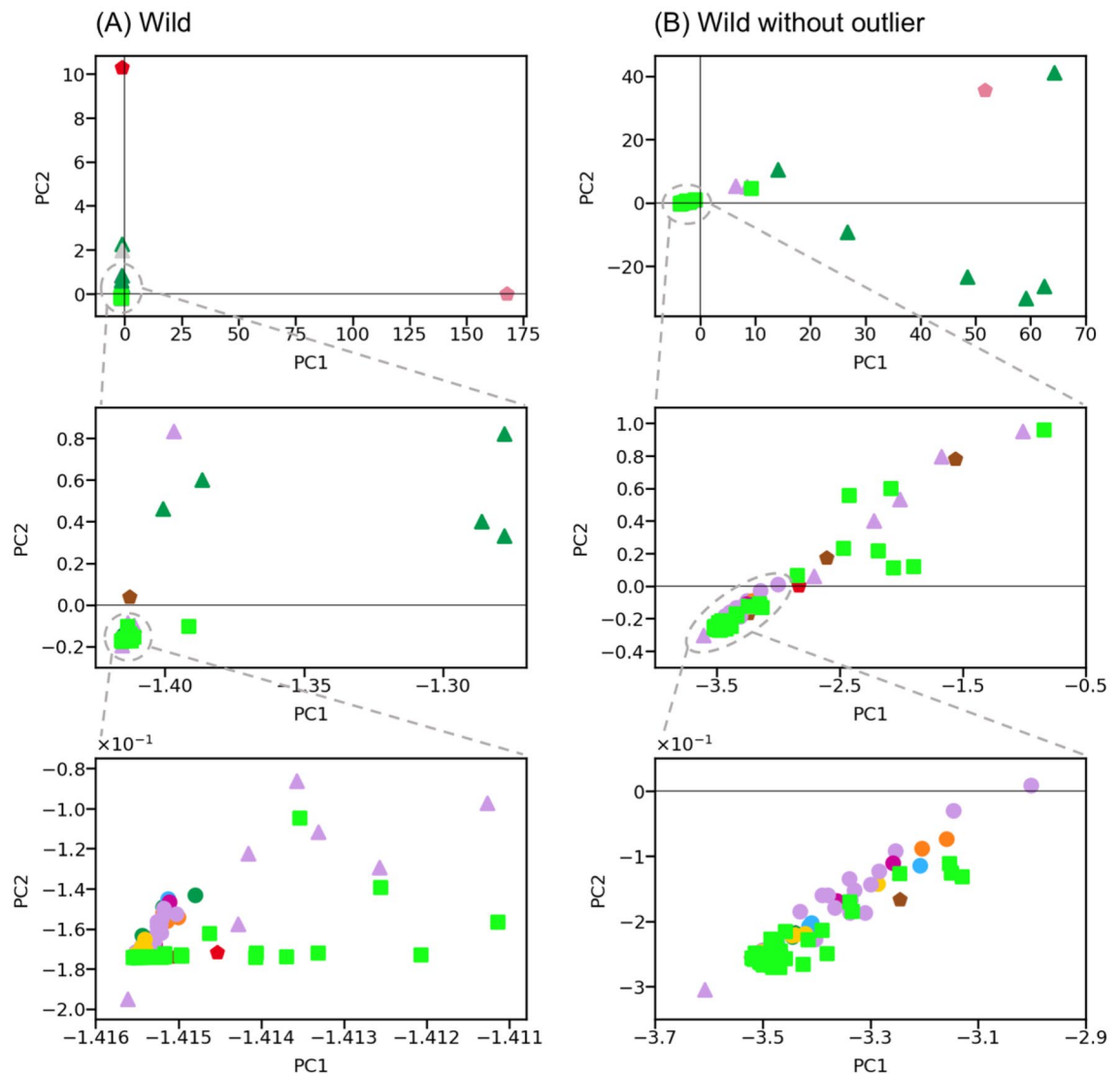


Figure 3. PCA results for the wild mouse data. Projection of the first (PC1) and the second (PC2) principal components of the immune measures (A) with and (B) without the outlier (number of spleen cells), with three levels of expansion (shown by the dotted lines) to reveal detail. Nodes are coded as in Fig. 1A.

or 12 communities with some nodes not belonging to any communities) and with the CR nodes overwhelmingly concentrated in just 3 or 4 communities (Supplementary Fig. S3).

As for the wild mouse network, the different categories of immune measures are distributed among the communities; three communities are almost entirely composed of CR measures (as for the wild mouse network); many communities are connected to each other, though L1 and L6, L3 and L7, and L3 and L6 are not connected; and there are multiple within-community and among-community edges (as for the wild mouse network). Applying PCA to the laboratory mouse data showed that most of the data variance was encompassed in the first principal component (with the second principal component accounting for only 6×10^{-6} of the variance), and the analysis did not reveal any strong patterns, nor recapitulate features revealed by the network analysis (Supplementary Text S3; Supplementary Fig. S6).

In comparing the wild and laboratory immune networks, it is notable that both have three communities in which most CR nodes are concentrated (henceforth CR communities). Specifically, of the 45 CR nodes, 42 and 41 nodes belong to CR communities in the wild (W5, W6, W7) and laboratory (L5, L6, L7) mouse networks, respectively. The concentration of the CR nodes in these three communities was significantly higher than in a random assignment of the nodes among the seven communities, for both wild and laboratory networks (Supplementary Text S5). Moreover, the interconnections between the three CR communities is qualitatively the same for both networks—there is one CR community that is densely connected to the other two CR communities, which are in turn sparsely connected to each other. In other words, the three CR communities have a chain-like connection. In the wild network, community W7 has edge densities of 0.72 and 0.82 with communities W5 and W6, respectively, whereas the edge density between W5 and W6 is 0.06. Analogously, in the laboratory network, L7 has edge densities of 0.60 and 0.83 with L5 and L6, respectively, whereas the edge density between L5 and L6

is 0.03. These three CR communities have a significantly more chain-like connection than do randomly rewired CR subnetworks (Supplementary Text S6).

But, despite these structural similarities, the wild and laboratory mouse CR communities consist of different sets of nodes. We examined this further by computing the number of common nodes and the Jaccard index between each pair of communities in the wild and laboratory networks (Supplementary Text S7; Supplementary Fig. S7). Communities W7 and L7, which are strongly connected to the other two CR communities in the wild and laboratory networks, respectively, have only two CR nodes in common with a Jaccard index of 0.07 (Supplementary Fig. S7). In fact, this number (i.e., two) is not significantly different from the number of CR nodes in common when the nodes are randomly assigned to the three CR communities in the wild and laboratory networks (Supplementary Text S7). Moreover, the number of common nodes between W5 and W6 and L5 and L6 in all pair-wise W to L comparisons is not different from the number in randomized networks (Supplementary Text S7). Therefore, the nodes in the CR communities W5 and W6 (which are both strongly connected to W7, but not strongly connected to each other) are not concentrated in either L5 or L6 (which are both strongly connected to L7). Rather, many nodes in W6 and W7 are found in L7 and L6, respectively (Supplementary Text S7). For example, in the wild network all of the IL-1 β , IL-12p70, and IL-13 nodes are each within one community (i.e., W7, W6, and W6, respectively), whereas in the laboratory network only the IL-12p70 nodes are all within one community (i.e., L7) (Supplementary Table S2). In summary, although the mesoscale structure of the CR communities is similar between the wild and laboratory mouse networks, the composition of each CR community is substantially different between the two networks. This difference might represent functional immunological differences between the cytokines of the wild and laboratory mice.

We also compared the immune networks from the perspective of the Th1 vs. Th2 cytokines, finding that the laboratory mouse network differs from that of the wild mice. Specifically, the IFN- γ Th1-marker cytokine nodes are in L2, L5 and L6, and the Th2-marker cytokine nodes are in L5, L6 and L7. We find that the Th1 or Th2 cytokine nodes in the laboratory network are not more concentrated than in a random assignment of the nodes into communities (Supplementary Text S4). We suggest that this difference between wild and laboratory mice represents a functional immunological difference between such mice. Thus, in wild mice extensive antigen exposure has stimulated both Th1 and Th2 immune responses that, with their cross-regulation, are manifest as separate communities within the immune network. In contrast, the laboratory mice will have had very limited antigenic exposure such that meaningful Th1 vs. Th2 immune responses have not been generated, hence their absence from the immune network.

We also investigated the distribution of 12 humoral response nodes, consisting of antibody concentration, and FACS and MFI of B cells (Fig. 1). We found that six of these are in L2 and three in L1 (Fig. 2), which is suggestive of some community concentration of these humoral nodes. This pattern contrasts with the wild network, where these 12 nodes were in communities W1-4. These nodes are significantly more concentrated in these laboratory network communities compared with random assignment of these nodes; in the wild network the observed distribution of these nodes was not different from a random assignment (Supplementary Text S8). A different arrangement of these humoral response nodes between the wild and laboratory mouse networks may be due to their different humoral responses, with wild mice having very high concentrations of antibodies compared with laboratory mice⁷.

There are other similarities and differences among the wild and laboratory mouse networks. In both, a majority of NK cell nodes (i.e. FACS and MFI nodes) belong to two communities, W1 and W2, and L1 and L2. Communities W1 and L1 are predominately composed of NK cell nodes and all the eight nodes that they share are FACS NK cell nodes. Among the seven nodes shared by communities W2 and L2, five are NK cell nodes (three FACS NK cells and two MFI NK cells). In both networks, the edges within each of these two NK cell communities are more dense than between them. Consistent with this observation, the pairwise correlations between nodes within L1 and L2 are mostly strongly positive whereas those between L1 and L2 are mostly strongly negative (Supplementary Fig. S8). This result may imply negative biological regulation among NK cell nodes.

T cell nodes differ among the wild and laboratory mouse networks. In the wild network, T cells are confined to a single community, W2, whereas they are dispersed among four communities in the laboratory mouse network (Fig. 2; Supplementary Figs. S2 and S5). This result is likely due to greater functional interaction among these cells in wild, antigen-experienced mice, compared with laboratory mice³³.

There are therefore some similarities among the wild and laboratory networks, but this is not completely unexpected since both wild and laboratory mice are the same species, and so have the same immune system with the same components, which will necessarily have some, even substantial, shared function. Despite the mesoscale similarity, the node composition of the individual communities differs substantially between the wild and laboratory networks. We find considerable differences between the two networks in: the identity of nodes constituting the three CR communities; evidence for separation of Th1 vs. Th2 cytokine nodes; and different arrangement of the T cell nodes. However, when we compare the overall community structure of the two networks, i.e., how all the nodes are partitioned into communities, we were not able to find statistical support for differences in the structure of the wild and laboratory mouse networks (Supplementary Text S9; Supplementary Table S5).

Negative correlations. We have not considered negative pairwise correlations because the thresholds for creating the networks from the correlation matrices were positive. However, mindful that in biological systems negative regulation is common, we also briefly analysed how the immune measures are connected through negative correlations. In these negative networks, node pairs that are strongly negatively correlated were assumed to form the edges. The negative networks have substantially more communities (9 and 17 for the wild and laboratory, respectively) than the positive networks such that their community structure is more difficult to interpret (Supplementary Fig. S9). However, the CR nodes are predominately grouped together within several commu-

nities in both negative networks, similarly to the positive networks. Specifically, in the wild mouse negative network, 43 out of the 45 CR nodes are in one of two communities; for the laboratory mouse negative network 39 CR nodes belong to one of the four communities. Because CR nodes tend to be strongly positively correlated with each other, there is no direct connection between pairs of CR nodes in the negative networks. However, in the SBM, a pair of nodes tend to belong to the same community if they have similar patterns of connectivity to the other nodes in the network. This is the case even if the direct connectivity between the two nodes is not strong, or even absent. This property of the SBM explains why various CR nodes belong to the same community despite the absence of the direct connections among them. The results for the negative networks suggest that, for both wild and laboratory mice, the groups of CR nodes have relatively similar patterns of negative connectivity towards other CR or non-CR nodes.

Conclusion

We present a network analysis of the immune state of mice based on a large, comprehensive set of immune measures and a large sample of mice. This analysis therefore considerably expands the previous use of network analysis for investigating the immune system and immune responses. Our analysis is able to recover some known features of immune function (the Th1 vs. Th2 cytokine polarisation) suggesting, by extension, that the other communities that we have identified may also represent immunological functional units. We show that the network analysis is superior to PCA for our data. However, formal statistical analysis failed to identify differences between the wild and laboratory mouse networks. We suggest that network analyses such as the one we have presented complement other non-network methods of analysis, by providing measures not available from other methods.

The present study has some limitations. First, we necessarily discarded a considerable portion of data during our pre-processing steps due to missing values. Studies of wild animals will routinely confront such limitations. Therefore, the pre-processing of data that we have undertaken may provide a model for how this matter can be addressed in similar analyses of analogous data of other populations. A larger data set would significantly enhance the precision of the present findings. Second, we were not able to find statistical support for differences between the wild and laboratory mouse networks. As above, this may be due to the relatively small size of the data set, compared to the number of nodes. Sampling wild mice and making these multiple immune measures is a substantial piece of work and we note that this data set is the largest existing immunological data set for wild rodents. Third, we applied a threshold to the correlation matrices to create the correlation networks. This procedure loses potentially useful information contained in the value of the correlation. The choice of the threshold values we used was not determined a priori, but rather determined functionally by using the highest value that resulted in a single connected component^{34,35}. Alternative options to select the threshold value could be based on an optimization method³⁵ or the use of network quantities directly applied to correlation matrices^{36–38}. Fourth, we have constructed networks based on either positive or negative correlations, which therefore does not allow a network based on correlations that are in different directions. Biological systems are typically both positively and negatively regulated, and so a network-style analysis that would allow both positive and negative correlations could be more valuable.

The network analysis of wild mouse populations presented here is a novel analysis of a unique data set. It generates a holistic, data-driven view of the mammalian immune system and the dynamics of its function as the animals themselves interact with diverse and dynamic environments. Humans are, arguably, immunologically more akin to wild mice than to laboratory mice^{7,8,39}, and the diversity of immune state among people could also be analysed using network analyses. Such an approach may usefully be used to better understand the mammalian immune system and its function in protecting individuals from infection and disease.

Received: 11 August 2020; Accepted: 21 January 2021

Published online: 22 February 2021

References

1. Finkelman, F. D. *et al.* Cytokine regulation of host defense against parasitic gastrointestinal nematodes: Lessons from studies with rodent models. *Annu. Rev. Immunol.* **15**, 505 (1997).
2. Graham, A. L., Allen, J. E. & Read, A. F. Evolutionary causes and consequences of immunopathology. *Annu. Rev. Ecol. Evol. Syst.* **36**, 373 (2005).
3. Callard, R., George, A. J. T. & Stark, J. Cytokines, chaos, and complexity. *Immunity* **11**, 507 (1999).
4. Yates, A., Chan, C. C. W., Callard, R. E., George, A. J. T. & Stark, J. An approach to modelling in immunology. *Brief. Bioinform.* **2**, 245 (2001).
5. Subramanian, N., Torabi-Parizi, P., Gottschalk, R. A., Germain, R. N. & Dutta, B. Network representations of immune system complexity. *WIREs Syst. Biol. Med.* **7**, 13 (2015).
6. Balkwill, F. R. & Burke, F. The cytokine network. *Immunol. Today* **10**, 299 (1989).
7. Abolins, S. *et al.* The comparative immunology of wild and laboratory mice, *Mus musculus domesticus*. *Nat. Commun.* **8**, 14811 (2017).
8. Viney, M. & Riley, E. M. The immunology of wild rodents: Current status and future prospects. *Front. Immunol.* **8**, 1481 (2017).
9. Barabási, A.-L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101 (2004).
10. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947 (2005).
11. Broderick, G. *et al.* A formal analysis of cytokine networks in chronic fatigue syndrome. *Brain Behav. Immun.* **24**, 1209 (2010).
12. Romero, R. *et al.* Evidence of perturbations of the cytokine network in preterm labor. *Am. J. Obstet. Gynecol.* **213**, 836 (2015).
13. Broderick, G. *et al.* A pilot study of immune network remodeling under challenge in Gulf War illness. *Brain Behav. Immun.* **25**, 302 (2011).
14. Tieri, P. *et al.* Charting the NF- κ B pathway interactome map. *PLoS ONE* **7**, e32678 (2012).
15. Sweatt, A. J. *et al.* Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circ. Res.* **124**, 904 (2019).

16. Rieckmann, J. C. *et al.* Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* **18**, 583 (2017).
17. Tieri, P. *et al.* Quantifying the relevance of different mediators in the human immune cell network. *Bioinformatics* **21**, 1639 (2005).
18. Frankenstein, Z., Alon, U. & Cohen, I. R. The immune-body cytokine network defines a social architecture of cell interactions. *Biol. Direct* **1**, 32 (2006).
19. Arnold, K. B., Szeto, G. L., Alter, G., Irvine, D. J. & Lauffenburger, D. A. CD4+ T cell-dependent and CD4+ T cell-independent cytokine-chemokine network changes in the immune responses of HIV-infected individuals. *Sci. Signal.* **8**, ra104 (2015).
20. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75 (2010).
21. Zalesky, A., Fornito, A. & Bullmore, E. On the use of correlation as a measure of network connectivity. *Neuroimage* **60**, 2096 (2012).
22. de Vico Fallani, F., Richiardi, J., Chavez, M. & Achard, S. Graph analysis of functional brain networks: Practical issues in translational neuroscience. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **369**, 20130521 (2014).
23. Junker, B. H. Schreiber, F. *Analysis of Biological Networks*. (Wiley, 2008).
24. Vidal, M., Cusick, M. E. & Barabási, A.-L. Interactome networks and human disease. *Cell* **144**, 986 (2011).
25. Borsboom, D. & Cramer, A. O. J. Network analysis: An integrative approach to the structure of psychopathology. *Annu. Rev. Clin. Psychol.* **9**, 91 (2013).
26. Tsonis, A. A., Swanson, K. L. & Roebber, P. J. What do networks have to do with climate? *Bull. Am. Meteorol. Soc.* **87**, 585 (2006).
27. Mantegna, R. N., & Stanley, H. E. *Introduction to Econophysics: Correlations and Complexity in Finance*. (Cambridge University Press, 1999).
28. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **5**, 109 (1983).
29. Peixoto, T. P. Bayesian stochastic blockmodeling. in (P. Doreian, V. Batagelj, A. Ferligoj, eds). *Advances in Network Clustering and Blockmodeling*. (Wiley, 2020).
30. Peixoto, T. Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models. *Phys. Rev. E* **89**, 012804 (2014).
31. Peixoto, T. The graph-tool python library. Figshare, 2014 (updated 2019 Jul 13, cited 2020 Feb 10). <https://graph-tool.skewed.de>.
32. Barbaric, I., Miller, G. & Dear, T. N. Appearances can be deceiving: phenotypes of knockout mice. *Brief. Funct. Genomics* **6**, 91 (2007).
33. Abolins, S. *et al.* The ecology of immune state in a wild mammal, *Mus musculus domesticus*. *PLoS Biol.* **4**, e2003538 (2018).
34. Rubinov, M. & Sporns, O. Weight-conserving characterization of complex functional brain networks. *NeuroImage* **56**, 2068 (2011).
35. De Vico Fallani, F., Latora, V., Chavez, M. A topological criterion for filtering information in complex brain networks. *PLoS Comput. Biol.* **13**, e1005305 (2017).
36. MacMahon, M. & Garlaschelli, D. Community detection for correlation matrices. *Phys. Rev. X* **5**, 021006 (2015).
37. Masuda, N., Sakaki, M., Ezaki, T. & Watanabe, T. Clustering coefficients for correlation networks. *Front. Neuroinform.* **12**, 7 (2018).
38. Masuda, N., Kojaku, S. & Sano, Y. Configuration model for correlation matrices preserving the node strength. *Phys. Rev. E* **98**, 012312 (2018).
39. Beura, L. K. *et al.* Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* **532**, 512 (2016).

Acknowledgements

The wild and laboratory mouse data set was generated during the tenure of the NERC grant NE/I022892/1. We would like to thank the University of Bristol's Jean Golding Institute for seed-corn funding. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. We would also like to thank Steve Abolins, Liz King, Laura Weldon, Louise Hughes, Paul Drescher, John Raynes and Julius Hafalla for their generation of the data, and Eleanor Riley for useful discussions.

Author contributions

N.M. and M.V. conceived the study, E.F.d.R. conducted the analysis and interpreted the results with N.M. and M.V.; E.F.d.R., N.M. and M.V. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83139-7>.

Correspondence and requests for materials should be addressed to N.M.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021