Blood Groups

Integration of Pathway Commons data helps distill gene sets tailored to immune tissues that form the basis of predictors of protective vaccine responses

Quick Summary

o   Molecular predictors of vaccine responses are lacking

o   Immune tissue data are used to distill gene sets called 'blood transcriptional modules'

o   Modules represent shared immune cell processes that correlate with vaccine protection

Author Profile

Shuzhao Li, Nadine Rouphael and Sai Duraisingham are at the Emory Vaccine Center, Emory University in Atlanta, Georgia, USA. **Bali Pulendran** is in the Department of Pathology, Emory University School of Medicine.

Context

Each October, health organizations gear up to get everyone to literally "role up their sleeves" and get a flu shot. Getting children vaccinated at the clinic or pets at the veterinarian is a standard part of life in the West. However, the impact and effectiveness of vaccines, beginning with Pasteur's pioneering work in the late 1800s, cannot be understated: Their ability to mitigate disease is unparalleled in medical science. Nevertheless, our grasp of how vaccines harness the immune system towards a protective response is lacking. 'Systems vaccinology' aims to provide an integrated perspective of the mechanisms that link vaccines and immune protection[1].

Question

Do shared predictors of vaccine-induced immunity exist?

Researcher Goals

Why do some vaccines succeed while others fail? A primary obstacle in vaccine development is the lack of early, robust predictors of a protective response. On one hand, predictors would accelerate clinical trials, possibly as much as years[2]. On the other hand, vaccines would prospectively identify poor- or non-responders within days, rather than months.

Li *et al.*[3] use large-scale integration of biological data resources to distill gene sets tailor-made for the immune tissues. These 'blood transcriptome modules' (BTM) provide a fresh basis upon which to identify predictors common to the transcriptomes of vaccine-treated individuals.

Approach

Overall, Li *et al.* set out to devise gene-expression based predictors of vaccine response, often correlated with the production of antibodies active against antigen. In all, five clinically-approved vaccines were surveyed: Two against *Neisseria meningitidis* (MPSV4 and MCV4), one against yellow fever (YF-17D), and two for influenza (live attenuated influenza vaccine [LAIV], and trivalent influenza vaccine [TIV]). Transcriptome and antibody levels in peripheral blood were monitored at several points up to 30 days to capture events useful in predicting long-term protection.

An examination of vaccine transcriptomes identified few striking commonalities that might mark antibody responses. The authors turned to a pathway analysis approach to determine what 'gene sets' comprising pathways were being altered. Pathways are often more interpretable than lists of genes and can often detect pathway alterations where individual genes may not reach statistical significance. While a handful of pathways modulated by each vaccine demonstrated strong correlation with antibody production, the results left much to be desired in terms of unifying principles.

> *"Systems-level approaches for the immune system present unique challenges."*

Study of the immune system presents unique challenges. First, it is a hierarchy of highly-interdependent cells that are anatomically distributed. Second, biological knowledge databases are biased towards oncology and often contain data generated under conditions that do not reflect healthy immune systems. Li *et al.* sought an alternative knowledge base tailored to the idiosyncrasies of immune tissue.

A large-scale data integration approach was used to generate a collection of gene sets called 'blood transcriptional modules' (BTM) to address the goal of finding common principles underlying vaccine gene expression (Fig. 1). Briefly, over 32 000 transcriptomes from 540 studies were extracted from the NCBI gene expression omnibus (GEO). A master network of over 7 000 genes was constructed where edges connect gene pairs with elevated 'mutual information' across GEO data sets. In this case, mutual information effectively measures gene co-expression and such collections are referred to as 'relevance' networks.
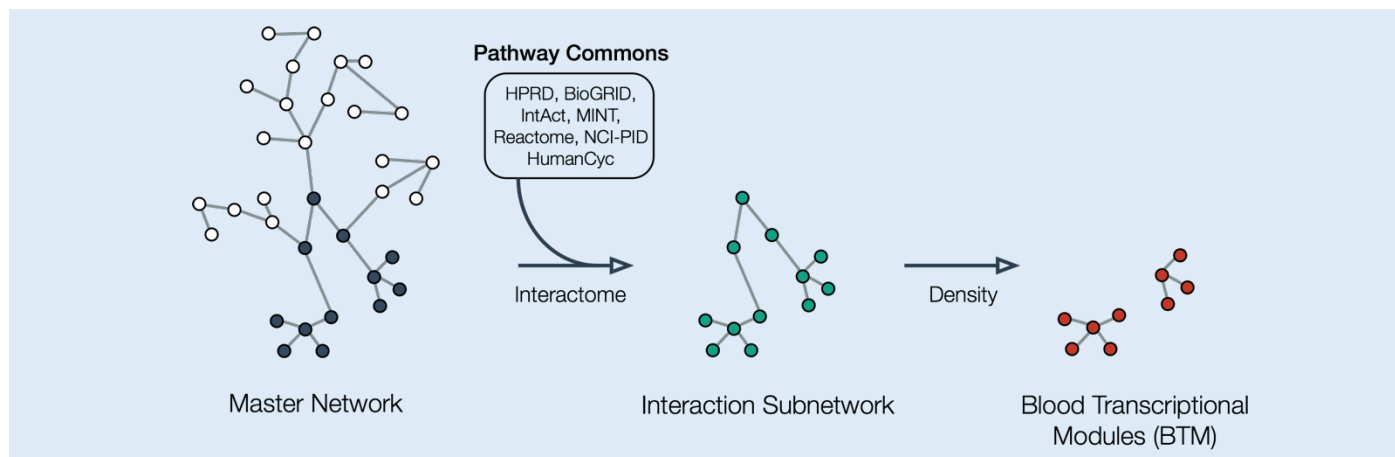
Figure 1. Overview of blood transcription module (BTM) creation. A master network is created from blood transcriptomes gathered from Gene Expression Omnibus; Edges connect genes with high mutual information. An interaction subnetwork is identified by intersecting the master with an 'interactome' network consisting of gene interactions drawn from Pathway Commons. Finally, densely connected BTM are identified within the subnetwork. Process is repeated for other context-specific subnetworks.

This 'master' relevance network is limited in that it is composed of data from heterogeneous studies. To address this, 77 context-specific subnetworks were identified by intersecting related sets of genes from existing knowledge databases (e.g. Gene Ontology, PubMed). For example, A previous survey of transcriptomes from different immune cells[4] was used to identify subnetworks specific for each of 12 immune cell types (e.g. naive B, memory B, T-helper). Li et al. also defined an 'interactome' context, which is a network of gene interactions with over 10 000 genes and 141 000 edges. The interactome used in this study was retrieved from Pathway Commons[5] and includes source data from HPRD, BioGRID, IntAct, MINT, Reactome, NCI-PID, and HumanCyc. Cross-referencing the edges between the master and interactome resulted in a subnetwork with almost 4000 edges.

> *"The interactome used in this study was retrieved from Pathway Commons"*

In the final step, BTM were distilled from the various context-specific subnetworks using a previously developed algorithm[6] to extract regions with relatively high levels of interconnectedness. In all, 334 BTM were identified and many had a biological significance that could be gleaned from inspecting their constituent genes (e.g. 'spliceosome'). Importantly, many BTM represented processes specific to immune tissue and not previously described in other pathway databases, for example, 'T-cell differentiation'.

Do BTM provide any value in describing vaccine responses? For each vaccine transcriptome, Li et al. sifted out the BTM that correlated with antibody production. Unexpectedly, the BTM divided the vaccines into three groups: Those containing a protein conjugate (TIV and MCV4); Those containing a polysaccharide (MPS-V4 and MCV4); and those eliciting a viral response (YF-17D). This supports the notion that predictors of vaccine-induced antibody response are unlikely to be 'universal' but rather, different predictors may be shared between specific vaccine subsets.

Summary

Pathway Commons databases was used for constructing an 'interactome' integrated into a pipeline to distill blood transcriptional modules - gene sets modulated in the response to vaccines. These modules represent processes that underlie a protective antibody response.

References

1.   Pulendran, B., Li, S. & Nakaya, H. I. Systems Vaccinology. *Immunity* **33,** 516–529 (2010).

2.   Haining, W. N. & Pulendran, B. Identifying gnostic predictors of the vaccine response. *Current Opinion in Immunology* **24,** 332–336 (2012).

3.   Li, S. *et al.* Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* **15,** 195–204 (2013).

4.   Abbas, A. R. *et al.* Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* **6,** 319–331 (2005).

5.   Cerami, E. G. *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Research* **39,** D685–D690 (2010).

6.   Bader, G. D. & Hogue, C. W. V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,** 2 (2003).