Finding patterns in protein chip data to probe the immunological mechanisms of nut allergies

Agustin Calatroni, David Hall and Herman Mitchell
Rho, Inc.
Chapel Hill, NC

Emily Lei Kang
Department of Mathematical Sciences
University of Cincinnati
Cincinnati, OH

Project Description

Allergies are caused by the immune system reacting to common harmless substances, like pollen, as if they were pathogenic. These reactions trigger a cascade of physiological responses that may be mild in some individuals and life threatening in others. Medications, such as antihistamines and corticosteroids, can be used to reduce symptoms and can be life-saving during acute episodes. Immunotherapy, e.g. regular injections of increasing amounts of allergen, can enable the immune system to become tolerant of an allergen over time. This is an effective treatment for many environmental allergies. However, effective treatments for food allergies, which are on the rise globally, are lacking. By understanding the molecular mechanisms of the interactions between the immune system and food allergens, we hope to develop immunotherapy-based treatments.

The immune system recognizes portions of proteins called epitopes. For a given allergen, there is typically a finite set of epitopes within the population although different allergic individuals may recognize different subsets of the set of possible epitopes. In this SAMSI project you will help identify the “global” set of epitopes within a group of tree nut and peanut allergic sufferers sampled from multiple cohorts and clinical trials in Europe and the US. The dataset consists of peptide microarray data. Proteins are linear chains of amino acids that adopt a unique 3-dimensional structure, which is determined by the primary sequence. Peptides are subsequences of a protein’s primary sequence. The protein microarrays were created by synthesizing overlapping peptides that span across the entire protein sequence of known peanut and tree nut allergens. These peptides were then fixed to a glass slide and incubated with blood serum collected from subjects. After treating with a fluorescent compound, a microarray reader quantifies the level of reactivity of each subject’s serum to each of the peptides.

Peptide microarray data are very noisy due to nonspecific hybridization, batch effects, and array defects. Your task will be to separate the signal from background noise. The

dataset you will receive will consist of a matrix of hybridization intensities. One dimension represents subjects and the other corresponds to the individual peptides. You will also receive positional information, so you will know the order of peptides within a protein. This may be helpful in that epitopes may span adjacent peptides. Your method should normalize data first and then determine which hybridization readings constitute a true response of a person's immune system to an epitope.

Citation

Cracking Nut-Allergy Mechanisms published in the October 2013 issue of Agricultural Research magazine, United States Department of Agriculture
http://agresearchmag.ars.usda.gov/2013/oct/nut/