Glenn Tesler*, University of California, San Diego, Department of Mathematics, 9500 Gilman Dr., #0112, La Jolla, CA 92093-0112. A decomposition of the Pearson $X^2$ statistic, with applications to pairwise gene interactions.

Recent technological advances have made it feasible to conduct genome-wide scans in large populations to find genetic markers for common diseases and other traits. However, analyzing the data for associations between traits and combinations of multiple markers is computationally challenging. Data has the form of an $n \times m$ table ($n$ individuals, $m$ markers, and a finite number of values allowed in each cell, such as one of 2 haplotypes or one of 3 genotypes). We study a transformation of this table and a decomposition of the Pearson $X^2$ statistic for a contingency table by Irwin (1949) that allows us to efficiently cluster together highly correlated markers. The central idea is a mathematical transformation that maps statistical correlation between locus pairs to distance between two points in a Euclidean space. This enables the use of geometric properties to identify proximal points (correlated locus pairs), without testing each pair explicitly.

This work is joint with Vineet Bafna, Dumitru Brinza, and Matthew Schultz at the University of California, San Diego. (Received January 25, 2011)