

Reviewer's report

Title: Spatially Uniform ReliefF: Increasing the Power to Detect Epistasis in Genetic Association Studies

Version: 1 **Date:** 4 June 2009

Reviewer: Marylyn Ritchie

Reviewer's report:

This is a very interesting paper describing a novel approach for filtering SNPs in GWAS studies. It is very timely given the flood of GWAS data that many human geneticists are dealing with. Because this work is critically important, there are many areas of the current manuscript that should be revised to make it reach the human genetics audience with the most impact. In general, the paper uses a lot of phrases without defining them. A lot of information is in the supplement that should be in the main paper. The reader should not need to read the supplement to understand the main points of the method. Also, there are a lot of missing commas and closing parentheses in the paper.

Abstract:

It says that ReliefF has low power when the interaction effect is small. How small? Many people in the field expect the interaction effects to be small (while others think it will be large) and so they may be put off by that statement right in the abstract and may not read any further. They will think this approach cannot work.

What is the user-selectable parameter? This makes no sense here.

What is meant by "amining"? I cannot find a definition for this word.

The phrase "discover detecting" is awkward and does not make sense.

Background:

There are many more updated references than what is listed in 1,2. Those are ok but there should be newer ones as well.

In the third paragraph where it says "small data sets" it would be good to explain whether you mean small in terms of the number of samples or SNPs.

Page 3, second paragraph. The first sentence is awkwardly worded. It also says that Relief algorithms detect interacting pairs of attributes. Later in the paper it says that it does not detect interacting pairs (page 8). It would also be good to define relief algorithms in this paragraph. The general reader of this will not know.

The statement that stochastic approaches fail without additional information is highly dependent on the types of models being evaluated. Models with main

effects or main effects with interactions can be found without Relief algorithms. Also, purely epistatic models in data with linkage disequilibrium between noise loci and the functional loci can be detected in stochastic algorithms.

What is meant by "a set of assumptions regarding variance"?

When you say a single nearest neighbor, is that one person, one SNP, one person at all of their SNPs? Based on how many SNPs?

Does deleting SNPs with lowest Relief weights make it difficult to detect SNPs with purely interactive effects?

In general, this first Background section is very difficult to follow. A figure would be helpful to explain these different Relief algorithms in this section and the next.

Results:

All of the references to figures are listed as "figure ??"

In the third paragraph you discuss the powers when SURF is used. This is not power. It is really sensitivity or detection.

On page 8 it says "direct replacement of these methods". What methods?

Methods:

The use of 99th, 95th, 75th percentiles in the first paragraph is not clear. Percentiles of what? I can figure it out once I read further, but it would be good to describe it here.

On page 9, it says that we test each method with fixed parameters. What methods? What are the parameters?

How long does it take to run 1000 SNPs? Is it feasible to run 1M? Does the distance metric work in that space?

Tables/Figures

Where are the other 30 penetrance models? They should be provided.

Did you evaluate everything from 99th - 75th percentile?

Figure 2 legend says figure ??

Are these plots an average across the 30 models with 5 shown in each plot?

Figure 1 is not clear. The bar plots need to be better defined. This does not get the point across.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.