Yale University

MB&B
260/266 Whitney Avenue
PO Box 208114
New Haven, CT 06520-8114

Telephone:
203 432 6105
360 838 7861 (fax)
Mark.Gerstein@yale.edu
http://bioinfo.mbb.yale.edu

EASILY ?

17 August 2015

Dear Editor of Nature Methods,

We are submitting a revised version of our manuscript entitled "Analysis of Information Leakage in Phenotype and Genotype Datasets". We also include a revised set of figures, a supplementary material document, and a document with itemized responses to all of the reviewer comments. Briefly, in the revision we have added a number of experimental validations that show the general applicability of the extremity attack under different scenarios. These results show how realistically the extremity attack can be implemented. As per 1st and 3rd reviewer's suggestions, we performed analysis on whether the attacker can analyze the reliability of the linkings that he/she makes. This enables the attacker to focus on the more reliable, linkings, which can escalate the privacy concerns. We show that the attacker can link and characterize as large as 80% of the individuals with very high accuracy.

?

In response to the specific concern of Referee 1 with regard to Im et al 2012 study, we have included a detailed comparison of our manuscript with Im et al study. First of all, privacy is a multifaceted concept and we believe not one study can analyze all the aspects of it. The reviewer, however, has the interesting view that Im et al is the definitive study that encompasses all the privacy breaches related to eQTL studies, which we find rather incomplete. In response to the reviewer's comments, we explained how the two studies address significantly different scenarios in genomic privacy. We believe that as the number and size of genotype and phenotype datasets grow, the attacks that are exemplified in our manuscript will become much more prevalent. Most importantly, we explain that the machinery that is presented in Im et al study is not enough and not suitable for an attacker to perform the attacks that we are studying in our manuscript. In other words, the attack scenario that we are presenting is almost orthogonal to that presented in Im et al study. We also listed a number of technical differences. These show the novelty of our study compared to Im et al study.

WEAK

NOT APPLIC

In addition, we made a comparison to the Schadt et al study and show that our approach can utilize much less information and obtain very high and comparable individual characterization accuracy, which makes our attack scenario much more realistic and applicable. We finally made several schematic figures, added a supplementary materials document, and updates to the manuscript so as to clarify our contributions and different aspects of the linking attacks that we study in our manuscript.

We also added a paragraph to conclusions for summarizing how an individuals privacy get compromised as the result of linking attacks and discussed the possible risk management strategies.

We do realize that the manuscript is above the word limit and we can most certainly reorganize the manuscript to have it fit to the word limits.

Yours sincerely,

Mark Gerstein
Albert L. Williams Professor
of Biomedical Informatics