

From: Vitalii Kleshchevnikov vitalii@ebi.ac.uk  
Subject: Fwd: Benchmarking looks weird but there is some signal  
Date: 30 November 2017 at 16:53  
To:

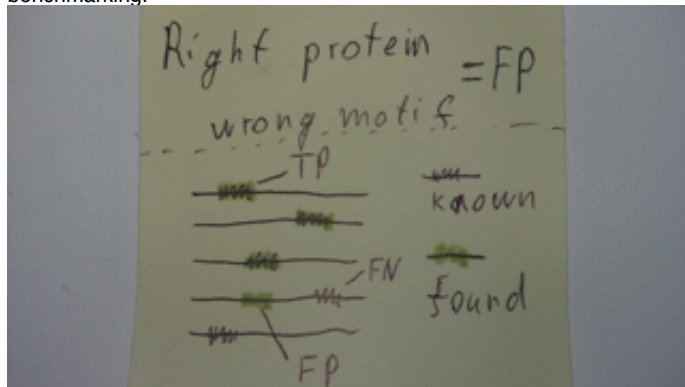
VK

Hello Evangelia,

I would like to give you a short update regarding how the benchmarking has worked.

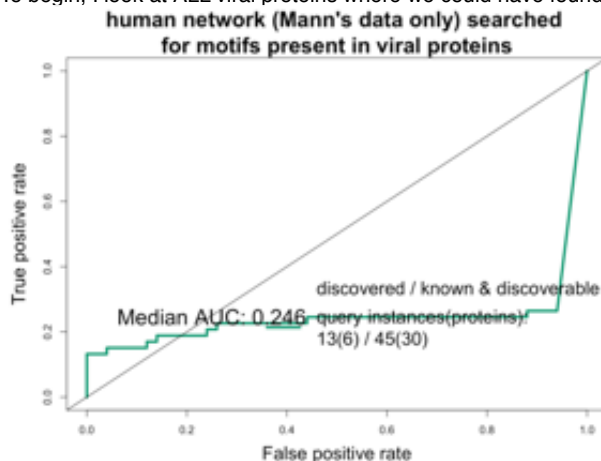
This is the way I did it:

The definition of true positives or false negatives is quite straightforward, false positives and true negatives are harder to define. I define false positives this way: we have found the right protein, but the wrong motif. True negatives are not present in this benchmarking.

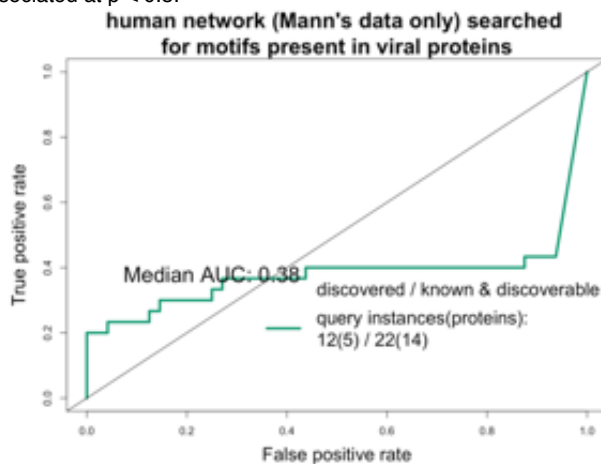


I show only result obtained using protein-centered motif search datasets and Mann's PPI dataset.

1. To begin, I look at ALL viral proteins where we could have found known motifs and where we do predict motifs.



2. Next, I filter motif search datasets based on domain overrepresentation. First, I keep only viral proteins that have a domain associated at  $p < 0.5$ .



3. Second, I keep only viral proteins that have a domain associated at  $p < 0.05$ . Nothing changes suggesting that viral proteins with known ELM instances have domains associated with high significance.

human network (Mann's data only) searched

