

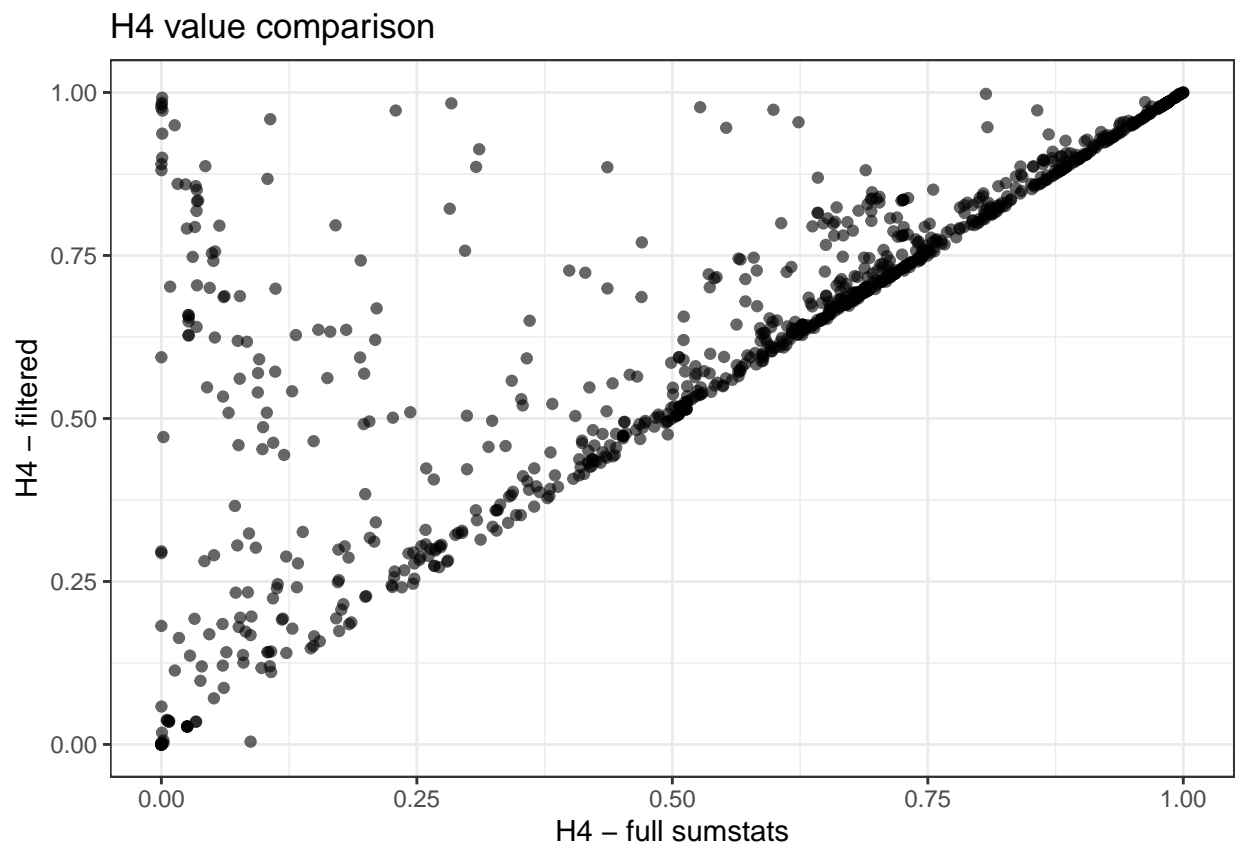
Explore drug QTL concordance

Jeremy Schwartzentruber

18/11/2021

Let's compare the coloc H4 values for the filtered and unfiltered sumstats. We would like them to be highly similar.

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
## [1] 0.8021161
```

Unfortunately, a significant fraction of the time, the H4 values computed from filtered sumstats are over-estimated. Presumably this happens because SNPs that were filtered out in one dataset ($p > 0.05$) had significant p-values in the other dataset, and this would have been informative regarding lack of colocalisation.

This experiment involved filtering sumstats in each dataset separately to those with $p < 0.05$. But the ideal approach would be to filter for cases where $p < 0.05$ in at least ONE of the datasets. But unfortunately,

doing it on a pairwise basis would remove most of the benefit, since you then still have to load full sumstats for each pair.

Therefore, it seems that the approach of pre-filtering sumstats individually will not be a useful optimisation.