Refining reference chemicals and signatures of activity using high throughput transcriptomics for advancing predictive toxicology

Laura Taylor[1], Bryant Chambers[1], Nancy Baker[2], Joshua Harrill[1], Logan Everett[1], Imran Shah[1], Richard Judson[1]

Affiliations: US EPA CCTE[1], Leidos[2]

Objective

Refined gene signatures

=

More effective predictive screening

Having sensitive and specific signatures will help us screen new chemicals for their activity

Methods Overview

Start with:

- Chemical target annotations for about 2,500 chemicals (such as a gene, e.g. ER Estrogen Receptor, AR Androgen Receptor, PR Progesterone Receptor)
- Signature target annotations for around 22,300 gene signatures from MSigDB, BioPlanet, DisGeNET, and CMAP that were annotated as being associated with a specific target (a gene, cell process, illness, etc).

Use high-throughput transcriptomics (HTTr) to optimize for true (T) and false (F) positives (P) and negatives (N):

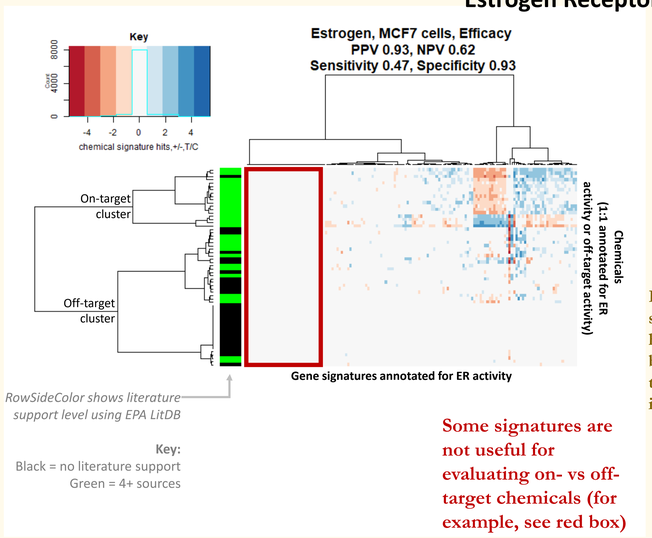
- Sensitivity: (TP / TP + FN)
- Specificity: (TN / FP + TN)
- Positive predictive value (PPV): (TP / TP + FP)
- Negative predictive value (NPV): (TN / TN + FN)

Improve predictive toxicology:

- Increase confidence in reference chemicals for the targets
- Use refined signatures for screening targets of new chemicals

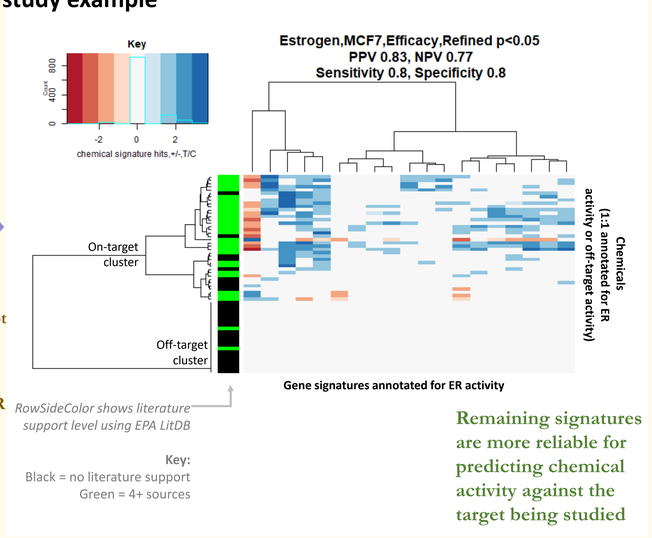
Results

Estrogen Receptor, ER, as a case study example



Some signatures are not useful for evaluating on- vs off-target chemicals (for example, see red box)

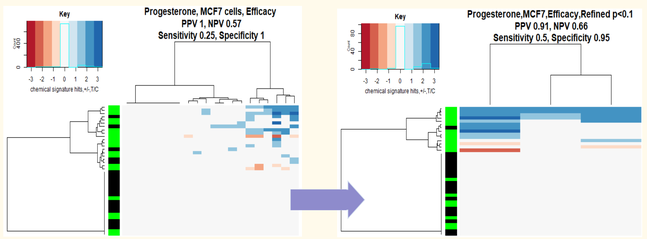
Removed gene signatures that do not help distinguish between chemicals that do versus don't impact the target, ER

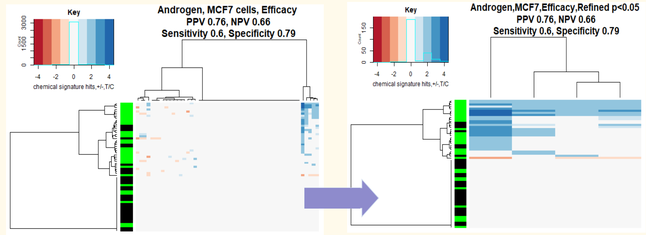


Remaining signatures are more reliable for predicting chemical activity against the target being studied

Figure 1. This heatmap has gene signature sets as the columns and chemicals as rows. Signature scoring was used to evaluate each chemical's activity against the target (ER). The colors correspond to the efficacy level. The row side colors show the level of literature support for that chemical-target association. QC: T/C between -1 and 1 were set to 0, when hitcall was < 0.90, T/C was set to 0.

Figure 2. This heatmap has refined the signature space to just the signatures that are more useful for screening chemicals. The p-values are for t-tests between the on-target and off-target chemicals and were $p < 0.05$ for all retained signatures. Specificity decreased slightly but sensitivity was greatly improved.





Conclusions and Future Work

- Conclusions: this work will help to aid predictions of target genes and cellular pathways that are perturbed by chemicals and thus help to increase the speed and efficiency of screening chemicals for biological hazards when perturbation of those target pathways is known to result in toxic effects.
- CMAP signatures appear to be the most reliable signature source for high-throughput transcriptomics (HTTr).
- Future work: some targets may not be appropriate to evaluate or predict using HTTr data. Thus, future work will include confirming whether specific types of gene targets (for example, nuclear receptors, GPCRs, etc) cause downstream gene signature impacts that are able to detected using this HTTr method.