Evaluating methods for computing points-of-departure with multi-omics data

Jessica Ewald[1], Jo Nyffeler[2], Joshua Harrill[3], Shantanu Singh[1], Anne Carpenter[1]

Affiliations: 1 - Broad Institute of MIT and Harvard, Cambridge, USA, 2 - Helmholtz Centre for Environmental Research, Leipzig, Germany, 3 - US EPA Center for Computational Toxicology and Exposure. **Contact:** jewald@broadinstitute.org

Background

- Transcriptomic points-of-departure (PODs) are a promising approach for integrating toxicogenomics data into risk assessment practices[1].
- There is growing interest in deriving PODs using other omics platforms, for example morphological profiling (Cell Painting) data[2].
- There are many different proposed methods for computing PODs.
- Most POD estimates are a single number with no confidence interval, limiting a statistically rigorous evaluation of modifications made to the experimental design and analytical pipeline. One solution is to use sampling to derive bootstrapped estimates of POD variability[3].
- Here we use a sampling approach to compare five established and two novel methods for computing PODs across:
  - Omics platforms (transcriptomics and Cell Painting)
  - Number of replicates
  - Feature sets
- We compared the technical specifications, failure rate and variability of bootstrapped POD estimates to inform potential use of these methods in risk assessment.

Materials & Methods

Dataset: Transcriptomics and Cell Painting dose-response data[2] were collected from U2OS cells after exposure to 11 model compounds, along with DMSO controls. Each compound concentration-response series had 7 dose groups, with 4 replicates per group. Cytotoxic dose groups were removed.

Methods: The general experimental design and data analysis pipeline for computing PODs (Fig 1A) was modified in four different ways (Fig 1B), resulting in 76 unique "replicate-platform-filter-POD type" scenarios. Replicates were sampled 30 times per compound for each scenario (Fig 1C), and PODs were computed from all samples (76 scenarios * 11 compounds * 30 samples = 25,080 PODs). The resulting POD distributions were compared across scenarios to evaluate different POD methods and to draw conclusions about the impact of various experimental design choices.

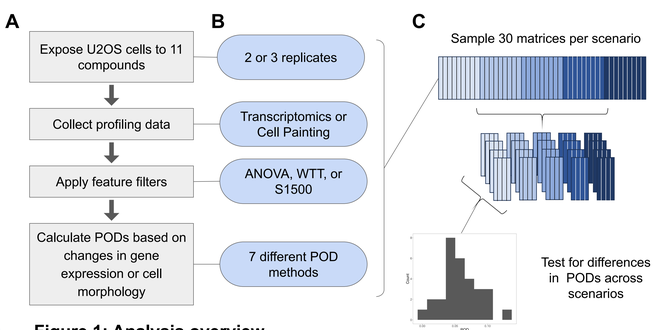
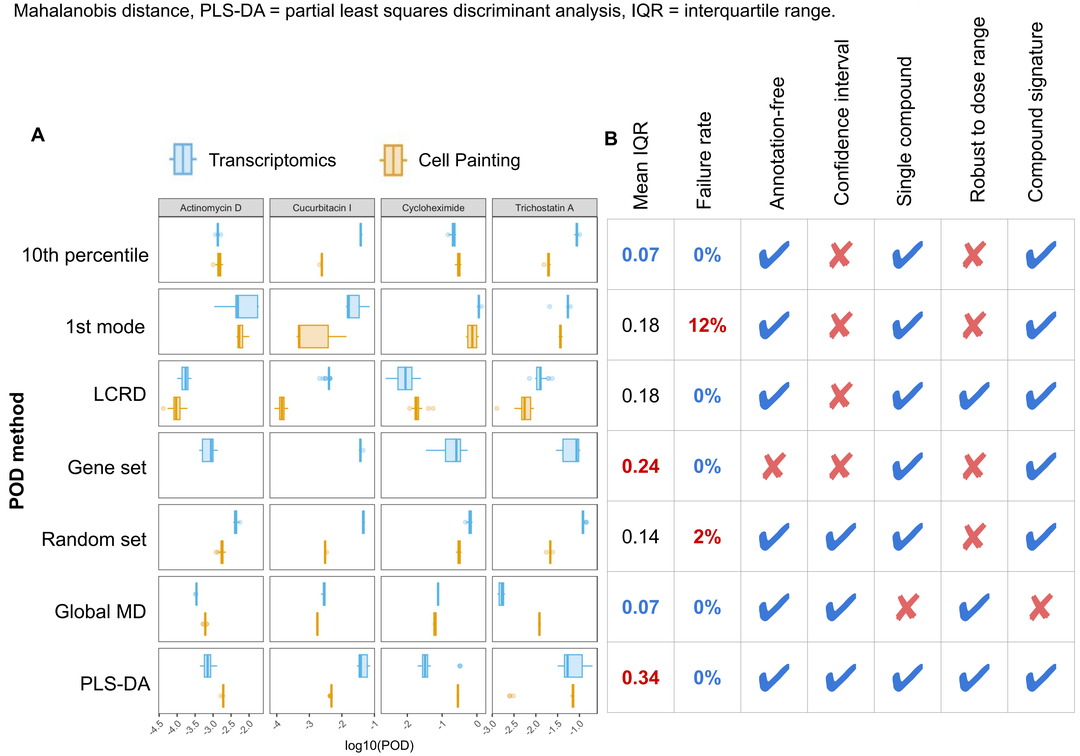


Figure 1: Analysis overview

Results

Figure 2. Comparison and evaluation of POD methods. Distributions of 7 POD types for a representative 4 of 11 compounds, grouped by platform (A) and evaluated according to 7 criteria (B). Both A and B represent scenarios with 3 replicates, the WTT feature filter for transcriptomics data, and the full feature set for Cell Painting data. LCRD = lowest consistent response dose, MD = Mahalanobis distance, PLS-DA = partial least squares discriminant analysis, IQR = interquartile range.



- No single POD method excelled in all evaluation categories, although the **10th percentile, global MD, and random set** methods performed well across both Cell Painting and transcriptomics data for most criteria.
- Cell Painting PODs tended to be lower than the transcriptomic PODs, except for when considering the multivariate methods (global MD and PLS-DA). These methods project features onto an orthogonal basis, suggesting that there could be a technical explanation (multicollinearity) rather than a biological explanation for the observed difference.

Find detailed methods, PODs from the full compound set, and definitions of the POD types and evaluation criteria here:

http://broad.io/dlz0wz

References: [1] Johnson et al., 2022. *Toxicol Sci* 190 (2); [2] Nyffeler et al., 2022. *Toxicol Appl Pharmacol* (444); [3] Ewald et al., 2022. *ES&T* 56 (22)

Results

Table 1. Impact of experimental design modifications on the POD. Significant differences in the mean POD and mean POD IQR between different experimental designs were evaluated with paired t-tests, where each pair had identical experimental design parameters (compound, POD type, replicates, filter, omics platform, and starting feature set), other than the one specified in the "Comparison" column.

Comparison	Mean POD	Mean POD IQR
Replicates (2 vs. 3)	No significant difference	IQR is 3-fold greater for 2 replicates (p-value = 7.4e-12)
Feature filter (ANOVA vs. WTT)	No significant difference	No significant difference
Omics platform (Cell Painting vs. Transcriptomics)	Mean is 1.2-fold lower for Cell Painting (p-value = 2.2e-16)	No significant difference
Starting features (S1500 vs. whole transcriptome)	No significant difference	IQR is 5-fold greater for S1500 (p-value = 0.016)

Conclusions

- Using sampling to derive estimates of POD variability enabled evaluation of significant differences between POD types and experimental design modifications.
- Each POD method had strengths and weaknesses; community consensus on the ideal POD characteristics for different regulatory applications would promote the development of "fit-for-purpose" POD pipelines.
- Future work will investigate explanations for the lower Cell Painting PODs, conduct qIVIVE to make comparisons between omic and apical PODs, and evaluate POD methods on a larger number of compounds, including some compounds with low signal.

Funding Sources

JE is funded by the OASIS Consortium, an SOT Syngenta Fellowship, and a Banting Fellowship.