NTP
National Toxicology Program

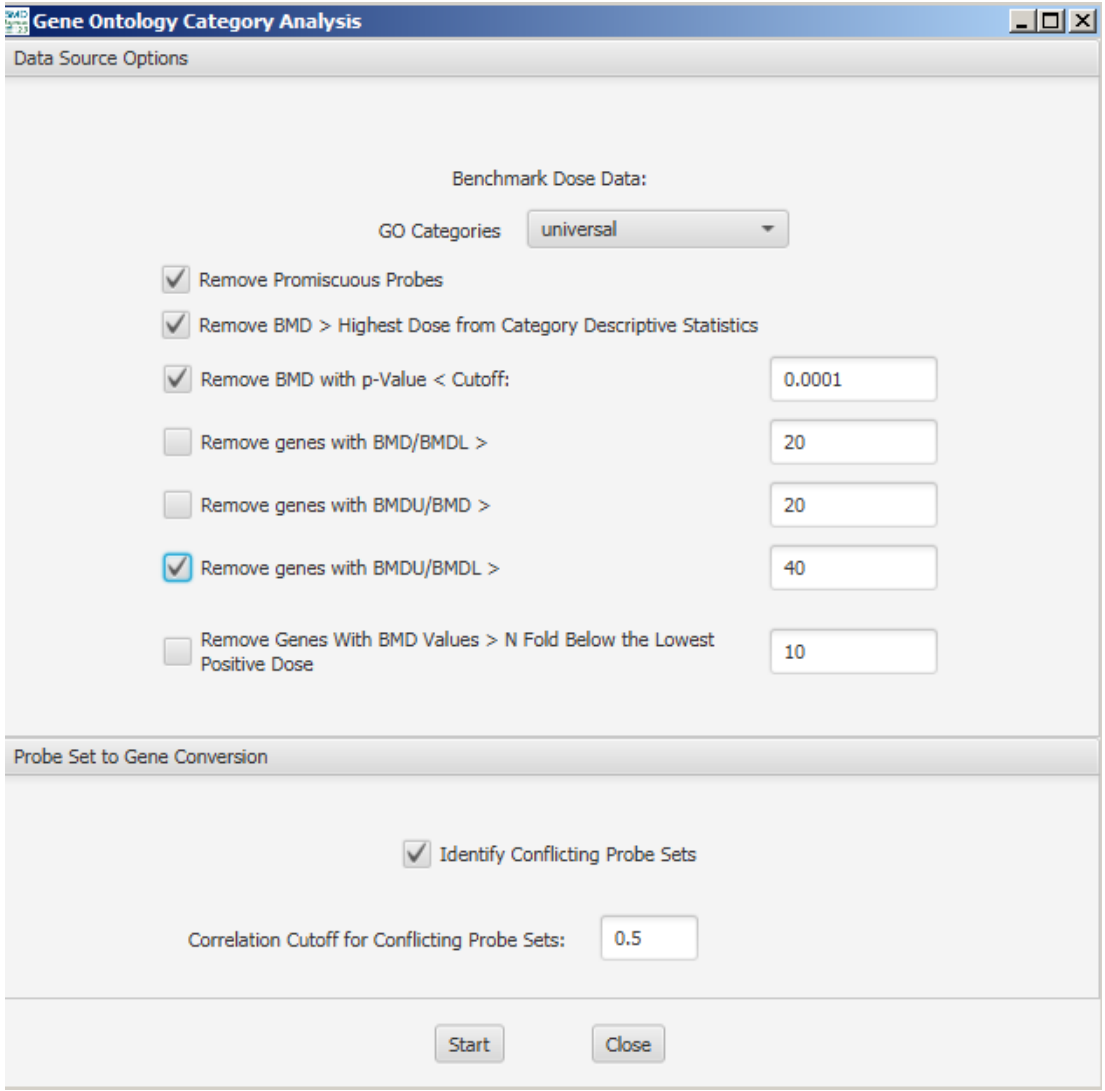
NTP's Proposed Approach to Estimating Gene Set Level Potencies

Scott S. Auerbach Ph.D., DABT
Biomolecular Screening Branch
National Institute of Environmental Health Science

Expert Panel Meeting on the Peer Review of Draft NTP Approach to Genomic Dose-Response Modeling
October 24, 2017

Populating Gene Sets

- For a feature to be considered its best model must…
  - Have convergent BMD, BMD_L and BMD_U values
    - Indicates model parameters are optimized
    - Ensures complete representation of the uncertainty around the BMD
  - Not map to more than one gene
    - Removes features with uncertain gene association
  - Not have a BMD> highest dose
    - Avoids model extrapolation
  - Have a nominal global goodness of fit p-value >0.0001
    - Higher values indicate better fit
    - Ensures a minimum (albeit liberal) fit of the model to the data
  - BMD_U/BMD_L < 40
    - Removes features with highly uncertain BMDs



Global Goodness of Fit P-value

- EPA Guidance
  - Prior model hypothesis, fit p-value > 0.05
  - No prior hypothesis, multiple models, fit p-value > 0.1
- Justification for the lower threshold fit p-value
  - A number of orthogonal filters for removing non-responsive or noisy data are included in the analysis pipeline
    - Fold change and ANOVA
    - BMDU/BMDL ratio <40
    - Gene set level filters - 3 genes, 5% populated, Fisher Exact test p<0.05
  - We use agglomerative estimates of potency
  - Loss of critical information particularly for moderate signal test articles

Identifying Active Gene Sets and Potency

Gene Set 1 (15 genes)

Gene Name	BMD	BMD_L	BMD_U
Gene 1	10	5	25
Gene 2	50	25	70
Gene 3	100	75	120
Gene 4	150	100	175
Gene 5	200	100	210
Gene 6	Failed fit filter	Failed fit filter	Failed fit filter
Gene 7	Failed fit filter	Failed fit filter	Failed fit filter
Gene 8	Failed fit filter	Failed fit filter	Failed fit filter
Gene 9	Failed fit filter	Failed fit filter	Failed fit filter
Gene 10	Failed fit filter	Failed fit filter	Failed fit filter
Gene 11	Failed fit filter	Failed fit filter	Failed fit filter
Gene 12	Failed fit filter	Failed fit filter	Failed fit filter
Gene 13	Failed fit filter	Failed fit filter	Failed fit filter
Gene 14	Failed fit filter	Failed fit filter	Failed fit filter
Gene 15	Failed fit filter	Failed fit filter	Failed fit filter

( ) = Median value = Gene Set BMD, $BMD_{L,U}$

- At least 3 genes
  - Ensure that small gene sets are minimally populated
  - Assuming no prior knowledge, minimum number of genes required to indicate a pathway or gene set is responding to treatment
  - Minimum number of genes from which you can identify a median value
- At least 5% populated
  - Ensure larger gene sets require more than 3 genes
- Fisher Exact Test ($p<0.05$)
  - Nominal Statistical Filter

Points to Consider

- Fit p-value threshold >0.0001
- BMDU/BMDL ratio threshold of <40
- Threshold for “active” gene sets
  - 3 genes, 5% populated and Fisher Exact P-value<0.05
- Determining potency of a gene set
  - Median and Mean BMD
- Other variables to consider
  - GSEA-based approach
  - Bayesian alternative to enrichment
  - Focus only on use of select biomarker genes from AOPNs
  - What if only 2 biomarker genes are active (e.g., p21 and Ccng1)? Ignore?
  - Data supporting use of thresholds of “3 genes, 5% populated and Fisher Exact P-value<0.05”
  - Pathway agnostic agglomerative BMD
    - Median BMD of the 20 most differentially expressed genes