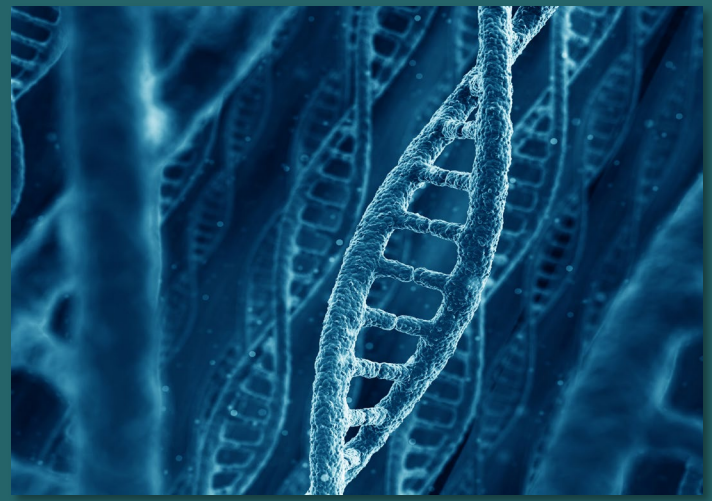
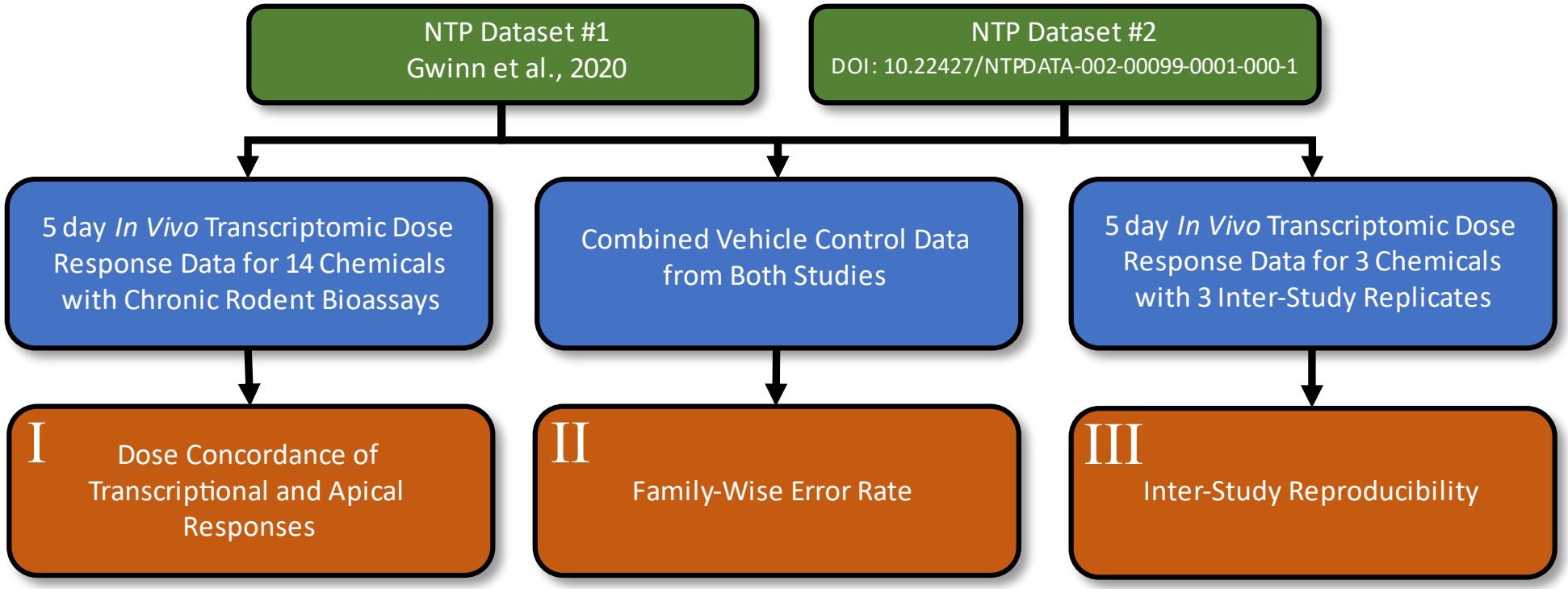
Dose Response Methods & Parameter Refinement

Logan J. Everett, Ph.D. – Bioinformatics Scientist



The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the U.S. EPA.
Company or product names do not constitute endorsement by U.S. EPA.

Overview of Scientific Support Analysis

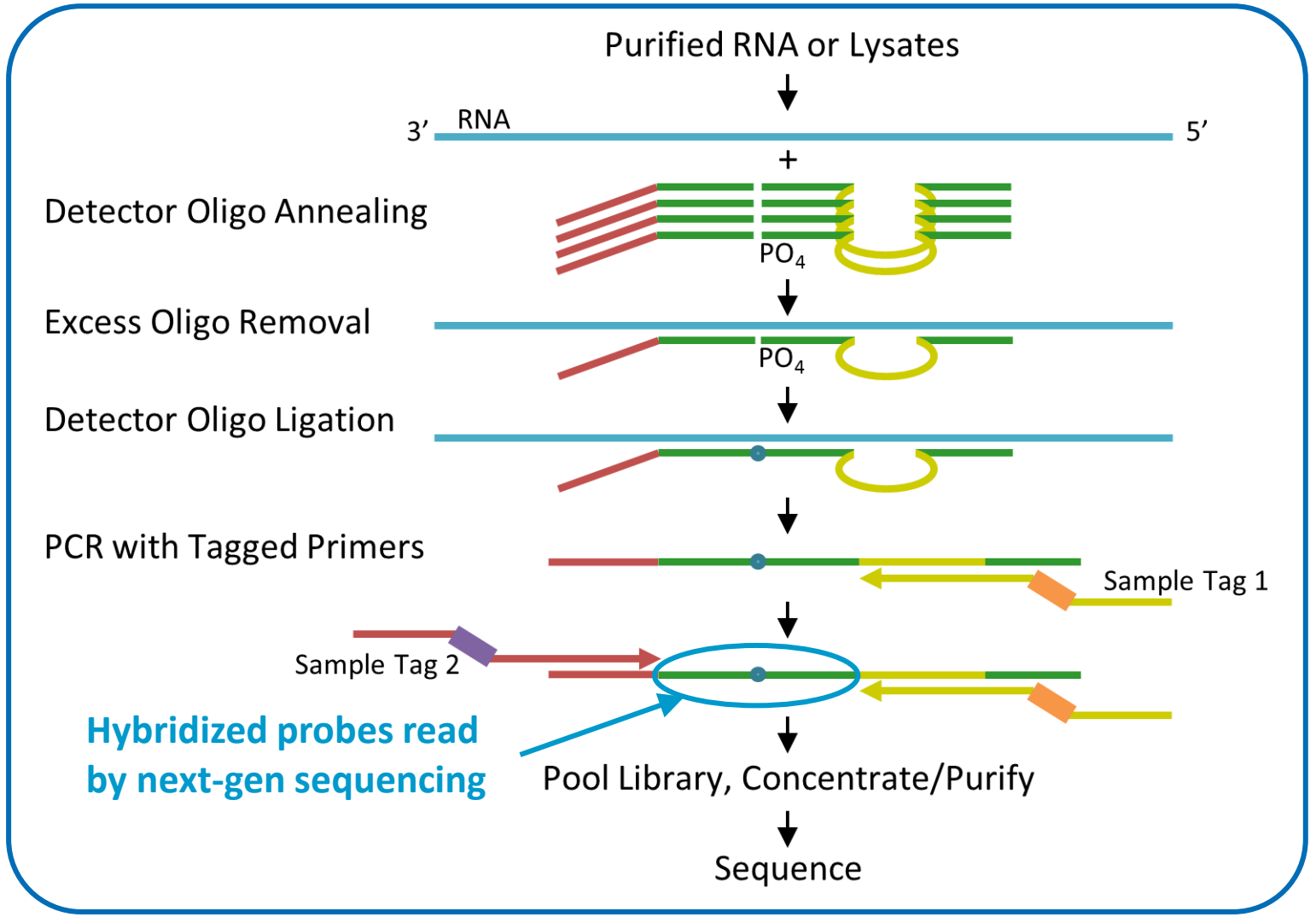


(Figure 4-1)

- Same platform (TempO-seq rat S1500+) and general design (5-day repeat dose) as proposed ETAP method
- All transcriptomic data obtained from NTP in raw FASTQ format and processed using established EPA TempO-seq pipeline *(Harrill, et al. 2021)* & same outlier removal process in standard method document

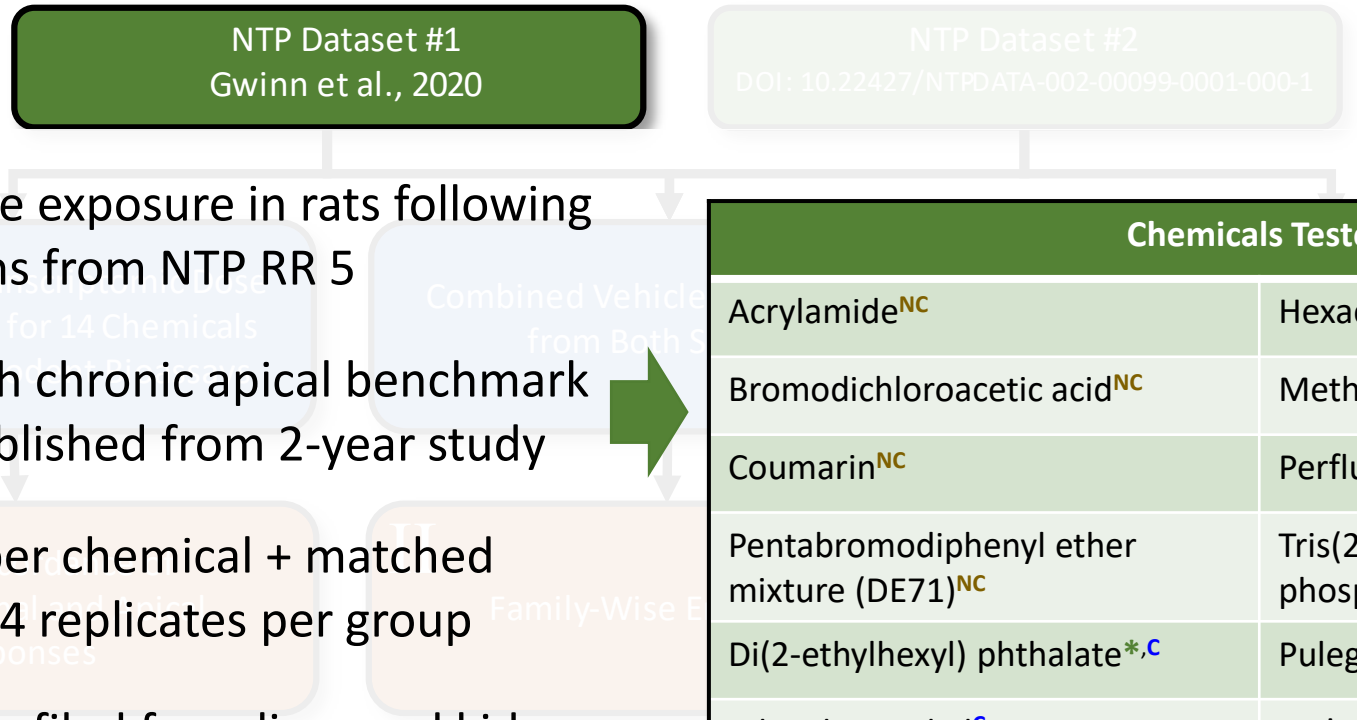
Targeted RNA-seq Assay (TempO-seq)

- Next-gen sequencing of targeted probes hybridized to expressed transcripts
- Captures gene expression at lower cost than RNA-seq or microarrays
- S1500+ probe set designed to maximize biological coverage with ~2,700 genes
  - Mav, et al. PLoS ONE 2018, DOI: 10.1371/journal.pone.0191105
- Using same assay technology for high-throughput *in vitro* screening and other research within ORD
  - Standardized pre-processing & normalization methods
  - J Harrill, et al. *Tox Sci* 2021, DOI: 10.1093/toxsci/kfab009



Yeakley, *et al. PLoS ONE* (2017) DOI: 10.1371/journal.pone.0178302

Overview of Datasets



- 5-day repeat dose exposure in rats following recommendations from NTP RR 5
- 14 chemicals with chronic apical benchmark dose (BMD) established from 2-year study
- 8+ dose groups per chemical + matched vehicle controls, 4 replicates per group
- Transcriptome profiled from liver and kidney in each animal
- Updated 2-year study results for 2 chemicals (marked with *)

Chemicals Tested	
Acrylamide[NC]	Hexachlorobenzene[NC]
Bromodichloroacetic acid[NC]	Methyl eugenol[C]
Coumarin[NC]	Perfluorooctanoic acid[NC]
Pentabromodiphenyl ether mixture (DE71)[NC]	Tris(2-chloroisopropyl) phosphate*,[C]
Di(2-ethylhexyl) phthalate*,[C]	Pulegone[NC]
Ethinyl estradiol[C]	3,3',4,4,'-Tetrachloroazobenzene[C]
Furan[NC]	α,β-Thujone[NC]

C *indicates cancer endpoint was most sensitive BMD*
NC *indicates non-cancer endpoint was most sensitive BMD*
See Table 4-1 for additional details on chronic bioassay results

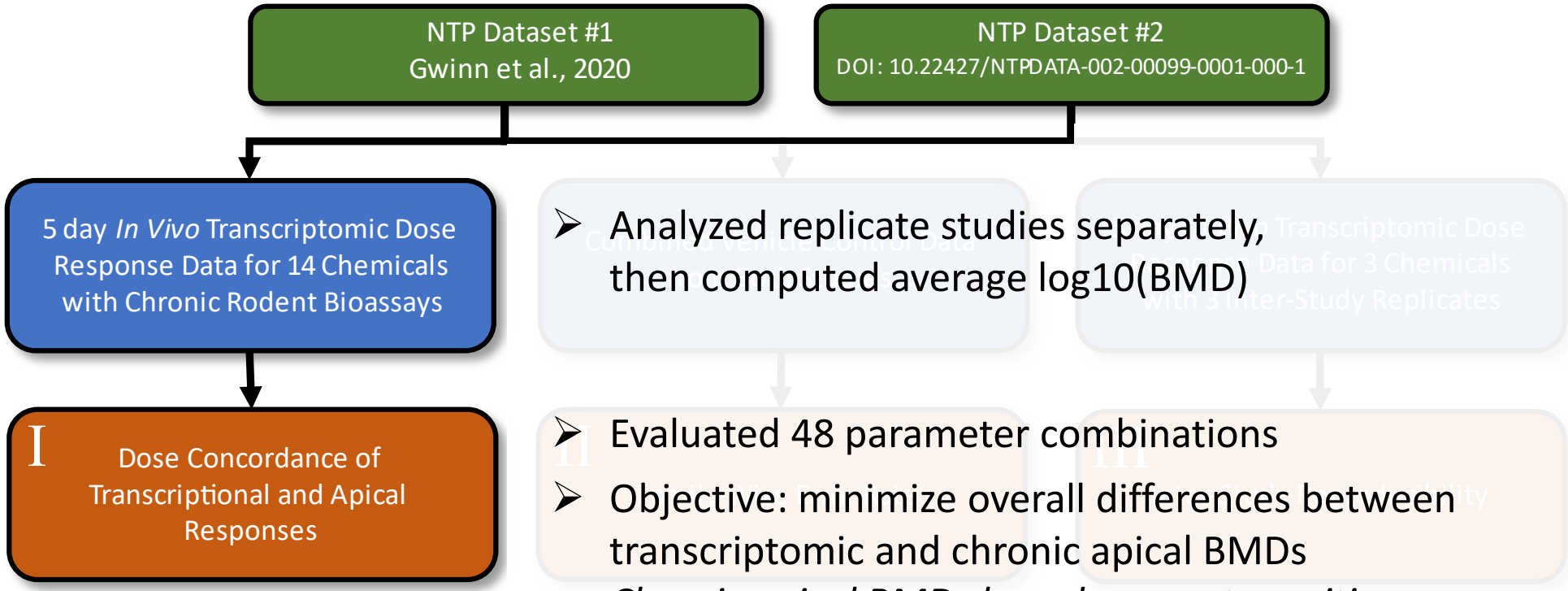
Overview of Datasets

- 5-day transcriptomic studies replicated for 3 chemicals in Gwinn, et al.
- 2 additional replicate studies per chemical
- All replicate studies performed with same doses, in same contract lab

NTP Dataset #2
DOI: 10.22427/NTPDATA-002-00099-0001-000-1

Chemicals Replicated	
Acrylamide	Hexachlorobenzene
Bromodichloroacetic acid	Methyl eugenol
Coumarin	**Perfluorooctanoic acid**
Pentabromodiphenyl ether mixture (DE71)	Tris(2-chloroisopropyl) phosphate*
Di(2-ethylhexyl) phthalate*	Pulegone
Ethinyl estradiol	3,3',4,4,'-Tetrachloroazobenzene
Furan	α,β-Thujone

Part I: Dose Concordance



- Analyzed replicate studies separately, then computed average log10(BMD)
- Evaluated 48 parameter combinations
- Objective: minimize overall differences between transcriptomic and chronic apical BMDs
  *Chronic apical BMDs based on most sensitive **cancer or non-cancer** endpoint*

Analysis Workflow

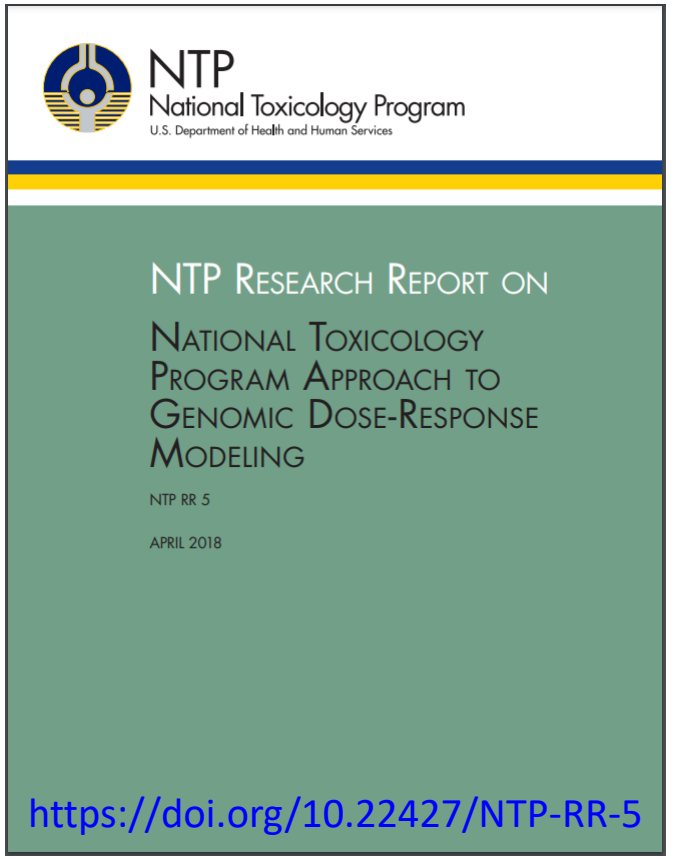
Aligned with the NTP research report, four main steps:

1. Evaluate dataset for adequate signal

2. Pre-modeling filtering for dose-responsive probes

3. Dose-response modeling of individual probes
   - Fit 8 different parametric models
   - Best-fit model selected for each probe based on AIC

4. Summarization of BMD(L) for known gene sets
   - All Gene Ontology Biologic Process gene sets were used
   - Gene Set BMD(L) = median of all valid gene-level BMD(L) values within set
   - **Overall BMD(L) = Minimum Gene Set BMD(L)**



Concordance of Transcriptomic vs Chronic Apical BMDs

- For each chemical, transcriptomic BMD(L) = minimum gene set BMD from either tissue (liver, kidney) and corresponding BMDL

- Evaluated Root-Mean-Square Difference (RMSD):

$$\boldsymbol{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}(\boldsymbol{Y_i} - \boldsymbol{X_i})^2}{\boldsymbol{N}}}$$

  - X_i = log10 transcriptomic BMD(L)
  - Y_i = log10 chronic apical BMD(L)
  - N = 14 chemicals

- Also assessed Pearson Correlation of transcriptomic vs chronic apical log10 BMD(L)s

BMDExpress Parameter Space

Tested 48 different combinations of analysis parameters, focused on those most likely to be dependent on platform & study design:

- Pre-modeling probe filtering
  - William's Trend Test p-value ≤ 0.05 or 0.1
  - Minimum absolute fold-change ≥ 1.5 or 2
- Dose response modeling
  - BMR = 1.349 * S.D. (10% increase in risk when direction is unknown *a priori*)
  - Maximum uncertainty: BMD/BMDL ≤ 20 or BMDU/BMDL ≤ 40
- Gene set (GO Biological Process) summarization
  - Minimum genes per set: 3 or 5
  - Minimum percent coverage: 0%, 3%, or 5%

Concordance of Transcriptomic vs Chronic Apical BMDs

- 13 of 48 parameter combinations produced transcriptomic BMD values for all 14 chemicals
  - *Focused on these combinations to ensure sufficient sensitivity*
- Computed RMSD and correlation for all 13 combinations of BMDExpress parameters
- RMSD values ranged from **0.567** to **0.958** (log10 mg/kg-d)
- Pearson correlations ranged from **0.804** to **0.917**

Top 5 Parameter Combinations by RMSD

Rank	Pre-Modeling Probe Filtering, BMD Modeling, and Gene Set Summarization Parameter Combination	Pearson Correlation Coefficient (PCC)	RMSD (log_{10} mg/kg-d)
1	**Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%**	**0.910**	**0.567**
2	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%	0.907	0.571
3	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.905	0.578
4	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 5%	0.906	0.581
5	Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.905	0.593

(Table 4-3)

Top 5 Parameter Combinations by RMSD

Rank	Pre-Modeling Probe Filtering, BMD Modeling, and Gene Set Summarization Parameter Combination	Pearson Correlation Coefficient (PCC)	RMSD (log_{10} mg/kg-d)
1	**Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%**	**0.910**	**0.567**
2	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%	0.907	0.571
3	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.905	0.578
4	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 5%	0.906	0.581
5	Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.905	0.593

(Table 4-3)

Consistent parameters:

- Pre-filter for probes with maximum fold change (FC) > 1.5
- Maximum uncertainty in best-fit model: BMD/BMDL < 20
- Valid gene set BMD must have minimum of 3 valid gene BMDs

Top 5 Parameter Combinations by RMSD

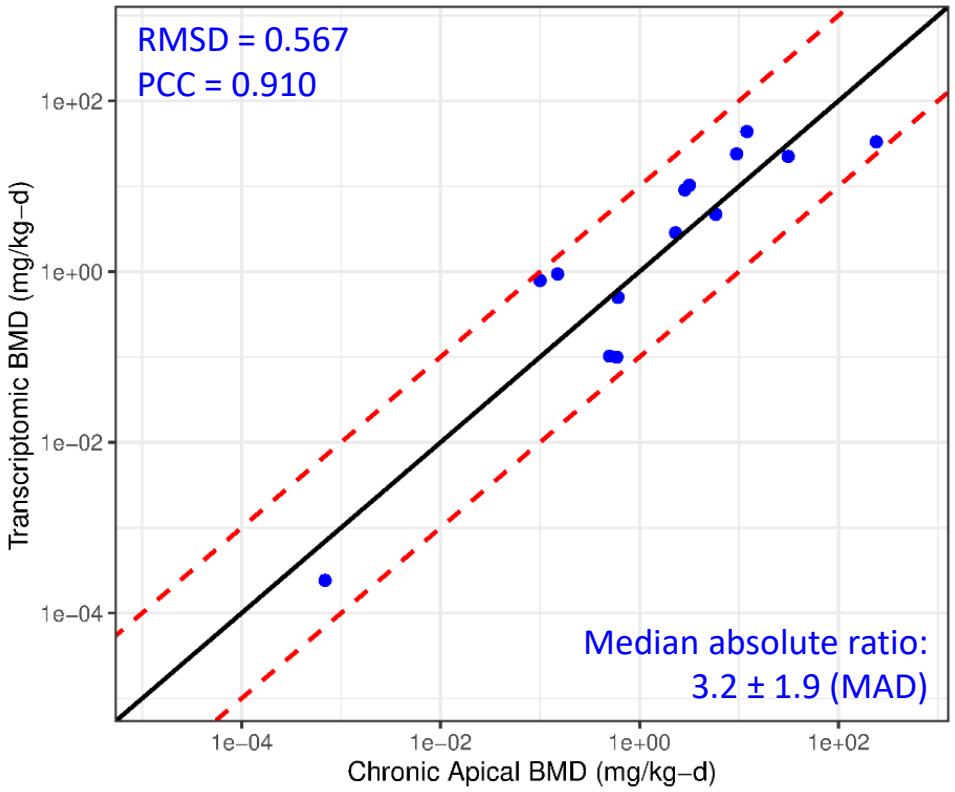
Rank	Pre-Modeling Probe Filtering, BMD Modeling, and Gene Set Summarization Parameter Combination	Pearson Correlation Coefficient (PCC)	RMSD (log_{10} mg/kg-d)
1	**Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%**	**0.910**	**0.567**
2	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%	0.907	0.571
3	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.905	0.578
4	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 5%	0.906	0.581
5	Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.905	0.593

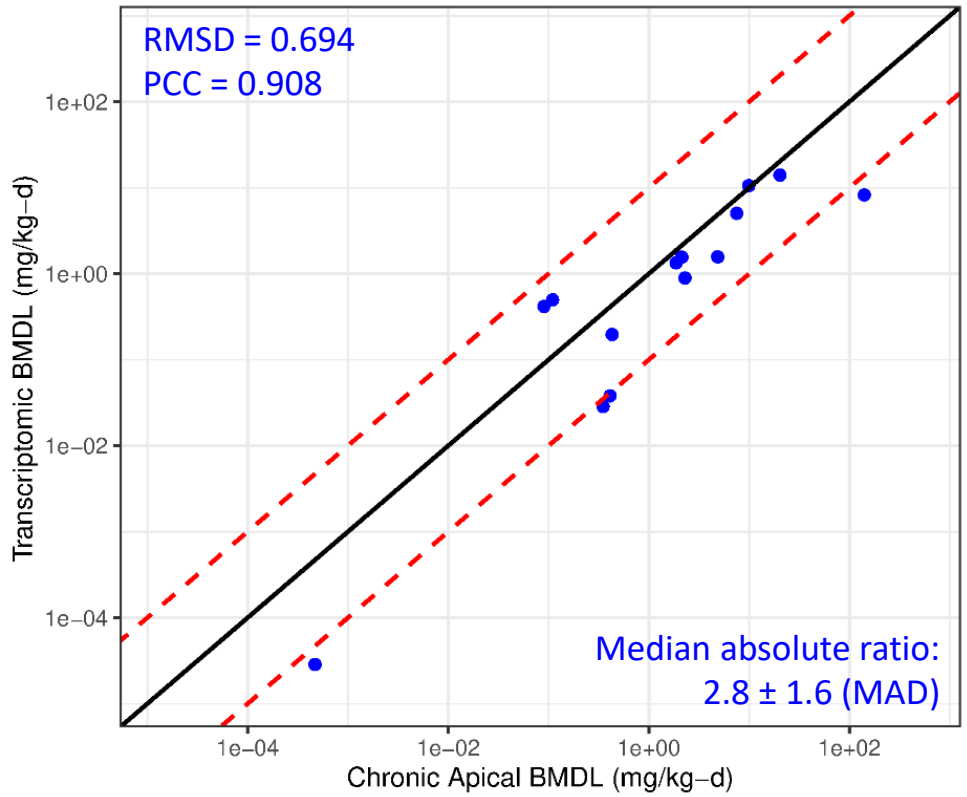
(Table 4-3)

Variable parameters:

- William's Trend Test p-value cutoff for probe pre-filtering
- Minimum percent coverage of valid gene set (0, 3, or 5%)

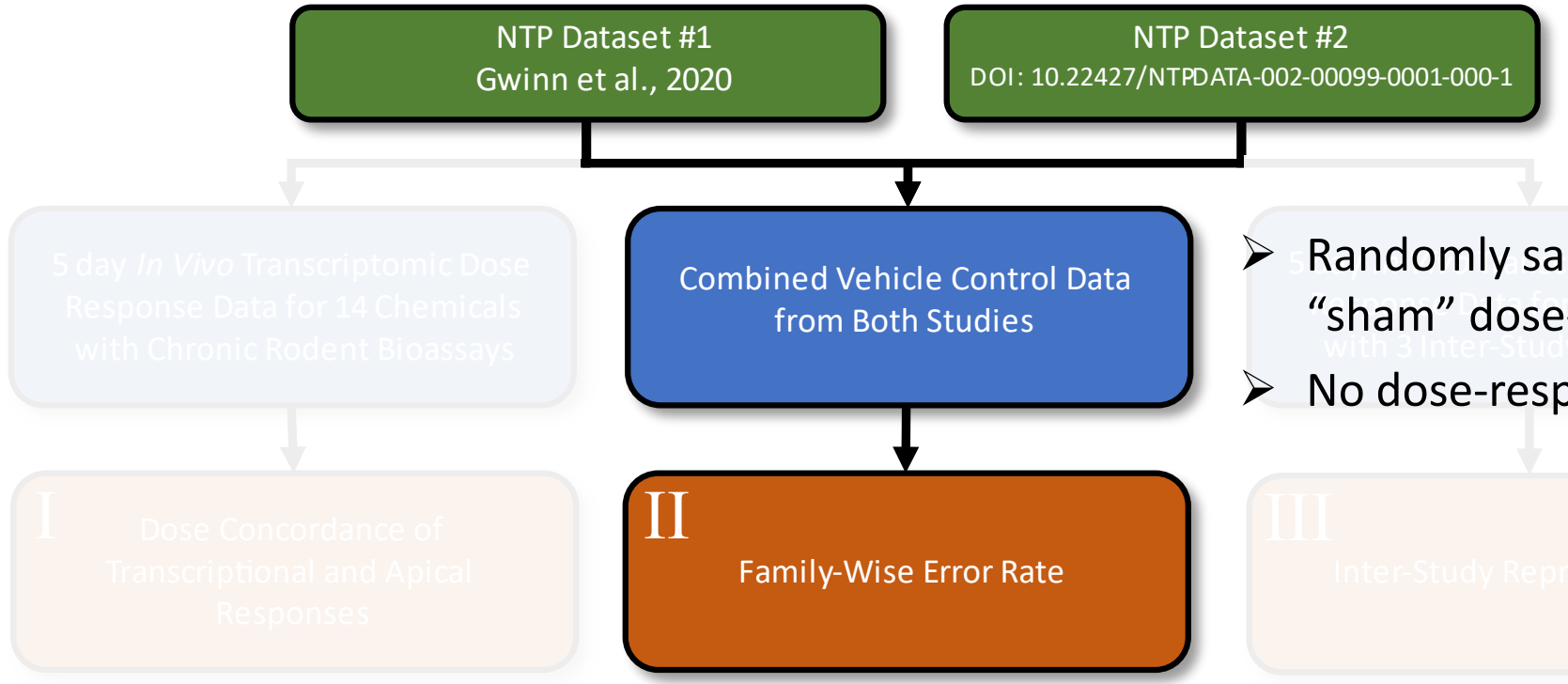
Concordance of Transcriptomic vs Chronic Apical PODs





Scatter plot of $\log_{10}$ transcriptomic BMD(L) versus chronic apical $\log_{10}$ BMD(L) values for the top ranked combination of parameters (Williams $p < 0.05$; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0% coverage). The black line is 1:1 concordance. The red lines are $\pm$ 10-fold. Values below the black line indicate the transcriptomic BMD(L) value is less than the chronic apical BMD(L) value. (Figures 4-2, 4-3)

Part II: Family-Wise Error Rate



- Randomly sampled to create "sham" dose-response series
- No dose-responsive genes expected

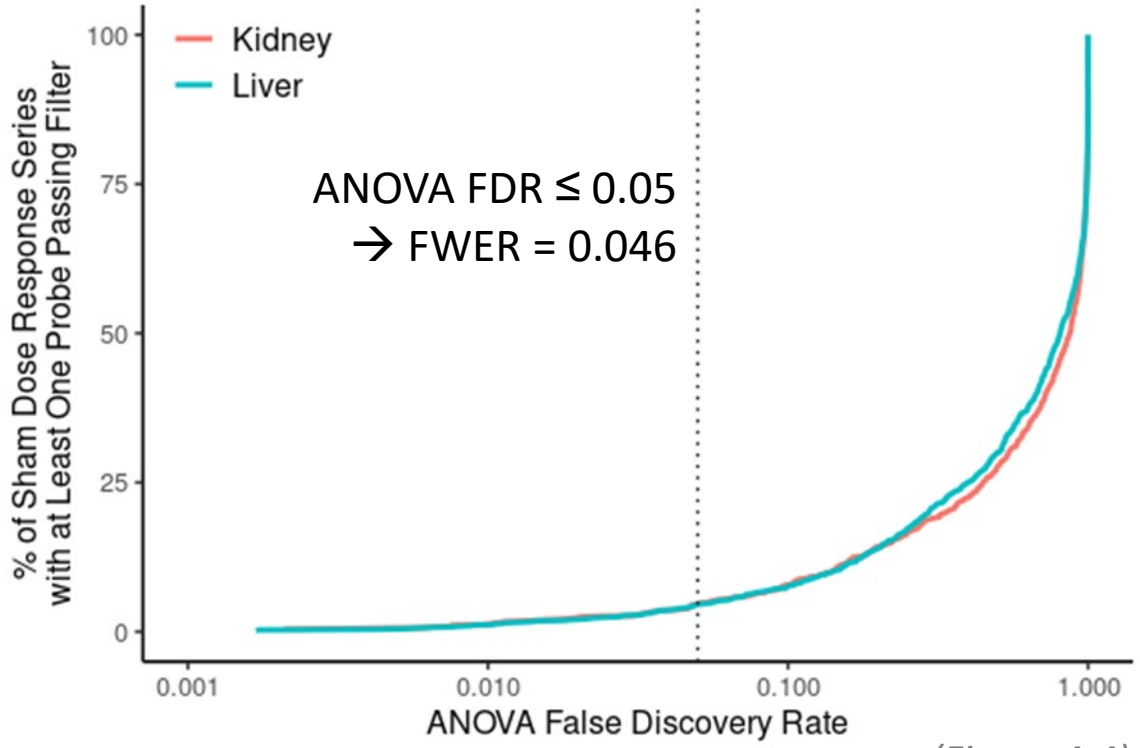
If no real dose-dependent effect, how often would a dataset:

- Pass dataset pre-filtering criteria?
- Generate a valid gene set BMD passing all filters?

Generating “Sham” Dose-Response Series

- Used corn oil vehicle control samples from 14 distinct studies
  - 53 total samples per tissue after sample-level QC
- Randomly sampled 36 vehicle control replicates from same tissue
  - Each series = vehicle controls + 8 dose groups, 4 replicates per group
  - 53 choose 36 = 3.2E13 possible combinations
  - Used dose values from each of the 14 chemical studies in Analysis Part I
- Generated 1,000 “sham” dose-response series for each tissue
  - Applied workflow to each sham series, starting with ANOVA test

Family-wise Error Rate (FWER)



(Figure 4-4)

Overall FWER:

Sham dose response series with at least one probe passing ANOVA 5% FDR filter were run through complete workflow to determine % of sham series producing at least one valid gene set BMD/L
(for the top 5 parameter combinations)

- Dataset-level FWER = % of sham dose-response series with 1+ probe passing ANOVA test
- Probe-level False Discovery Rate (FDR) based on Benjamini-Hochberg corrected p-values
- Dotted line marks FDR ≤ 0.05, corresponding FWER = 0.046

Overall Family-wise Error Rate (FWER)

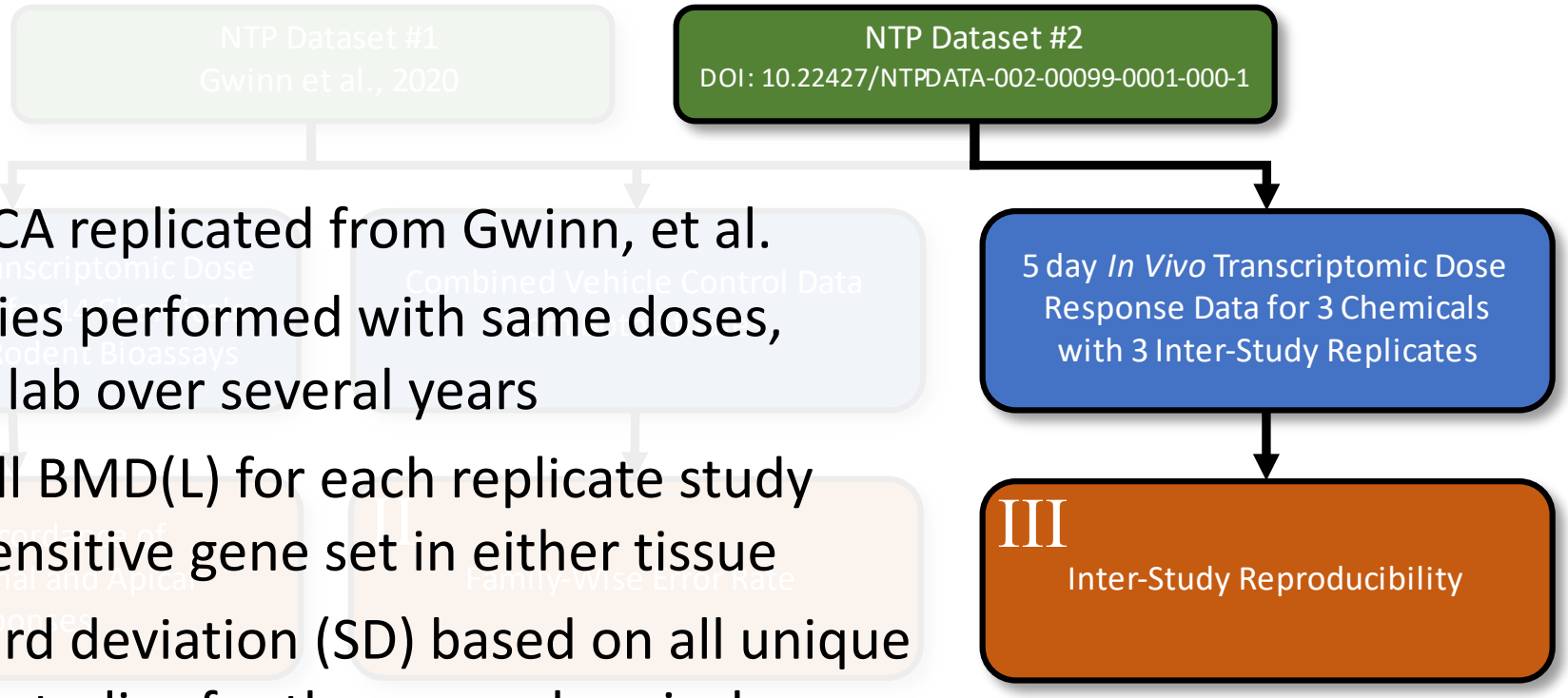
Rank	Pre-Modeling Probe Filtering, BMD Modeling, and Gene Set Summarization Parameter Combination	Overall Family-Wise Error Rate
1	**Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%**	**0.006**
2	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%	0.009
3	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.002
4	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 5%	0.002
5	Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.001

(Table 4-5)

- Overall FWER: If dataset contains no real dose-dependent effect, how often would we assign a final BMD from complete workflow?

- **Top 5 parameter combinations all have overall FWER < 1%**

Part III: Inter-Study Reproducibility



- Furan, PFOA, BDCA replicated from Gwinn, et al.
- All replicate studies performed with same doses, in same contract lab over several years
- Computed overall BMD(L) for each replicate study based on most sensitive gene set in either tissue
- Evaluated standard deviation (SD) based on all unique pairs of replicate studies for the same chemical:

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - X_i)^2}{2N}}$$

Inter-Study Reproducibility

Rank	Pre-Modeling Probe Filtering, BMD Modeling, and Gene Set Summarization Parameter Combination	Log_{10} BMD SD (log_{10} mg/kg-day)	Log_{10} BMDL SD (log_{10} mg/kg-day)
1	**Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%**	**0.242**	**0.295**
2	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 0%	0.247	0.292
3	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.245	0.290
4	Williams p < 0.1; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 5%	0.241	0.289
5	Williams p < 0.05; FC > 1.5; BMD/BMDL < 20; min 3 genes; min 3%	0.242	0.289

(Table 4-4)

- Evaluated Standard Deviation (SD) for the top 5 configurations from Analysis Part I
- Differences in SD were negligible between these configurations

Summary

Combined transcriptomic data for 14 chemicals from Gwinn, et al. 2020 and replicate studies, performed three analyses to refine & validate workflow:

I. Demonstrated concordance with chronic apical BMDs from 2-year studies & refined BMDExpress parameters to minimize RMSD

II. Evaluated family-wise error rate (FWER) using sham series, demonstrated FWER < 1% using all workflow filters

III. Evaluated inter-study reproducibility using replicate transcriptomic studies for 3 chemicals

- **Up Next:** Further evaluation of chronic apical vs transcriptomic concordance in the context of inter-study variability