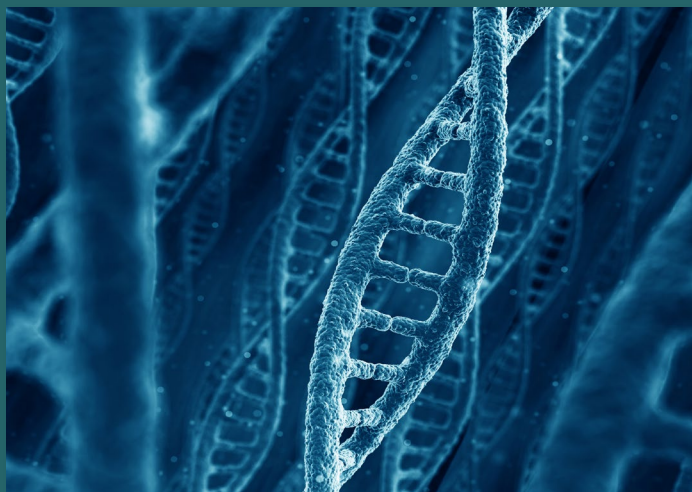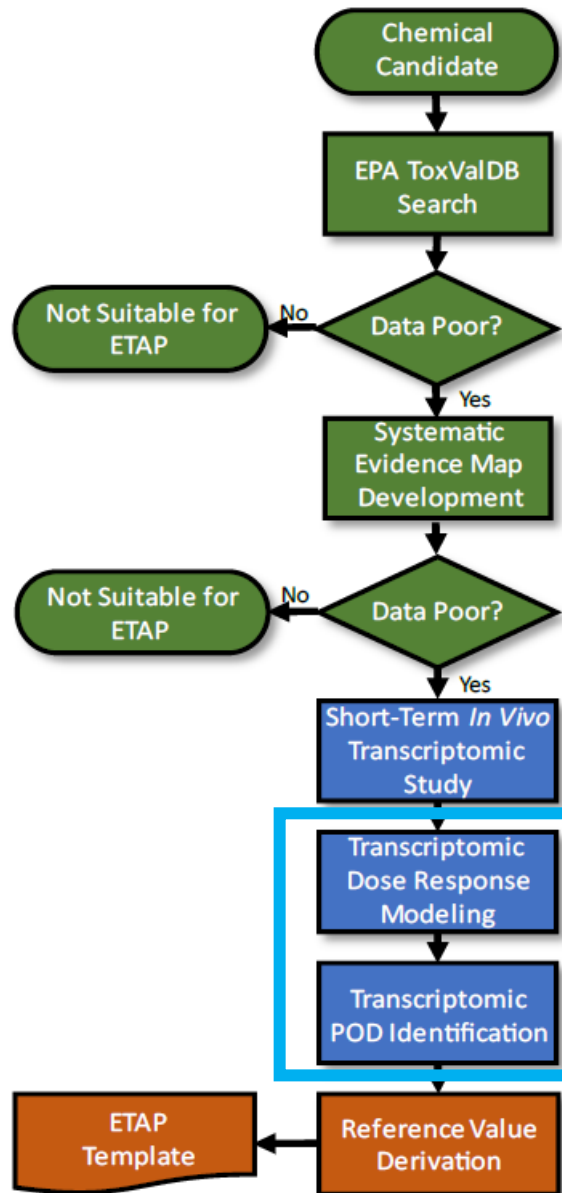# Transcriptomic Dose Response Analysis

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.

Analysis Overview

I. Data pre-processing and quality control
  1. Align reads to known probe sequences
  2. Compute sample-level QC metrics
  3. Normalize to log2(CPM) scale
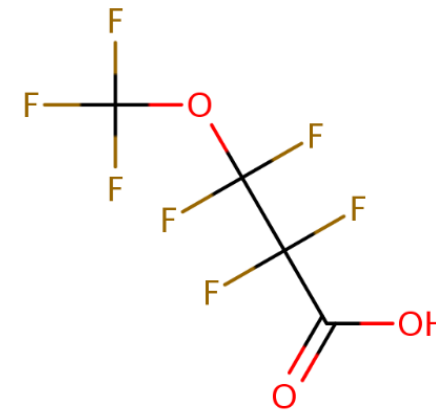 4. Identify and remove outliers

II. Dose Response Modeling & Summarization *(BMDExpress v2.3)*
  1. Evaluate dataset for adequate signal
  2. Filter for dose-responsive probes
  3. Dose-response modeling of filtered probes
  4. Summarization of BMD(L) for known gene sets

Example ETAP Report

Tested Chemical:
Perfluoro-3-Methoxypropanoic Acid (MOPA)

- Tested at 9 doses ranging from 0.01 to 300.0 mg/kg-day
  - 104 animals profiled total (52 male, 52 female)

- Transcriptomic profiling both sexes, 12 tissues total
  - TempO-seq S1500+ platform covering ~2,600 genes

I. Data Pre-processing & Quality Control

1. Raw sequencing reads aligned to known probe sequences (S1500+)
   - Reproducible pipeline with rigorous QC criteria developed for TempO-seq (Harrill et al. *Tox Sci* 2021 DOI: 10.1093/toxsci/kfab009)
   - Pipeline and QC procedures used across multiple EPA/CCTE projects

2. Sample-level quality metrics computed:
   - Total mapped reads → remove samples with < 10% of target depth
   - % reads uniquely aligned → Remove samples with < 50%
   - Total probes with 5+ reads → Remove samples with < 1,200 probes
   - Signal distribution – no cutoff applied, but used in outlier review
     - N_{80} = Minimum number of probes capturing 80% of total reads

Sequencing QC by Sample

Tissue	Animal ID	Sex	Dose Group	QC Issue
Liver	108	Male	Vehicle	< 50% reads aligned
Heart	123	Male	0.1 mg/kg/day	< 50% reads aligned
Brain	143	Male	10 mg/kg/day	< 50% reads aligned
Adrenal	149	Male	30 mg/kg/day	< 50% reads aligned
Spleen	150	Male	100 mg/kg/day	< 50% reads aligned
Thymus	180	Female	0.3 mg/kg/day	*Library preparation failed*
Kidney	211	Female	300 mg/kg/day	*Library preparation failed*

7 / 1,092 (< 1%) samples failed to yield high quality TempO-seq data

I. Data Pre-processing & Quality Control

3. Normalize counts to log2(CPM+1) for BMDExpress input
   - Jointly normalized all samples for same study, tissue, and sex
   - Remove all samples failing initial QC
   - Remove all probes with mean count < 5 across same study/tissue

4. Review PCA of log2(CPM+1) values for each study, tissue, and sex to identify outliers

Sample Outlier Detection

- Strong outliers always removed
  - Distance to all other samples >2x the span of all other samples (PC1 or PC2)

- Moderate outliers removed if:
  - Vehicle sample outliers on both vehicle PCA and dose-response PCA
  - QC metrics below matching samples
  - Outliers in both PC1 and PC2 with relatively large Euclidean distance
  - Larger distance from corresponding replicates or adjacent dose groups

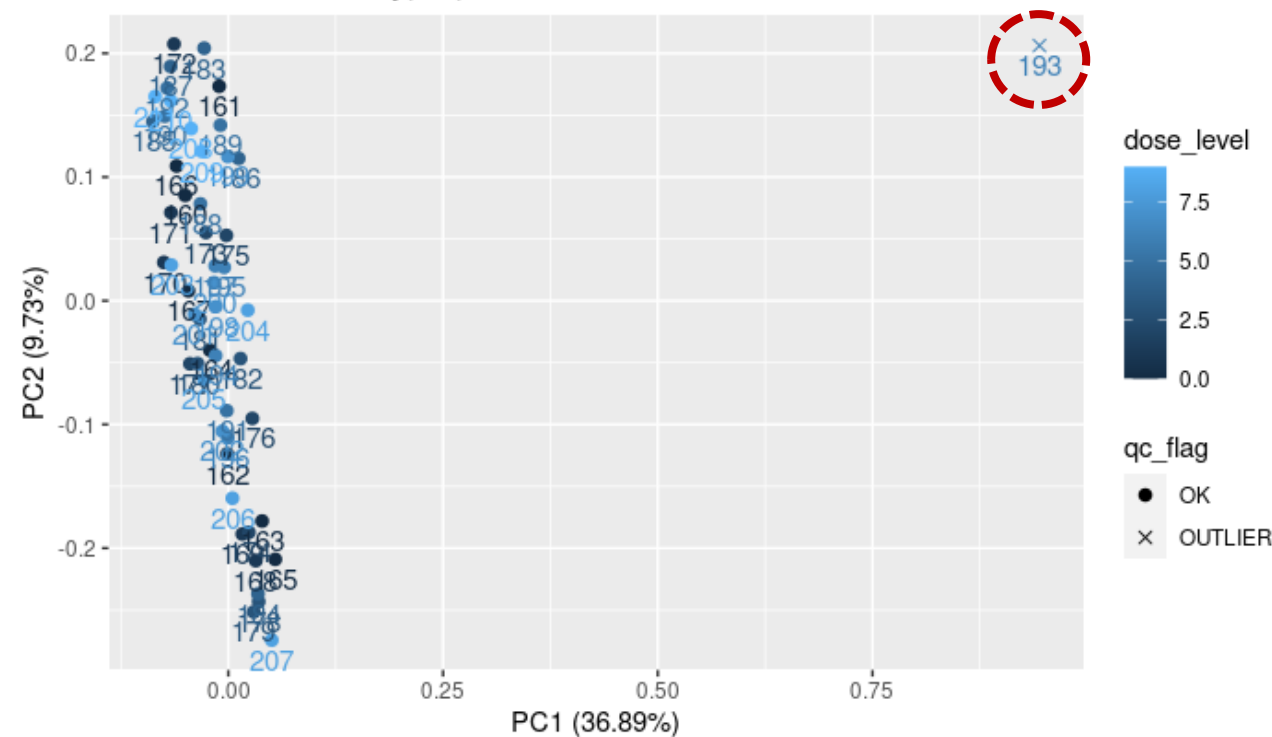**Removed 10 / 1,085 (< 1%) samples based on PCA outliers**

Tissue	Animal ID	Sex	Dose Group
Spleen	210	Female	300 mg/kg-day
Thyroid	162	Female	Vehicle
Adrenal Gland	128	Male	0.3 mg/kg-day
Adrenal Gland	193	Female	10 mg/kg-day
Thymus	115	Male	0.01 mg/kg-day
Thymus	167	Female	Vehicle
Brain	113	Male	Vehicle
Testis	131	Male	1 mg/kg-day
Ovary	191	Female	3 mg/kg-day
Uterus	165	Female	Vehicle

(Example ETAP, Table 8-5)

Sample Outlier Detection

- Strong outliers always removed
  - Distance to all other samples >2x the span of all other samples (PC1 or PC2)


Perfluoro-3-methoxypropanoic acid Female Adrenal Gland

Removed 10 / 1,085 (< 1%) samples based on PCA outliers

Tissue	Animal ID	Sex	Dose Group
Spleen	210	Female	300 mg/kg-day
Thyroid	162	Female	Vehicle
Adrenal Gland	128	Male	0.3 mg/kg-day
Adrenal Gland	193	Female	10 mg/kg-day
Thymus	115	Male	0.01 mg/kg-day
Thymus	167	Female	Vehicle
Brain	113	Male	Vehicle
Testis	131	Male	1 mg/kg-day
Ovary	191	Female	3 mg/kg-day
Uterus	165	Female	Vehicle

(Example ETAP, Table 8-5)
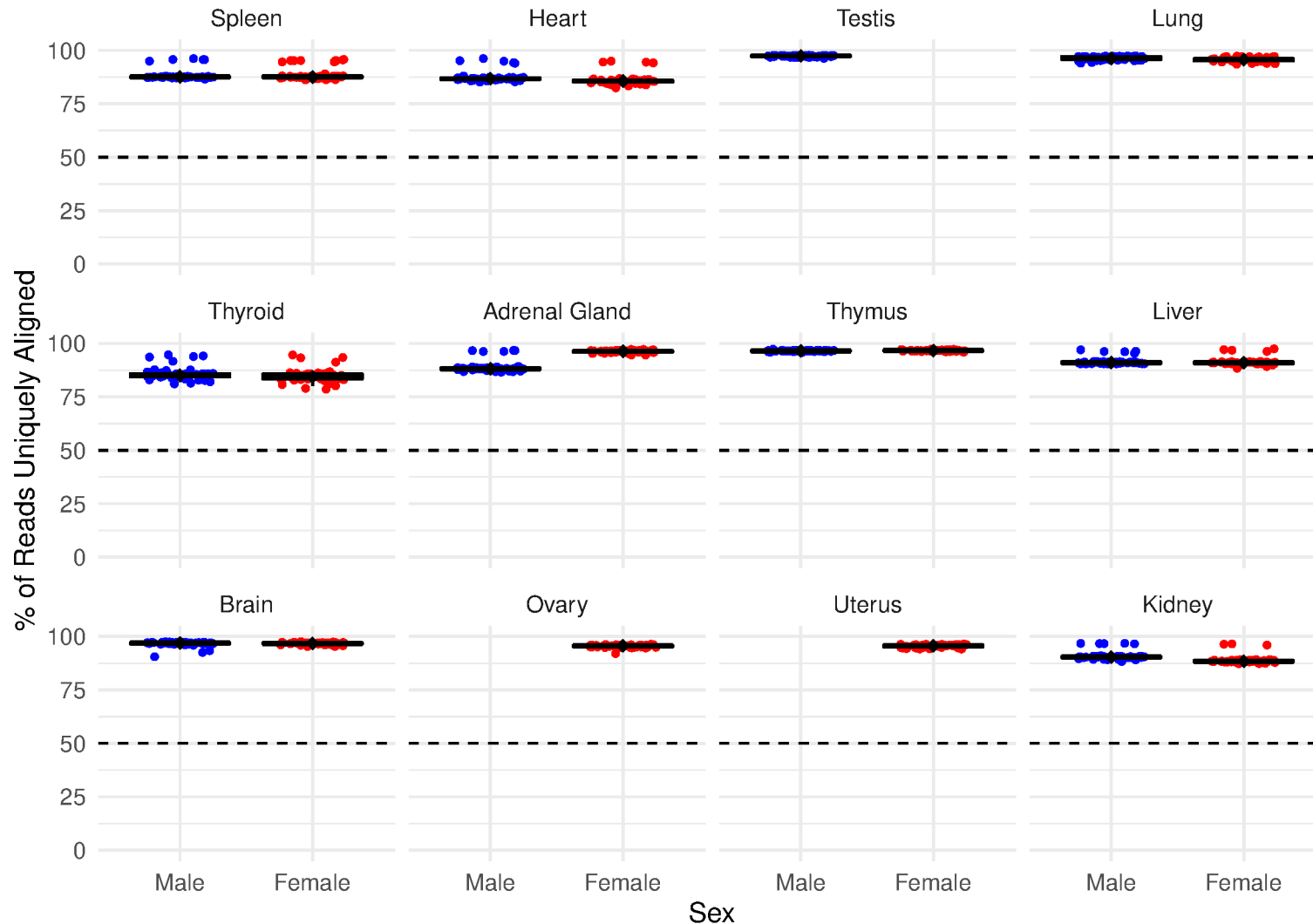
QC Summary – Sequencing Depth
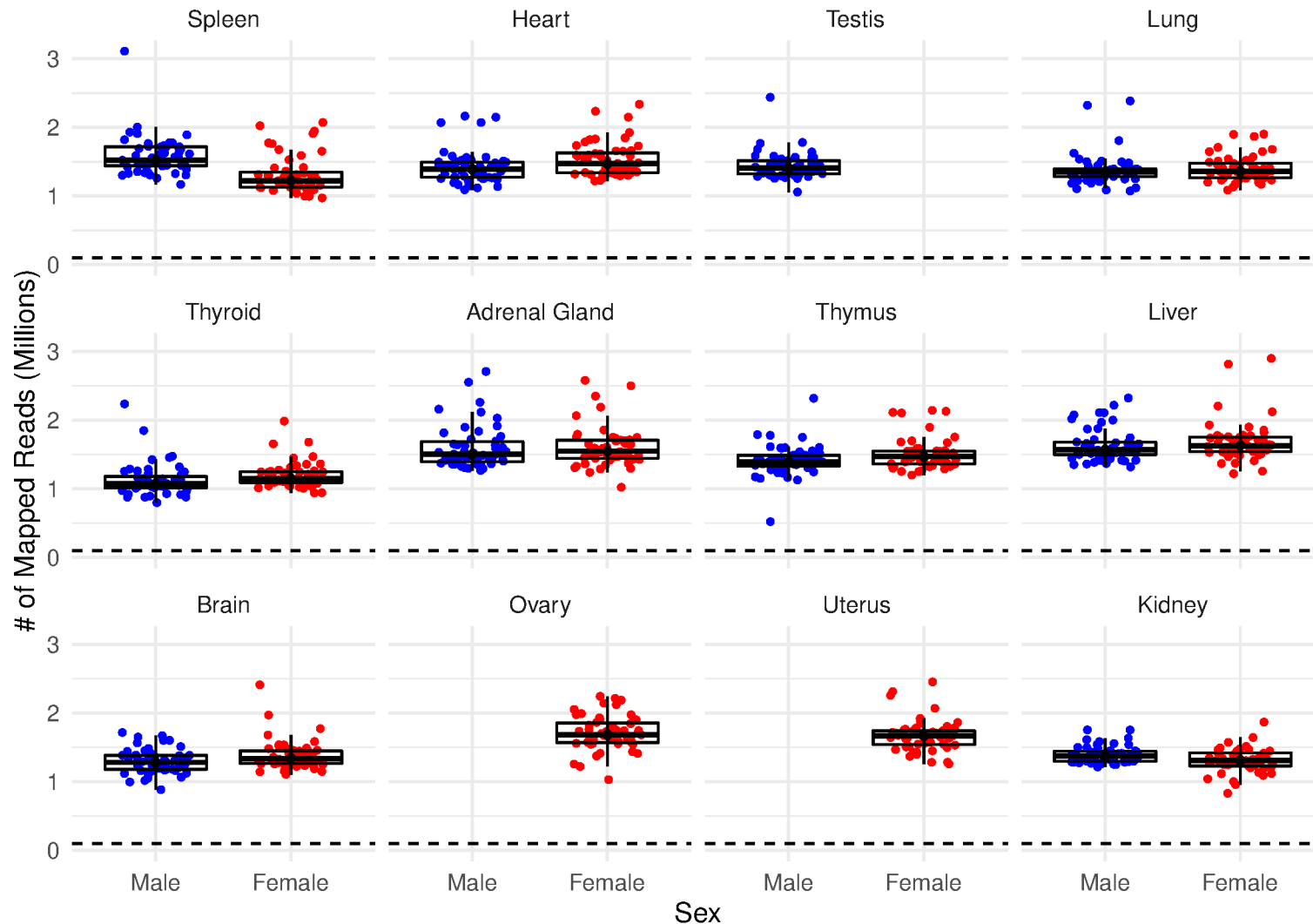
Majority of samples are above target depth

➢ Target depth = 1 million reads

➢ Remove any samples below 10% target depth *(100,000 reads)*

➢ Median depth > 1 million mapped reads for all tissues

(Example ETAP, Figure 8-2)

QC Summary – Alignment Rate

Distribution of mapping rate (% of reads uniquely aligned to probes) for each sample, grouped by tissue and sex.

➤ Removed samples < 50% *(5 samples, not shown here)*

➤ **Mapped > 80% of reads from the majority of samples**

(Example ETAP, Figure 8-3)

QC Summary – Probe Coverage
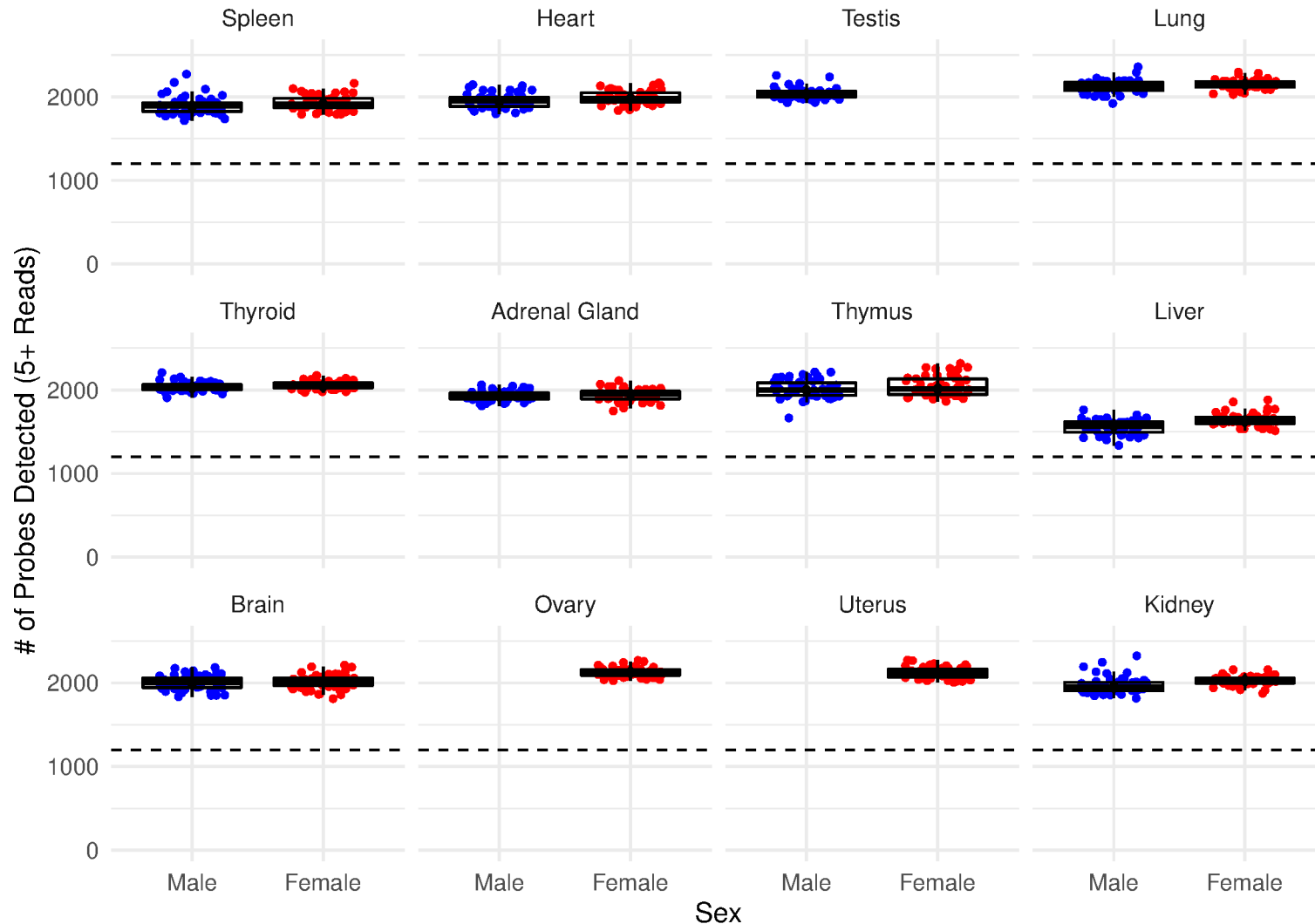
Distribution of probe coverage (number of probes detected with at least 5 reads) per sample, grouped by tissue and sex.

➢ Standard method removes any sample < 1,200 probes

➢ Distribution varies by tissue

➢ Profiling ~2,000 expressed genes for the majority of tissues

(Example ETAP, Figure 8-4)

II. Dose Response Modeling

1. Evaluate dataset for adequate signal
   - Evaluated separately for each tissue, sex
   - All tissues profiled passed ANOVA test for at least 1 probe

2. Pre-modeling filtering for dose-responsive probes ⟶
   - Williams trend test p-value < 0.05 and |Fold-Change| > 1.5

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 (GO:BP) gene sets were used
   - Gene Set BMD(L) = median of gene-level BMD(L) values
   - Overall BMD(L) = Minimum Gene Set BMD(L) across all tissues, both sexes

Tissue	Male	Female
Adrenal Gland	170	327
Brain	103	110
Heart	252	296
Kidney	540	153
Liver	347	158
Lung	340	86
Ovary	NA	168
Spleen	163	89
Testis	120	NA
Thymus	121	419
Thyroid	277	166
Uterus	NA	183

(Example ETAP, Table 5-3)

BMD Accumulation Plot
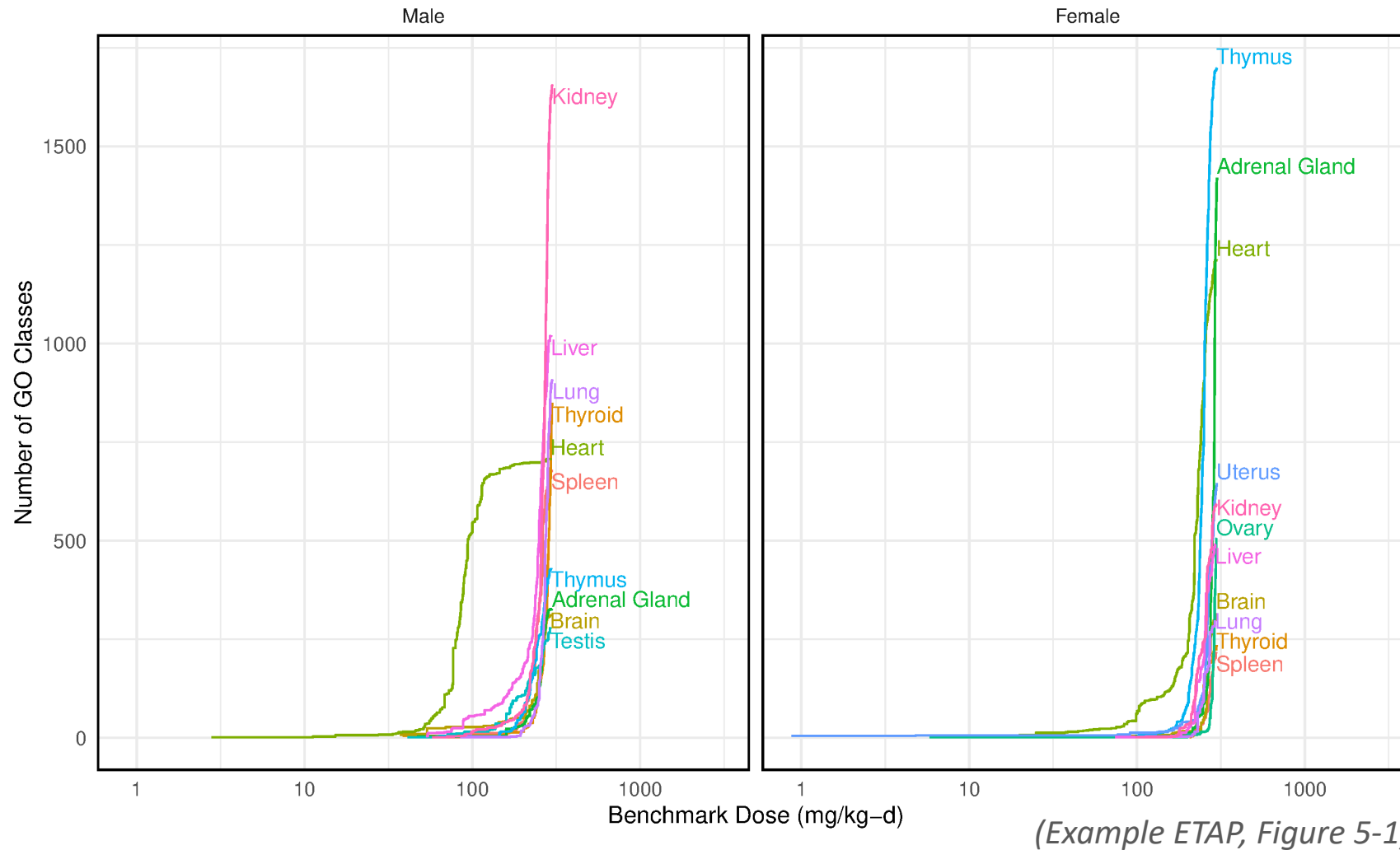
(Example ETAP, Figure 5-1)

Accumulation plots of GO biological process classes by median benchmark dose value for each tissue

POD Identification and Considerations

<u>Select GO:BP class with lowest BMD across all tissues, both sexes</u>

➢ Only consider GO:BP classes with 3 or more dose-responsive genes

➢ If there are multiple most sensitive GO:BP classes by BMD, select GO:BP class with lowest median BMDL

<u>"No Value" ETAP declared if:</u>

➢ No tissue produces any valid GO:BP classes

➢ Most sensitive GO:BP class BMD > 3-fold below lowest positive dose

(Report would include dose range tested in both cases)

Lowest BMD by Tissue (Males)

Tissue	GO Accession	Gene Ontology Biological Process Class	# of Genes with BMD	BMD (mg/kg-day)	BMDL (mg/kg-day)
Adrenal Gland	GO:0051248	negative regulation of protein metabolic process	3	4.68×10^1	1.08×10^1
Brain	GO:0033365	protein localization to organelle	3	3.70×10^1	8.35×10^0
Heart	**GO:0048608**	**reproductive structure development**	**3**	**2.80×10^0**	**8.93×10^{-1}**
Kidney	GO:1901568	fatty acid derivative metabolic process	3	5.76×10^1	3.77×10^1
Liver	GO:0006656	phosphatidylcholine biosynthetic process	3	5.30×10^1	3.53×10^1
Lung	GO:0070374	positive regulation of ERK1 and ERK2 cascade	4	8.45×10^1	1.13×10^1
Spleen	GO:0007519	skeletal muscle tissue development	4	6.30×10^1	3.87×10^1
Testis	GO:0090304	nucleic acid metabolic process	4	4.11×10^1	5.63×10^0
Thymus	GO:0010629	negative regulation of gene expression	3	6.64×10^1	4.09×10^1
Thyroid	GO:0071320	cellular response to cAMP	3	3.85×10^1	1.01×10^1

(Example ETAP, Table 5-4)

Lowest GO biological process class median benchmark dose values across tissues in male rats.
Heart was the most sensitive tissue in males with BMD = 2.8 mg/kg-day and BMDL = 0.893 mg/kg-day.

Lowest BMD by Tissue (Females)

Tissue	GO Accession	Gene Ontology Biological Process Class	# of Genes with BMD	BMD (mg/kg-day)	BMDL (mg/kg-day)
Adrenal Gland	GO:1901655	cellular response to ketone	5	9.78×10^1	5.15×10^1
Brain	GO:0097305	response to alcohol	3	8.09×10^1	4.49×10^1
Heart	GO:1901216	positive regulation of neuron death	3	1.86×10^1	3.15×10^0
Kidney	GO:0042594	response to starvation	3	9.79×10^1	7.96×10^1
Liver	GO:0034641	cellular nitrogen compound metabolic process	5	7.44×10^1	4.50×10^1
Lung	GO:0032355	response to estradiol	3	1.20×10^2	5.39×10^1
Ovary	GO:0060612	adipose tissue development	3	5.80×10^0	6.38×10^{-1}
Spleen	GO:0045597	positive regulation of cell differentiation	3	1.53×10^2	1.10×10^2
Thymus	GO:0060070	canonical Wnt signaling pathway	3	9.00×10^1	4.97×10^1
Thyroid	GO:0045597	positive regulation of cell differentiation	3	2.06×10^2	1.41×10^2
Uterus	**GO:0051271**	**negative regulation of cellular component movement**	**3**	**8.72×10^{-1}**	**1.21×10^{-1}**

(Example ETAP, Table 5-4)

Lowest GO biological process class median benchmark dose values across tissues in female rats.
Uterus was the most sensitive tissue in females with BMD = 0.872 mg/kg-day and BMDL = 0.121 mg/kg-day.
This was the most sensitive tissue overall, and was therefore used to determine the overall POD.

Summary

- A standardized transcriptomic dose response analysis and point-of-departure identification process was developed for ETAP

- In applying the ETAP process to MOPA, 1,075 of 1,092 samples (98 %) passed all quality control criteria

- All tissues passed initial criteria for dose-response modeling
  - 86 – 540 probes per tissue passed pre-filtering criteria

- Uterus was most sensitive tissue (lowest valid gene set BMD)
  - Median BMD = 0.872 mg/kg-day
  - **Median BMDL = 0.121 mg/kg-day used to compute Human Equivalent Dose**