OECD Extended Advisory Group for Molecular Screening and Toxicogenomics (EAGMST)

Guidance Document for Consistent Reporting of 'Omics Data From Various Sources

Transcriptomics Reporting Framework (TRF)

Kickoff Meeting
April 6th, 2018

Disclaimer

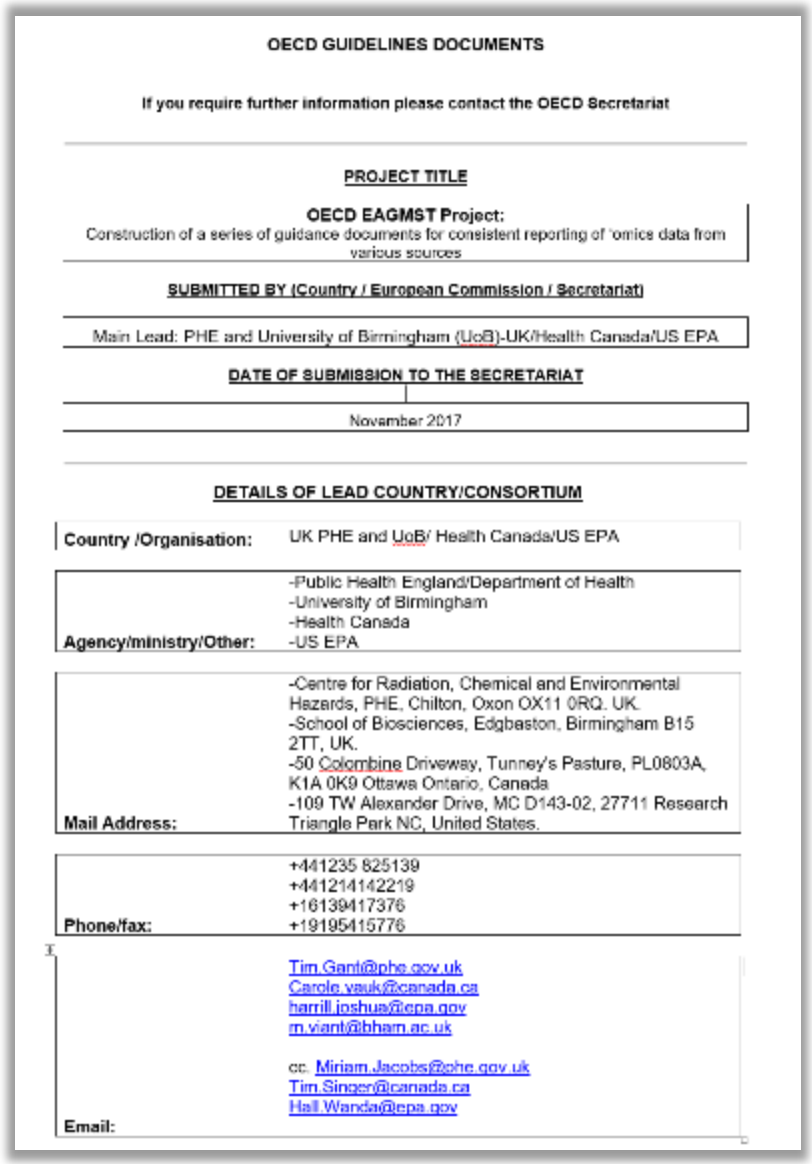
- *The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, Health Canada or Public Health England.*

Agenda

- Welcome and Introductions
- Background on the Transcriptomics Reporting Framework (TRF) Project
- Presentation of the TRF outline
- Project Timeline
- Establishment of Working Groups / Assignment of Writing Tasks
- Scheduling of Future Activities

Reasons for lack of regulatory acceptance of transcriptomic data

1) Poor experimental design and data quality (early days) = bad reputation for omics technologies
2) Lack of accepted quality control standards and data quality assessment tools
3) Lack of availability of metadata necessary for interpretation and regulatory application
4) Lack of transparency, public availability and best practices/standards for data processing methods
5) Variances in methods and prior knowledge used to analyse and interpret genomics data
6) Lack of standardized reporting frameworks to ensure that all required and appropriate data, metadata, and analytical processes are available.



OECD GUIDELINES DOCUMENTS

If you require further information please contact the OECD Secretariat

PROJECT TITLE

OECD EAGMST Project:
Construction of a series of guidance documents for consistent reporting of 'omics data from various sources

SUBMITTED BY (Country / European Commission / Secretariat)

Main Lead: PHE and University of Birmingham (UoB)-UK/Health Canada/US EPA

DATE OF SUBMISSION TO THE SECRETARIAT

November 2017

DETAILS OF LEAD COUNTRY/CONSORTIUM

Country /Organisation:	UK PHE and UoB/ Health Canada/US EPA
Agency/ministry/Other:	-Public Health England/Department of Health -University of Birmingham -Health Canada -US EPA
Mail Address:	-Centre for Radiation, Chemical and Environmental Hazards, PHE, Chilton, Oxon OX11 0RQ. UK. -School of Biosciences, Edgbaston, Birmingham B15 2TT, UK. -50 Colombine Driveway, Tunney's Pasture, PL0803A, K1A 0K9 Ottawa Ontario, Canada -109 TW Alexander Drive, MC D143-02, 27711 Research Triangle Park NC, United States.
Phone/fax:	+441235 825139 +441214142219 +16139417376 +19195415776
Email:	Tim.Gant@phe.gov.uk Carole.vauk@canada.ca harrill.joshua@epa.gov m.viant@bham.ac.uk cc. Miriam.Jacobs@phe.gov.uk Tim.Singer@canada.ca Hall.Wanda@epa.gov

ecetoc
EUROPEAN CENTRE FOR ECOTOXICOLOGY AND TOXICOLOGY OF CHEMICALS

Buesen et al. Reg Tox Pharm. 2017

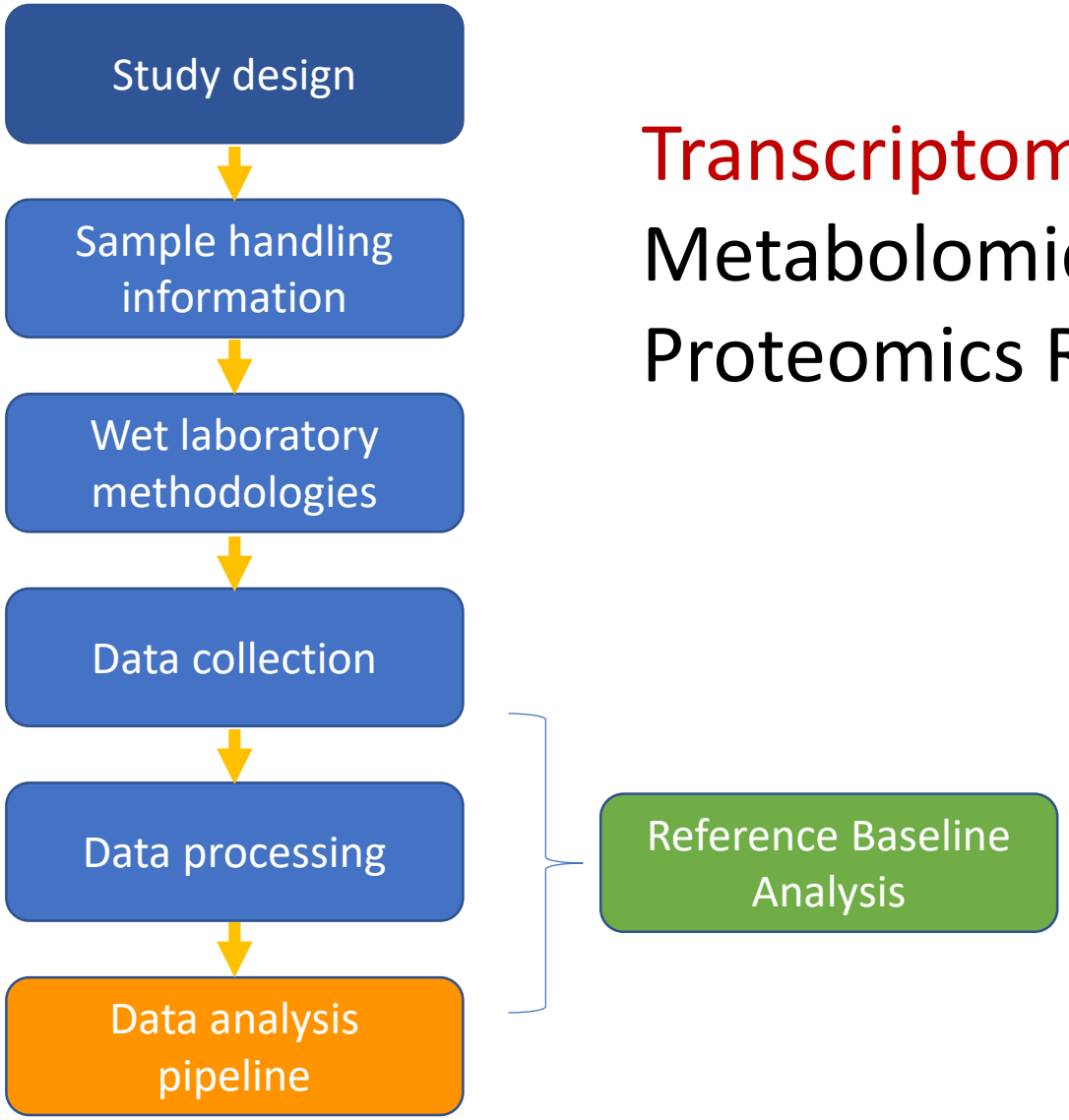
Objectives

To develop frameworks for the standardisation of reporting of ‘omics data generation and analysis, to ensure that all of the information required to understand, interpret and reproduce an ‘omics experiment and its results are available.

Purpose: to ensure that sufficient information is available to enable an evaluation of the quality of the experimental data and interpretation, and support reproducibility.

NOT to stipulate the methods of data analysis or interpretation.

Three Proposed Frameworks



Transcriptomics Reporting Framework (TRF)

Metabolomics Reporting Framework (MRF)

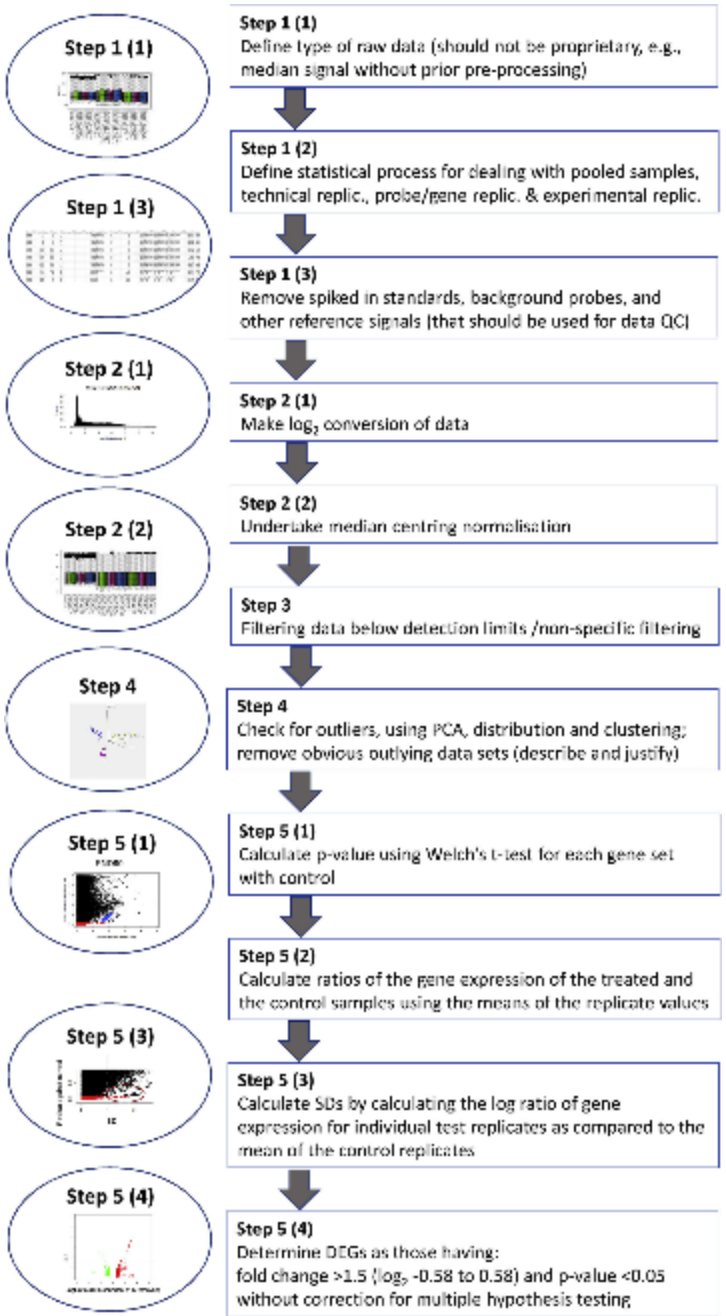
Proteomics Reporting Framework (PRF)

Not starting from scratch

ECETOC: A generic Transcriptomics Reporting Framework (TRF) for 'omics data processing and analysis. Gant et al. Reg Tox Pharm. 2017

MIAME: Brazma et al., 2001

SRA: Sequence Read Archives, www.ncbi.nlm.nih.gov/sra/docs/



Leads and roles

TRF: **Joshua Harrill (US)** and Carole Yauk (Canada)
RBA: Tim Gant (UK)
MRF: Mark Viant (UK)
OECD: Magda Sachana

Today's Objectives:

- Present draft outline of TRF
- Discuss outline and strategy
- Assign sub-group leads and members
- Roles and responsibilities
- Timelines

TRF Document, Objective & Topic Areas

OBJECTIVE: Development of a Transcriptomics Reporting Framework (TRF) for processing of 'omics data that will facilitate acceptance of transcriptomics studies in a regulatory setting.

MAJOR TOPIC AREAS:

EXPERIMENT:

- The experiment should be described in sufficient detail that would allow another researcher to replicate the experiment.
- Adapted from existing sources [CEBS, MIAME, ToxRTool, SOAR]
- Information in this section is independent of 'omics platform

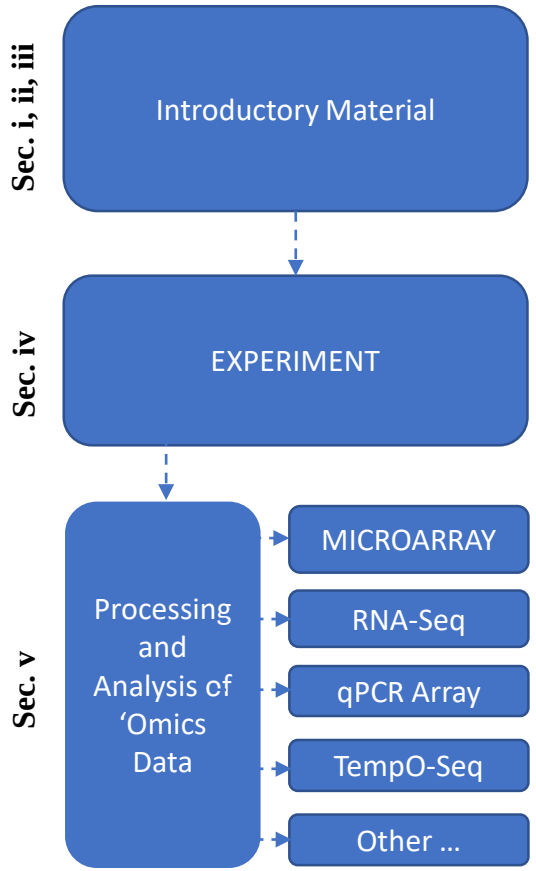
PROCESSING AND ANALYSIS OF 'OMICS DATA:

- The transcriptomics technology, sample processing procedures, methods used to collect raw data and methods used to generate processed data and identify differentially expressed genes (DEGs).
- Described in Gant et al. (2017).
- Information in this section is dependent on 'omics platform

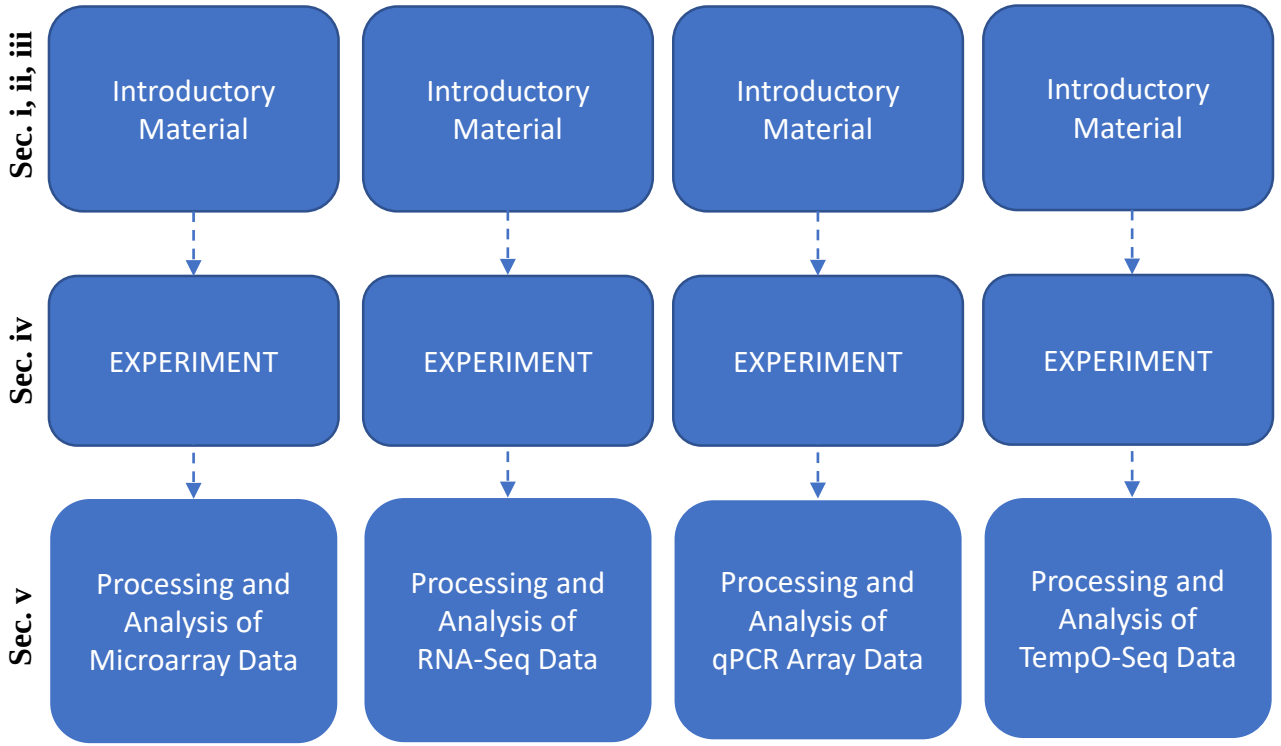
DOWNSTREAM ANALYSIS MODULES

TRF Document Structure, Options

- One large, linear document
- Contains info on all technologies



- Several short, technology-specific documents
- Redundancy in sections i-iv across documents



Downstream Analysis Modules:

- Separate documents with reporting standards for more complex analyses
- To be mixed and matched with TRF based on use case.

PCA | Hierarchical Clustering | BMD | GSEA etc.,

TRF Document Format

- **Stylistic alignment:**
  - Previous OECD guidance in the biological sciences (where applicable)
  - MRF (in progress)

- **Reporting Format (?)**
  - Narrative text / free form
  - Tabular entries

- **Database compatibility**

- **Consistent vocabulary across modules**

TRF Document Outline, Introductory Material

i. **ABSTRACT**

ii. **INTRODUCTION**
   I. Purpose / Aims
   II. Background
   III. Scope
   IV. Related ‘Omics Standard Projects

iii. **(TABLE OF) DEFINITIONS / ABBREVIATIONS**

These sections will be drafted by the leadership team and sent to the entire project group for comment.

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

Content is technology <u>independent</u>

Section to be drafted by a **section workgroup** under guidance of a **section leader**

Content leverages previously existing works

Experiment Module, Existing Resources

CEBS

- Waters, M., et al., (2003). Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics, 111*(1T), 15-28. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12735106
- Fostel, J. M., et al., (2007). Toward a checklist for exchange and interpretation of data from a toxicology study. *Toxicol Sci, 99*(1), 26-34. doi:10.1093/toxsci/kfm090

MIAME

- Brazma, A., et al., (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet, 29*(4), 365-371. doi:10.1038/ng1201-365

ToxRTool

- Schneider, K., et al., (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicol Lett, 189*(2), 138-144. doi:10.1016/j.toxlet.2009.05.013
- Segal, D., et al., (2015). Evaluation of the ToxRTool's ability to rate the reliability of toxicological data for human health hazard assessments. *Regul Toxicol Pharmacol, 72*(1), 94-101. doi:10.1016/j.yrtph.2015.03.005

SOAR

- McConnell, E. R., et al., (2014). Systematic Omics Analysis Review (SOAR) tool to support risk assessment. *PLoS One, 9*(12), e110379. doi:10.1371/journal.pone.0110379

TRF Document Outline, Experiment

iv. EXPERIMENT

- **I. Study Rationale**
- II. Study Design
- III. Subject / Test System Characteristics
- IV. Test Article
- V. Treatment Conditions
- VI. Study Exit
- VII. Sample Collection & Pre-processing
- VIII. Sample Identification Codes
- IX. Supporting Data Streams

I. Study Rationale

1. Background Information
2. Objectives

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

II. Study Design

1. Number and types of treatments
2. Individuals / samples per treatment
3. Description of block design (if applicable)
4. Descriptions of controls
5. Biologically relevant comparisons (e.g. treatment vs. control)
6. Time Table
   a. Treatments
   b. Sample Collection
   c. Etc.

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

III. Subject / Test System Charcteristics

1. Type (i.e. lab animal, alternative species, cell culture)
2. Taxonomic information (i.e. species, strain)
3. Gender
4. Age / Life Stage (i.e. in vivo age, passage number, etc.)
5. Source (i.e. commercial, derived)
6. *In vivo* housing conditions (if applicable)
7. *In vitro* culture conditions (if applicable)

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

IV. Test Article

1. Substance name
2. Formula and structure (if applicable)
3. Commercial source information (i.e. vendor, mfg #, lot #, purity, salt form)
4. Substance-specific identifiers (i.e. CAS, DTXSID, SMILES, InChi, etc.)

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

V. Treatment Conditions

1. Route of administration
2. Dose / concentration
3. Vehicle description and delivery volume
4. Exposure duration / schedule
5. Housing / culture conditions during treatment

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

VI. Study Exit

1. Method of euthanasia
2. *In vitro* termination

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

VII. Sample Collection & Pre-processing

1. Sample collection procedures
2. Pre-processing steps (if applicable)
3. Quality Control
   a. Type
   b. Applicability
   c. Performance Metric
   d. Acceptance Criteria
4. Sample storage conditions

There is a distinction between protocols used to collect and "pre-process" specimens as opposed to downstream preparation of a sample for a specific 'omics workflow.

RNA extraction vs. cDNA synthesis, fragmentation & labeling

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

VIII. Sample Identification Codes

1. Unique code for each sample
2. Relationship (per sample) with experimental metadata

TRF Document Outline, Experiment

iv. EXPERIMENT

I. Study Rationale
II. Study Design
III. Subject / Test System Characteristics
IV. Test Article
V. Treatment Conditions
VI. Study Exit
VII. Sample Collection & Pre-processing
VIII. Sample Identification Codes
IX. Supporting Data Streams

IX. Supporting Data Streams

1. *In vivo*
   a. Organ weights
   b. Histopathology
   c. Clinical Chemistry
   d. etc.
2. *In vitro*
   a. Cell viability
   b. Parallel assays
   c. etc.

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing **
III. Transcriptomics Study Design
IV. Specification of Raw Data **
V. Data Normalization **
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets **
VIII. Statistical Analysis for DEGs

Content is platform-specific

Sections to be drafted by a **section workgroup** under guidance of a **section leader**

** *Emphasis on the use and description of Quality Control procedures / samples / performance metrics.*

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. **Technology**
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

I. Technology

1. Type (e.g. microoarry, RNA-Seq, TempO-Seq, etc.)
2. Make and model (e.g. Affymetrix U133 Plus 2.0 Array)
3. Version
4. Feature type (e.g. spotted oligonucleotide, flow cell cluster)
5. Feature annotation (e.g. probe IDs)
6. Feature purpose (e.g. target gene expression, quality control, etc.)
7. Feature compostion (i.e. oligo sequence, ligated product sequence)
8. Instrument
   a. Brand, type & unique identifier
   b. Acquisition software version & config
   c. Control console operating system

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

II. Sample Processing

1. Detailed description of platform specific sample prep OR reference to standardized protocol with devitions noted.
   a. Microarray: cDNA synthesis, labeling, fragmentation, hybridization
   b. RNA-Seq: library prep
2. Sample pooling scheme (if applicable)
3. Quality Control
   a. Type
   b. Applicability
   c. Perforance metric
   d. Accept / reject criteria (sample level)

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

III. Transcriptomics Study Design

1. Sample Labeling
2. Association with metadata
3. Assignments of samples to physical unit of technology (e.g. array or flow cell serial number, flow cell lane)
   a. Association of physical unit identifiers with sample IDs.
   b. Association of molecular barcodes with sample IDs (if applicable)
4. Technical and experimental replicates
   a. Intended use
   b. Summary / aggregation strategy

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

III. Transcriptomics Study Design

1. Quality control samples, standards and spike-ins
   a. Type
   b. Source
   c. Applicability
      i. What aspect of sample processing are they designed to evaluate?
      ii. What level of the experimental design are they intended to address? (i.e. individual samples, batches)
   d. Performance metric
   e. Acceptance criteria

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

IV. Specification of Raw Data

1. Instrument Outputs
   a. Description of output type (e.g. microarray = image; RNA-Seq = fastq)
   b. File nomenclature
      i. Naming convention
      ii. File extension
      iii. Association with raw data
      iv. Association with experimental metadata
   c. Retention as part of experimental record ? (YES / NO)
      i. Archiving location

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

IV. Specification of Raw Data
2. Raw Data
   a. Description of raw data type
   b. File nomenclature
      i. Naming convention
      ii. File extension
      iii. Association with instrument output file

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

IV. Specification of Raw Data

c. Protocol for deriving raw data from instrument output
   i. Feature extraction tool / config
   ii. Background subtraction (if applicable)
   iii. Spot / probe level QC metrics
   iv. Sequence alignments (if applicable)
      1. Alignment tool & config
      2. Interpretatio of base-level quality scores
      3. Reference genome identity, version and accessibility path
d. Handling of replicates probes for calculation of gene level summaries.

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

V. Data Normalization

1. Detailed description of data normalization and transformation procedures OR reference to a standardized protocol with deviations notes.
2. Code for normalization is using an open-source package
3. Criteria for selection / exclusion of probes used in normalization procedures and list of selected and/or excluded probes
4. Quality assessment of normalization
   a. Type
   b. Applicability
   c. Performance metric
   d. Acceptance criteria

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Datsets
VIII. Statistical Analysis for DEGs

V. Data Filtering

1. Criteria for filtering probe / gene level
   a. Definition of a low abundance or absent probe / gene
      i. Threshold
      ii. How threshold was determined
   b. Process used to remove poor quality probes
   c. Use-case specific filtering of relevant genes (if applicable)

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Data
VIII. Statistical Analysis for DEGs

VII. Identification and Removal of Outlying Data

1. Description of how outlier samples were identified
   a. Detailed procedure
   b. Threshold
2. Processing step where exclusion occurs
3. List of samples excluded and per sample justification for exclusion.

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Data
VIII. Statistical Analysis for DEGs

VII. Statistical Analysis for DEG determination

1. Software use to identify DEGs
   a. Name
   b. Version
   c. Configuration
2. A list of biological comparisons for which DEGs were determined and association with sample level identifiers
3. Description of groups sizes, experimental replicates and treatment of those in the statistical analysis

TRF Document Outline, ‘OMICS DATA

iv. PROCESSING AND ANALYSIS OF ‘OMICS DATA

I. Technology
II. Sample Processing
III. Transcriptomics Study Design
IV. Specification of Raw Data
V. Data Normalization
VI. Data filtering
VII. Identification and Removal of Low Quality or Outlying Data
VIII. Statistical Analysis for DEGs

VII. Statistical Analysis for DEG determination

4. Description of statistical test used to identify DEGs
   a. Significance filters
   b. Fold change filters
   c. Multiple testing corrections
   d. Combination of these
5. Description of Output Tables
   a. File nomenclature and extension
   b. Description of output categories
6. Description of how DEGs are reported
7. Code for statistical analysis if using an open-source software package.

Workgroups

Each workgroup will consist of the following:

Title	Identity	Roles
Section Lead	TBD	Coordinate workgroup activities Maintain draft of section
Workgroup Members (n = 2-3)	TBD	Contribute text and content for sections
“Floating” Facilitator	Joshua Harrill & Carole Yauk	Ensure consistency and cross-talk with other workgroups. Monitor progress in accordance with project timeline Foster discussion.
OECD Secretariat	Magda Sachana	Project administration / OECD liason

Need to identify Section Leads and Working Group Members for:

- Section iv. EXPERIMENT
- Section v. PROCESSING AND ANALYSIS ‘OMICS DATA (**For each technology**)

All members of the project group will have the opportunity to comment on sections drafted by each workgroup.

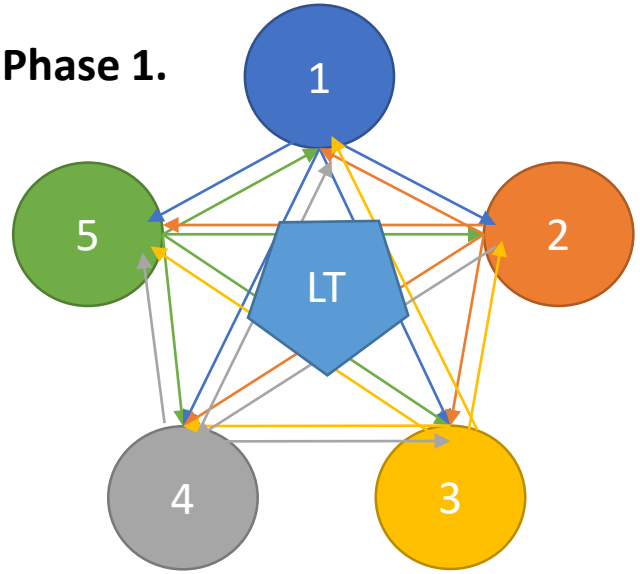
Project group leads (Harrill & Yauk) will integrate sections into the final document.

Round Robin Case Study

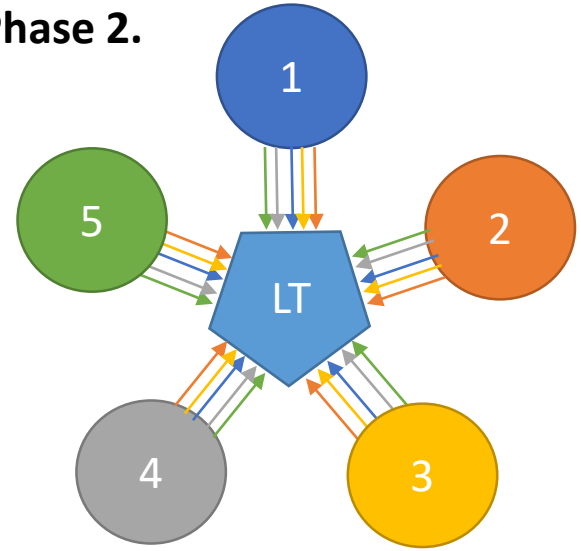
Objectives: Evaluate the utility of the TRF in fostering reproducibility of 'omics data analysis by different research groups.

Phase 1	**Step 1.**	Identify three (or more) analysis teams from various organizations.
	Step 2.	Coordinate with the leadership team to identify an existing dataset from each team
	Step 3.	Ask each team to: 1) Analyze their data & determine DEGs (no other instructions or restrictions). 2) Report DEGs and 3) Fill out the TRF describing what they did
Phase 2	**Step 4.**	Provide raw data and completed TRFs (blinded, sans DEG list) to other analysis teams
	Step 5.	Ask teams to: 1) Try and reproduce the analysis described in the TRF 2) Report DEGs to leadership team 3) Identify areas in the completed TRFs which were unclear
	Step 6.	Leadership team assesses concordance of DEG call results and report results back to analyses teams.
	Step 7.	Refine TRF (if necessary)

Phase 1.



Phase 2.



Project Timeline

Date	Milestone
April, 2018	Kickoff teleconference / recruiting for working groups
May – June, 2018	Begin work on Introduction, Experiment and Microarray modules
June 26-27th, 2018	OECD WPHA Meeting – Project update (presentation) and potential F2F for attendees
Dec, 2018	Advanced drafts of Introduction, Experiment and Microarray sections due OECD Winter Meeting
June, 2019	Finalization of Introduction, Experiment and Microarray sections Advanced drafts of RNA-Seq, PCR array, TempO-Seq due OECD Spring Meeting
Dec, 2019	Final document(s) – project completion OECD Winter Meeting

Leadership question: Where does Round Robin fit into this project timeline. Too ambitious?

Questions (???)