

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

**Update on Development of Guidance Document(s) for Consistent  
Reporting of 'Omics Data From Various Sources**

**Transcriptomics Reporting Framework (TRF)**

*EAGMST Meeting, Boulogne FR*

*June 21<sup>st</sup>, 2019*



# 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.*

# Background

- Transcriptomic methodologies are increasingly applied in research to support regulatory toxicology by providing a greater understanding of modes and mechanisms of toxicity, inform read-across and identify point-of-departure(s) for chemical bioactivity.
- To facilitate use in regulatory decision-making in the absence of a test guideline, a comprehensive reporting framework is necessary to thoroughly document the components of a transcriptomic study.
- Thus, in 2018, the OECD EAGMST undertook the development of a Transcriptomic Reporting Framework (TRF) to foster regulatory update of transcriptomic data.
- Previous frameworks for documenting transcriptomic studies focus primarily upon annotation of data (raw and normalized), samples, sample to data relationships and technology-specific feature annotation. The TRF includes all of these elements, but also provides a means to document the computational steps used to analyze the transcriptomic data and generate downstream results that may be of use in a regulatory decision-making context.
- Reporting using the TRF will maximize the probability that the results of a transcriptomic experiment can be reproduced, in parallel with providing the essential information needed to evaluate transcriptomic study design quality, interpretation and applicability to regulatory decision-making processes.

# Project Description

**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....**Rather**, provide guidance on reporting of information that fosters transparency and reproducibility.

Project Name	Project Lead
Metabolomics Reporting Framework (MRF)	Mark Viant (U. Birmingham, UK)
<b>Transcriptomics Reporting Framework (TRF)</b>	Joshua Harrill (USEPA) Carole Yauk (Health Canada)
Reference Baseline Analysis (RBA)	Tim Gant (PHE, UK)

# TRF Objective, Working Group Charge & Scope

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

**WORKING GROUP CHARGE:** The TRF working group is tasked with determining what information should be captured by the TRF to support interpretation and computational reproducibility of 'omics experiments by members of the regulatory community. Such information will also be of value to researchers in academia and industry.

**SCOPE:** The TRF was designed as a tool for documenting the details of laboratory-based toxicology studies that apply a transcriptomic technology: i.e. an assay that measures the abundance of many transcripts simultaneously and thus provides highly multiplexed outputs. The TRF is appropriate for use in documenting experiments involving the use of either in vivo or in vitro laboratory models. It is intended to facilitate the comprehensive and transparent documentation of a transcriptomic study including the experimental design, sample processing procedures, data collection, data normalization and downstream computational analyses, the results of which could be used in regulatory decision-making contexts. The information captured by the TRF should be of sufficient detail for end users to replicate all aspects of the transcriptomic experiment in each of these areas

# TRF Document, 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
- Information in this section is independent of 'omics platform
- Includes relevant fields from relevant OECD harmonized reporting templates (OHT 201)

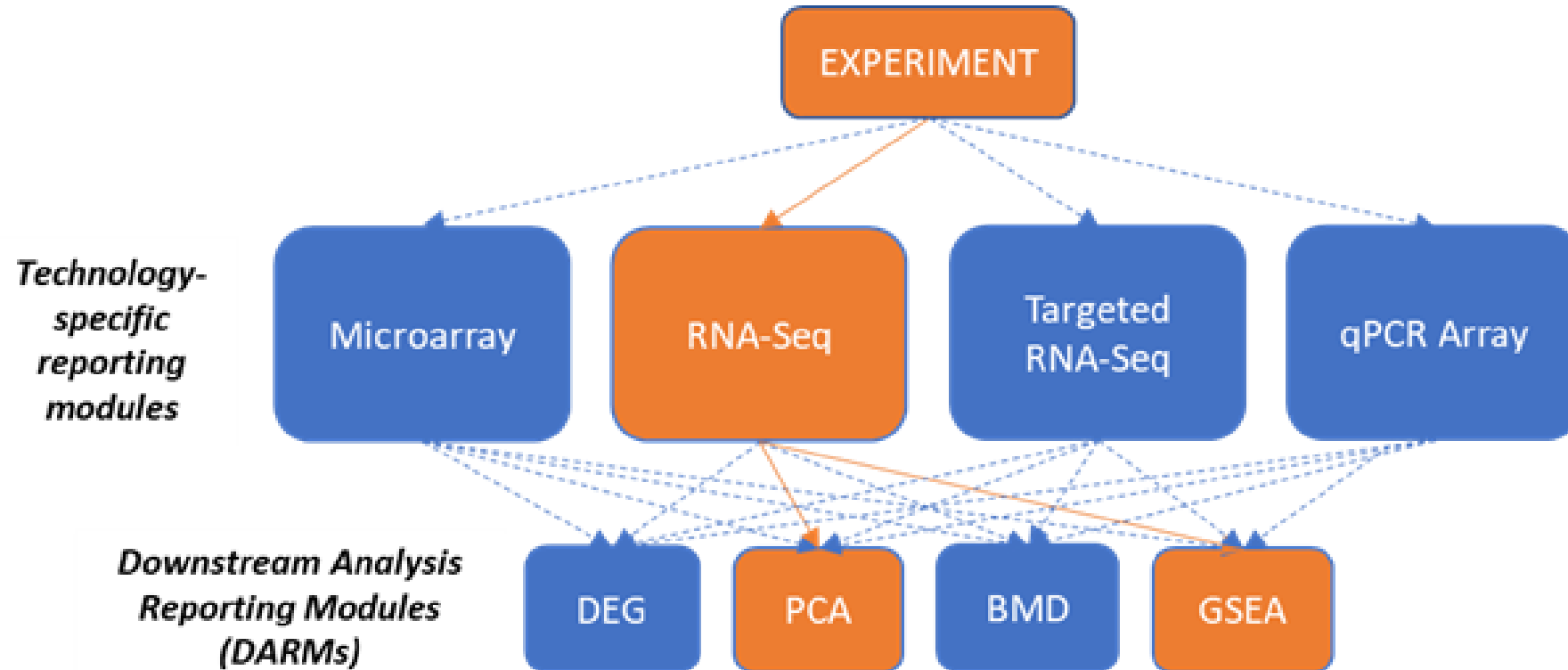
## **PROCESSING AND ANALYSIS OF 'OMICS DATA:**

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

## **DOWNSTREAM ANALYSIS REPORTING MODULES [DARMS]**

- Detail the steps and resources necessary to reproduce a computational analysis of the processed data.

# Modular Structure of Transcriptomics Reporting Framework



- To input information into the TRF, a researcher selects reporting modules relevant to the technology platform and computational analyses used to conduct a study.
- Report the information that would be required by an end-user to fully comprehend and replicate the analyses.

# Section Workgroups

Each workgroup will consist of the following:

Title	Identity	Roles	
Section Leads	<b>Experiment</b> <b>Microarray</b> RNA-Seq q-PCR array Targeted RNA-Seq <b>DARM.1 [DEG]</b>	<b>Raffaella Corvi [ JRC ]</b> <b>Vikrant Vijay [ NCTR ]</b> Florian Caiment [ Maastricht ] Jason O'Brien [ ECCC ] Scott Auerbach [ NTP ] <b>Lyle Burgoon [ ERDC ]</b>	Coordinate workgroup activities Maintain draft of section Manage timelines for deliverables
Workgroup Members (n = 2 - ?)	Various gov't, academic and industry scientists	Contribute text and content for sections	
"Floating" Facilitators	Joshua Harrill [ USEPA ] Carole Yauk [ Health Canada ]	Ensure consistency and cross-talk with other workgroups. Monitor progress in accordance with project timeline Foster discussion.	
OECD Secretariat	Magda Sachana	Project administration / OECD liaison	

All members of the TRF workgroup will have the opportunity to comment on each section.

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



# Format of the TRF

## 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

## PROCESSING 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 Datasets

## DARM.1 (DEGs)

- I. Statistical Software
- II. Contrasts for DEG Identification
- III. Assay Experimental Design
- IV. Statistical Analysis
- V. Outputs

## REPORT:

- 3.1. Type and version of the platform, manufacturer's name (e.g., Affymetrix U133 Plus 2.0 Array)
- 3.2. The unique identifier (e.g. serial number)
- 3.3. Feature type (e.g. spotted oligonucleotide)
- 3.4. Annotation (e.g. probe IDs)
- 3.5. Purpose (e.g. target gene expression, quality control, etc.)
- 3.6. Composition (i.e. oligo sequence, ligated product sequence)
- 3.7. Control console operating system
- 3.8. Any other relevant information

- The TRF provides narrative descriptions and basic background info for each reporting field (consistent with MERIT and MRF)

# Experiment Section Excerpt

## 2.3. Test System Characteristics

14. Transcriptomic studies suitable for use in a regulatory context may be performed in a wide range of model species and/or *in vitro* test systems. During the introduction of transcriptomic technologies into the field of toxicology, researchers were limited in terms of species selection by the small variety of species-specific platforms (mostly microarrays) available from commercial vendors. The traditional use of rodent species (i.e. mouse and rat) in toxicity testing and the availability of annotated genomes for these species partially drove commercial vendors to develop transcriptomic platforms specific for rodents. The proliferation of annotated genomes and the introduction of RNA-Seq, targeted RNA-Seq and PCR arrays into the field of toxicology greatly expanded the number of species from which transcriptomic data could be generated for use in evaluating chemical toxicity and also roughly coincided with increased emphasis on the use of alternative species (i.e. fish, invertebrates, etc.) and *in vitro* models (many of which are human-derived) for toxicity testing.

15. Similar to traditional toxicity testing, it is critical that regulatory scientists applying transcriptomic data for risk assessment be provided with comprehensive information regarding the characteristics of the test system from which the data are derived. Test system refers to the biological system that is exposed to the test items to obtain experimental data. There are numerous examples in the toxicological literature demonstrating differential susceptibility of different species, strains within a species and gender to chemical toxicity. Likewise, *in vitro* test systems derived from different species, tissues or even individuals vary in terms of relative sensitivity to toxicant exposure. The evaluator must be equipped with detailed and accurate information regarding the test species or *in vitro* test system used to generate the transcriptomics data in order to critically evaluate the results and accurately compare the results across studies and data types.

16. With respect to *in vivo* toxicology studies, researchers should include relevant taxonomic information (i.e. species and strain), sex, age (at onset of dosing and at study termination) and commercial source of all individuals included in a study. If determination of sex was not included in the study design (such as in the case of some types of alternative species studies), or pooled samples from multiple individuals were examined, then this should be explicitly described by the researcher. Researchers should also include detailed information on the housing conditions for all individuals included in a study including number of individuals housed per cage, type of bedding, type of food, type of water provided, food and water accessibility (i.e. ad libitum or defined quantities), light / dark cycle, relative humidity and other housing conditions the researcher may deem relevant for study interpretation. In general, information following the ARRIVE guidelines should be included ([Kilkenny, Browne et al. 2010](#)).

17. With respect to *in vitro* toxicology studies, researchers should include relevant information on culture type including species, strain (if applicable) and sex of the organism from which the cells were derived. Researchers should include detailed information on culture conditions used to conduct the study as applicable, including complete media formulations, culturing vessel, growth substrate, passage number, donor lot, source (including commercial vendor or academic source), incubator conditions and proof of cell line authentication if available. In studies using complex, multicellular culture models (e.g. 3D cell models, organoids, organ-on-chip, etc.), the researchers should report what types of cells the cultures are expected to contain, cite relevant literature characterizing the model system and describe any other relevant characteristic that might not be listed here.

## REPORT:

### 2.18. General characteristics of the test system or subject:

#### *In vivo*

- a. Animal species
- b. Strain
- c. Sex
- d. Age during study
- e. Developmental stage
- f. Individual weights/lengths at start
- g. Commercial source
- h. Any interventions that were carried out before or during the experiment
- i. Quality criteria before use
- j. Other ...

#### *In vitro*

- a. Cell type (cell line or primary cells, tumorous cells, etc...)
- b. Origin (animal/organ)
- c. Cell passage number
- d. Differentiation stage
- e. Absence of mycoplasma
- f. Metabolic competence
- g. Cell banking
- h. Supplier
- i. Quality criteria before use
- j. Other...

# Microarray Section Excerpt

## 3.1. Technology

33. This section describes the information a regulatory scientist requires for determining analytical sensitivity, limits of detection, interference, and precision (reproducibility and repeatability) of the microarray technology used in a transcriptomic experiment.

34. Documentation of the identity and number of probes measured and detection calls for individual probes is essential for interpretation of a microarray-based transcriptomics study. Microarray probes vary significantly in their hybridization properties, and arrays are limited to interrogating only those genes for which probes are designed. In addition, a potential limitation of microarray technology is background hybridization that limits the accuracy of expression measurements, mainly for transcripts present in low abundance.

35. A “platform” defines the microarray template and requires documentation of the sequence identity tracking information for each feature on the microarray. Probes (or oligoprobes) are short DNA sequences targeting a region of a transcript used to detect the presence of nucleotide sequences through hybridization of complementary labeled target to single-stranded nucleic acid. Whichever platform is used, the underlying mapping of the probes to biological entities (i.e. transcripts/genes/proteins) must be annotated. While probe sequences don’t change, genome assemblies (e.g. chromosomal sequences) and annotation of biological entities are both subject to change over time. Given the iterative improvement of genome annotations, a certain microarray probe that mapped to gene X in one instance could be mapped to gene Y in another instance because gene X has been made obsolete by a genome annotation update, or its exon-intron structure has changed in light of new supporting evidence. Therefore, the accurate reporting of version information, both in terms of microarray platform and the reference genome used for biological interpretation, is essential for understanding the results of a microarray-based transcriptomics study.

36. Manufacturers are vital for supplying testing laboratories with reliable products and probe sequences / annotation. Ideally, users and manufacturers communicate so that substantial changes to the product are conveyed to users. Finally, the hardware and software packages that generate and process microarray data represent a wide assortment of data styles and formats. Therefore, information on hardware and software versions and configurations used to collect microarray data should be reported to facilitate assessment of data provenance and quality.

### REPORT:

- 3.1. Type and version of the platform, manufacturer’s name (e.g., Affymetrix U133 Plus 2.0 Array)
- 3.2. The unique identifier (e.g. serial number)
- 3.3. Feature type (e.g. spotted oligonucleotide)
- 3.4. Annotation (e.g. probe IDs)
- 3.5. Purpose (e.g. target gene expression, quality control, etc.)
- 3.6. Composition (i.e. oligo sequence, ligated product sequence)
- 3.7. Control console operating system
- 3.8. Any other relevant information

# Challenges for TRF Development

Challenge Area	Comment
<b>Comprehensiveness</b>	Does the TRF include all of the information that would be required by an end-user (i.e. regulator) to judge the quality of an 'omics study and appropriateness for use in regulatory decision making?
<b>Level of Detail</b>	What level of detail is necessary for a regulator to have confidence that an 'omics study is suitable for using in a regulatory decision making process? Is too much detail a hinderance for the end user?
<b>Mandatory / Optional Reporting Fields</b>	'Omics studies are inherently associated with large amounts of experimental metadata. Is all of the information necessary for a regulator to judge whether a study is suitable for use in regulatory application?
<b>Reporting vs. Best Practice</b>	The TRF is a tool for data capture. Should be informative with regards to information that must be captured, but not prescriptive in terms of 'best practices' for study design or analysis
<b>Ease of Use</b>	The narrative format of the TRF is informative, but lengthy, and may be a barrier for use by researchers. <b>Is there a way to improve ease of use?</b>
<b>Interplay with Existing Reporting Structures</b>	Many elements of the TRF (i.e. EXPERIMENT section) overlap with certain OHT templates (i.e. OHT 201). However, not all data submitters and end users in the regulatory community use OHT / IUCLID compatible reporting templates. What is the best approach for maintaining the TRF as a stand alone document with broad applicability and also facilitating efficient use by the OHT user community? What is the interplay with other 'omics reporting structures (i.e. GEO, ArrayExpress, etc.)
<b>Downstream Analysis Reporting Modules (DARMs)</b>	DARMs was a concept that emerged during development of the TRF workplan. Critical for informing end users how 'omics data was analyzed to characterize the biological activity of chemicals. What DARMs would be of greatest value for the regulatory community? <i>Myriad analysis approaches for 'omics data → DARM development could go on indefinitely → "When is enough, enough?"</i>

# TRF Tabular Reporting Companion Document

- To ease reporting efforts, a tabular companion document to the TRF (i.e. spreadsheet) is under development.
- Data providers report using the tabular structure and refer to the narrative TRF for guidance on reporting fields.
- Will indicate **MANDATORY** and **OPTIONAL** reporting fields as determined by the TRF working groups.

71	<b>3. Test system characteristics</b>	
72	<b>3.1 General characteristics of the test system or subject</b>	<b>REPORT</b>
73	<i>In vivo</i>	
74	a. Animal species	
75	b. Strain	
76	c. Sex	
77	d. Age during study	
78	e. Developmental stage	
79	f. Individual weights/lengths at start	
80	g. Commercial source	
81	h. Any interventions that were carried out before or during the experiment	
82	i. Quality criteria before use	
83	j. Other ...	
84		
85	<i>In vitro</i>	
86	a. Cell type (cell line or primary cells, tumorous cells, etc...)	
87	b. Animal of origin	
88	c. Organ of origin	
89	d. Cell passage number	
90	e. Differentiation stage	
91	f. Absence of mycoplasma	
92	g. Metabolic competence	
93	h. Cell banking	
94	i. Supplier	
95	j. Quality criteria before use	
96	k. Other...	
97		

Navigation: Experiment | Processing microarray data | DARM DEG | (+)

1	<b>Reporting on a microarray experiment: data generation, handling, normalization and filtering.</b>	
2		
3	<b>1. Technology</b>	<b>REPORT</b>
4	Type and version of the platform, manufacturer's name (e.g. Affymetrix U133 Plus 2.0 Array)	
5	The unique identifier (e.g. serial number)	
6	Feature type (e.g. spotted oligonucleotide )	
7	Annotation (e.g. probe IDs)	
8	Purpose (e.g. target gene expression, quality control, etc.)	
9	Composition (e.g. oligo sequence, ligated product sequence)	
10	Control console operating system	
11		
12	<b>2. Transcriptomic study design</b>	
13	Sample pooling protocol (if applicable)	
14		
15	<b>2.1 Controls and limits on their acceptability</b>	
16	Quality control approach/sample type (text description, with ID of controls)	
17	Applicability (use of the controls)	
18	Performance metric (pass/fail criteria)	
19	Summary measures of the negative control spots (mean and standard deviation)	
20	Reproducibility of replicate probes (the median CV of replicate probes)	
21	Evaluation metrics for spike-in controls	
22		
23	<b>2.2 Linear amplification, labelling, cDNA/cRNA preparation</b>	
	Labeling protocol including information pertaining to: (a) when label was added (i.e. was the label added during reverse transcription, or following	

Navigation: Experiment | Processing microarray data | DARM DEG | (+)

# Round Robin Case Study

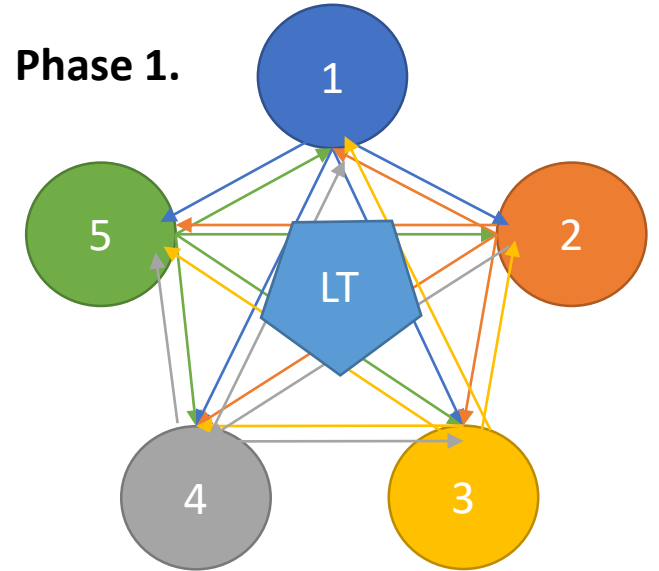
**Objectives:** Evaluate the utility of the TRF in fostering reproducibility of ‘omics data analysis by different research groups.

Phase 1

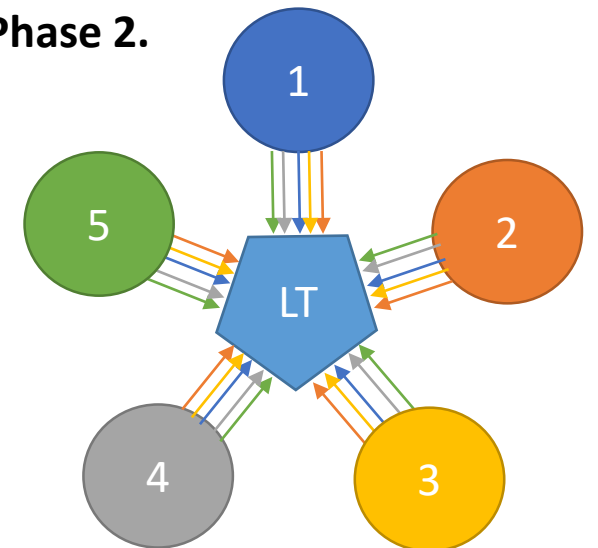
Phase 2

<b>Step 1.</b>	Identify multiple analysis teams across various organizations.
<b>Step 2.</b>	Coordinate with the leadership team to identify an existing microarray dataset from each team
<b>Step 3.</b>	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
<b>Step 4.</b>	Provide raw data and completed TRFs (blinded, sans DEG list) to the other analysis teams
<b>Step 5.</b>	Ask teams to: 1) Try and reproduce the analysis described in each TRF 2) Report DEGs to leadership team 3) Identify areas in the completed TRFs which were unclear and may have lead to inconsistencies.
<b>Step 6.</b>	Leadership team assesses concordance of DEG call results and report results back to analyses teams.
<b>Step 7.</b>	Refine TRF (if necessary)

Phase 1.



Phase 2.



# Project Timeline (Revised)

Date	Milestone
<b>April, 2018</b>	Kickoff teleconference / recruiting for workgroups
<b>May – June, 2018</b>	Begin work on Introduction, Experiment, Microarray and DARM.1 modules
<b>June, 2018</b>	OECD WPHA & EAGMST Meeting – Project update (presentation)
<b>Dec, 2018</b>	First drafts of Introduction, Experiment and Microarray sections due OECD Winter Meeting
<b>June, 2019</b>	Near Final Draft of Introduction, Experiment and Microarray sections Submit TRF drafts to EAGMST TRF Working Group members for review OECD Summer Meeting
<b>July-Aug, 2019</b>	Initiate RNA-Seq, targeted RNA-Seq and PCR array working groups Recruit for working members for additional DARM modules (BMD, PCA, etc.) Kickoff of Round Robin Case Study for Microarray
<b>Dec, 2019</b>	Near final drafts of RNA-Seq, targeted RNA-Seq and PCR array documents due OECD Winter Meeting

Project will likely require extension to the summer of 2020.

# Acknowledgements

## OECD TRF Contributors

<b>EXPERIMENT</b>	<b>MICROARRAY</b>	<b>DARM.1</b>
<b>Rafaella Corvi</b>	<b>Vikrant Vijay</b>	<b>Lyle Burgoon</b>
Gina Hilton	Tim Gant	Andrew Williams
Ian Cotgreave	Laura Gribaldo	Natalia Reyero
Ed Perkins	Andrew Williams	Andy White
Mirjam Luitjen		
John Colbourne		
Tao Chen		
Monica Vaccari		
Sofia Batista-Leite		
Ivana Campia		

## OHT Mapping

Alberta Martin-Aparicio  
Raffaella Corvi  
Sofia Batista-Leite  
Ivana Campia

## TRF Leadership Team

Carole Yauk  
Tim Gant  
Magda Sachana

*Questions?*