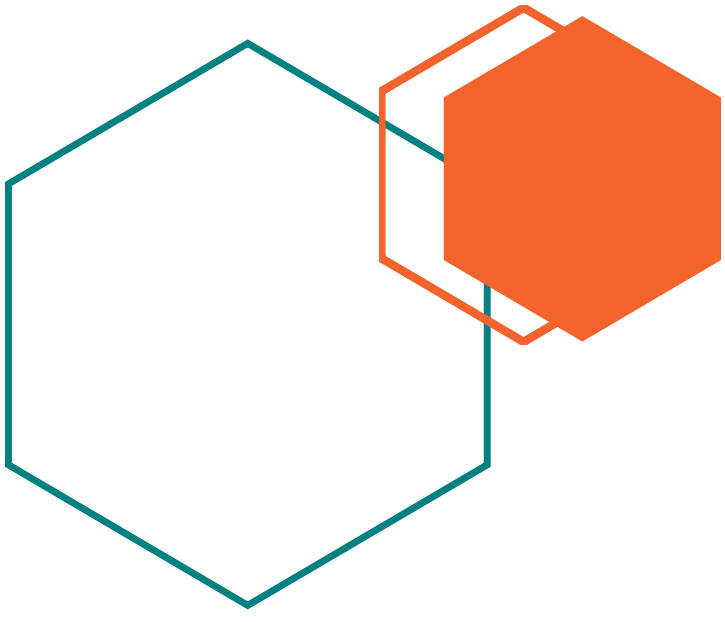
Technology Advancing Rapid Gene Expression-based Testing (TARGET)

A US Federal Challenge Competition, Sponsored by the U.S. Environmental Protection Agency

U.S. EPA and partners are seeking high quality, low cost, technologies for measuring global gene expression in samples from non-human organisms including fish, invertebrates, and plants/algae. The technology will serve as a foundation for the next generation of high throughput ecological toxicity tests for use in environmental safety evaluation of chemicals.



$300,000 will be awarded for the submission demonstrating the best technical performance at the lowest cost.

Problem

Chemical Safety Landscape

Tens of thousands of chemicals are currently in use and hundreds more are introduced to the market every year. Only a small fraction has been thoroughly evaluated for potential risks to human health and the environment.

High Throughput Screening

To aid hazard evaluation, U.S. EPA and partners have generated high throughput screening data for thousands of chemicals via the ToxCast and Tox21 programs.

Data Gap

However, to date, these programs have focused on human health. Coverage of biological pathways and physiological functions that are not conserved with mammals (for example photosynthesis in plants; molting in invertebrates) is lacking. This leaves critical gaps in the current high throughput screening battery relative to identifying ecological hazards and protecting the environment.

Proposed Solution

Coupling global gene expression profiling with high throughput assays that employ representatives of plant and animal diversity can address these gaps, allowing us to better protect the environment from harmful effects of chemicals.

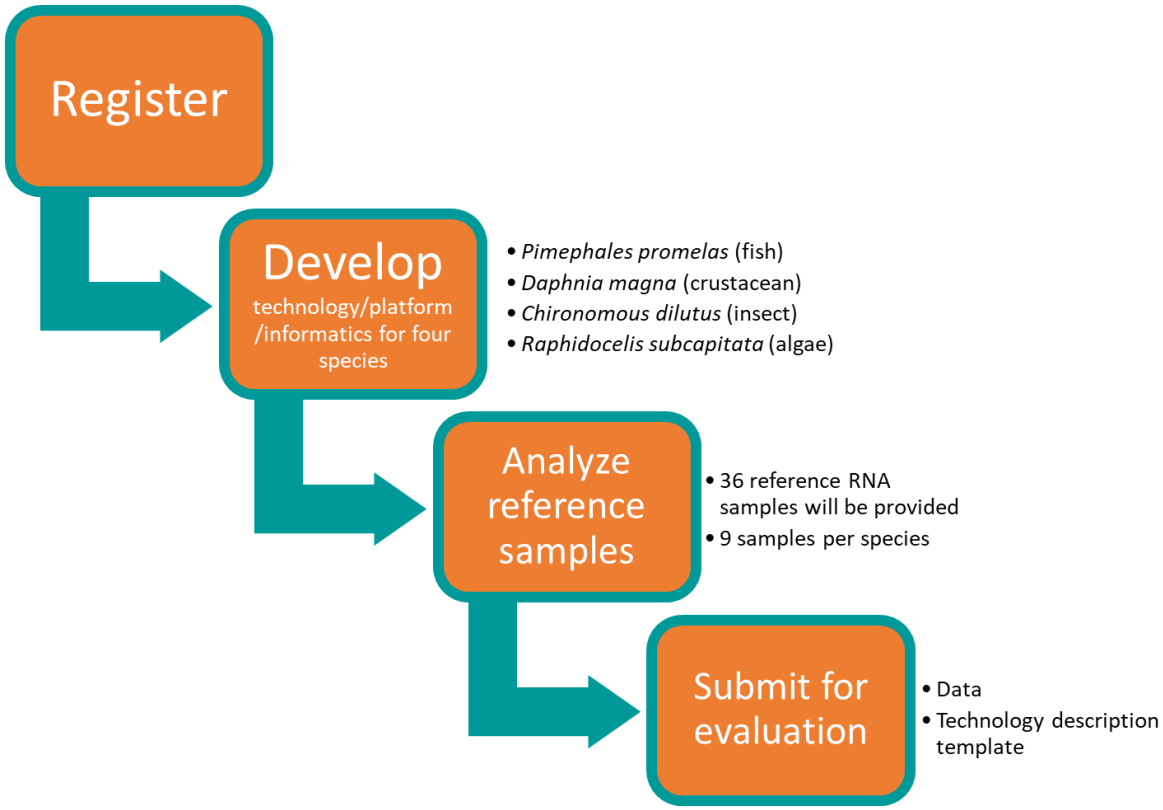
Challenge Overview

What we need:

U.S. EPA and partners seek high quality, low cost, technologies/platforms for evaluating global gene expression in samples (RNA or tissue homogenates) from non-human organisms. The technology must be capable of the following:

- Precisely quantifying the expression individual genes as reflected by relative mRNA abundance.
- Providing complete or near-complete pathway coverage for the expressed genome. This may be achieved through direct measurement of all expressed genes, or using validated sentinel gene sets whose expression can be confidently used to infer the response of the rest of the genome.
- Facilitating gene-specific annotation suitable for linking target expression with target functions, and readily accommodating updates as annotations evolve and improve over time.
- Performing analyses with limited sample masses necessary to make them compatible with high throughput testing protocols with small organisms or cells (e.g., <0.5 µg total RNA per sample).
- Generating data in formats compatible with a standardized and automated quality assurance and data analysis workflow that can be used for dose-response modeling and differentially expressed gene/pathway analysis.
- Meeting a level of quality, performance, and transcriptome coverage that is also economically and commercially viable for high throughput screening applications.

Challenge Process for Solvers:



Challenge Details

Step 1. Register

To be eligible to compete for the award, prospective solvers must register at: https://www.epa.gov/innovation/ecotox-target-challenge, no later than March 16, 2020. .

Eligibility:

- Eligible: Individuals, or teams from private companies, academic institutions, non-governmental organizations, or independent research or technological institutes. The competition is open to both US and foreign citizens/organizations.
- Not eligible: U.S. or foreign government organizations.
- Not eligible: Individuals involved in development of award selection criteria or reference sample generation.

Terminology:

- **Solver(s)**: Eligible individuals, groups, or organizations competing for the award.
- **Sponsors**: The U.S. EPA and partners involved in the development, implementation, and administration of the Challenge.

Required Information:

1. **Technical point of contact** for application (name, position, title, and contact information). This is the individual that will manage communications and coordination between the challenge sponsors (e.g., US EPA and partners) and the solver(s).
2. **Team members** (including affiliations) and partner organizations (as appropriate) that will be contributing to a registered application and will share the prize for that solution (if awarded).
3. **Financial point of contact**: Name, full address, and contact information for the organization or individual that will receive the award money and manage distribution of the funds if your solution is selected as the winner.
    - Note: If you receive the award, a single lump sum will be provided to a single point of contact, identified above. The financial point of contact will be responsible for any further dispersal or distribution of the award money to the Solver's institution and/or identified team members. Solvers are encouraged to establish and agree to an award distribution plan, in writing, prior to identifying the financial point of contact.
4. **Acceptance of the terms of participation**.
    - Solvers will not receive compensation for resources or time invested in addressing the challenge. Only the winning solution will receive a cash award.
    - Only the top-ranked solution will receive the award.
    - Solvers retain their rights to all intellectual property (e.g., details and design of their technology) that may be disclosed to the sponsors over the course of the challenge. Technical details and designs will not be disclosed or published without permission from the technical point of contact named in the registration.
    - Sponsors retain the right to disclose reference sample data, performance criteria, and other evaluation criteria summarized in the *technology description template* (see p. 8) to provide a transparent reporting of how the winning solution was selected.
    - Sponsors retain the right to publish, present, and/or otherwise publicize results of the challenge competition that does not involve the disclosure of intellectual property of the Solver(s). Solvers will be afforded opportunity to review publications, presentation, or other publicity to protect against unwanted disclosure of intellectual property.

- Solvers reserve the right to remove themselves from the competition at any time, up to final submission of results for evaluation, by notifying the sponsor in writing. The technical point of contact must make the request in writing on behalf of his/her team.
- Registration for the challenge does not confer any obligation to deliver results. However, any solvers removing themselves from the competition prior to evaluation forfeit the rights to publish results obtained for the reference samples supplied for the competition unless they obtain written consent from the challenge sponsors.
- Solvers that do not submit their results and technology description template by the submission deadline will be automatically removed from the competition and subject to the same terms as if they had forfeited in writing. The submission deadline may be extended at the discretion of the sponsors, but any extension will apply to all registered solvers.

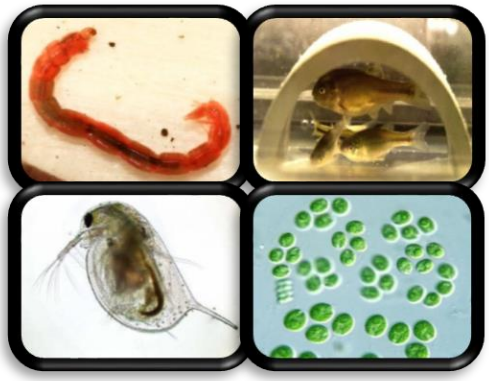
Challenge Kick-off:

All registered Solvers will be invited to an on-line Challenge kick-off meeting/webinar on March 19, 2020. The technical point of contact and/or appropriate designees should plan to attend. The kick-off meeting will provide an overview of the Challenge process and time-line and provide Solvers with an opportunity to ask questions.

The kick-off meeting will serve as the official start date for initiation of the Challenge. Registered Solvers are welcome to begin development of their solution at any time. However, the time-line for availability of method development and reference samples from the sponsors will be set relative to the kick-off date. All registered Solvers will receive reference samples at the same time (within the same 7 d period to the extent that shipping times allow).

Step 2. Develop

Each registered solver will develop and/or adapt their transcriptomic assay technology for use with the following four species:



Pimephales promelas

Daphnia magna

Chironomous dilutus

Raphidocelis subcapitata

An assay platform should consist of any means necessary to facilitate global gene expression analysis of an RNA sample input including sample pre-processing, analysis and data collection/extraction. The assay platform, consisting of supplies, reagents, processes, instrumentation, and software can remain proprietary

Solvers are challenged to provide complete or near-complete pathway coverage of the expressed transcriptome for all four species. This may be achieved through direct measurement of all expressed genes or using validated sentinel gene sets whose expression can be confidently used to infer the response of the rest of the genome. Technology development should include development of an automated and standardized pipeline or process for annotating the transcripts detected and quantified with respect to their identity (e.g., gene name or unique identifier) and biological function (if known).

Sponsors will furnish up to 1 µg of pooled RNA from each species for the Solver to use for method development. Pooled RNA for method development will be supplied within 60 d of Challenge kick-off. The technical point of contact for the Solver(s) should provide an overnight shipping address and desired date(s) for delivery of the method development samples.

The assumed starting material for each technology/solution is purified RNA. At the request of the Solver, alternative sample matrices such as crude homogenates may be provided for Solvers that may want to demonstrate unique capabilities of their technology. However, evaluation will be based on reference purified RNA samples, unless such samples are incompatible with the solution developed.

Genomic/Transcriptomic Resources for Species of Interest:

For annotation and development purposes, the sponsors recommend the following sources:

Species	Source(s) for Genome/Transcriptome Information
Pimephales promelas	A reference genome assembly and annotations are available from US EPA; contact Biales.Adam@epa.gov
Daphnia magna	Genome assemblies: Strain Xinb3 – NCBI BioProject PRJNA298946 Strain SK – NCBI BioProject PRJNA490418 Other resources: Wfleabase.org
Chironomous dilutus	Chironomous tentans (Chironomous dilutus) transcriptome NCBI BioProject PRJEB1888 Kutsenko A, Svensson T, Nystedt B, Lundeberg J, Björk P, Sonnhammer E, Giacomello S, Visa N, Wieslander L. The Chironomus tentans genome sequence and the organization of the Balbiani ring genes. BMC Genomics. 2014 Sep 27;15:819. doi: 10.1186/1471-2164-15-819.
Raphidocelis subcapitata	Raphidocelis subcapitata strain NIES-35 NCBI BioProject PRJDB5653 Suzuki S, Yamaguchi H, Nakajima N, Kawachi M. Raphidocelis subcapitata (=Pseudokirchneriella subcapitata) provides an insight into genome evolution and environmental adaptations in the Sphaeropleales. Sci Rep. 2018 May 23;8(1):8058. doi: 10.1038/s41598-018-26331-6

Solvers are welcome to use alternative or additional genomic and transcriptomic resources at their discretion.

Step 3. Analyze reference samples

Around 180 days after kick-off, the Sponsors will provide each registered Solver with a set of 36 reference samples. Nine reference samples will be provided for each species. Samples will be provided blind with respect to experimental conditions used to generate each sample pool but will be labeled with respect to species of origin. All registered solvers will receive aliquots from the same stock reference sample pools and each reference sample pool will be pre-qualified by RNAseq prior to distribution to the Solvers.

Each reference sample will consist of 1 µg purified total RNA. Solvers are encouraged to test smaller amounts of RNA if they wish to demonstrate the capability of their technology to function with smaller sample masses/volumes.

Solvers should analyze each of the 36 samples using their technology. Processed data should be provided as a single table with a column for each test sample and a row for each gene or measurement feature (e.g. probe), with values representing a measure of absolute or relative expression of each gene/feature in each sample, intended for differential expression and signal:noise analysis. Solvers must also provide a clear, unambiguous mapping from columns to provided sample IDs/method variants, and from rows to gene IDs in the transcriptome annotation for each species. This table should be provided as either an Excel document, tab-delimited text, or CSV file. Solvers should also document in detail the steps taken to generate the final processed data table from the raw data and note any software or internal data that is proprietary (e.g. that solver would not include when publishing results from this platform; see Technology Description below). Solvers should be willing to furnish raw data upon request by the Sponsors.

Step 4. Submit for Evaluation

To be considered for the award, Solvers must submit processed data file(s) for the reference samples, annotation files, and a technology description document by the Challenge end date (June 14, 2021). Late submissions will not be granted, unless the deadline has been extended for all eligible Solvers.

File submission process

Data files to be used for Judging and Award selection should must be submitted to the Sponsors by the end date (June 14, 2021). Sponsors will communicate with the technical contact for each project to determine a viable mechanism for file transmission based on file size and IT security restrictions and policies of the participating organizations.

Solvers are responsible for contacting the Sponsors at least 1 month before the challenge end date (i.e., no later than May 14, 2021) to work out the appropriate mechanism for file submission. There can be no exceptions for late submission due to file transmission problems, unless contact was made on or before May 14, 2021.

Technology Description Template

Along with the data submission, each Solver needs to submit a technology description that includes the following information:

1	Transcriptomic assay platform **Title [50 characters or less]** • This title will be used for reporting results of the Challenge and for publicity relation to the winning Solution.

2	Transcriptomic assay platform **Description [1-2 pages]**
	• Briefly describe the conceptual/theoretical basis of the assay approach.
	  ○ For example, is a targeted or non-targeted approach employed?
	  ○ What is the transcript detection/and quantification technology?
	  ○ How are data collected/extracted (e.g., image analysis; fluorescence detection, other)
	  ○ Whether data output are in absolute quantities or relative values compared to reference
	  ○ Emphasize any unique technological approaches/capabilities.
	• Use of one or more figures to help illustrate the conceptual/theoretical basis of the assay approach is recommended.
	• Note the details of your assay platform description that should remain proprietary/confidential.

3	**Quality control features** • Describe quality control features that are incorporated into the assay platform and how they are used to assure data quality. • Include features for assuring data quality both within and across samples
4	**Sample requirements** • Solvers should report the amount of reference RNA used to generate the data submitted for evaluation. • Solvers have the option to submit supplementary data demonstrating assay performance trade-offs if smaller sample amounts were to be used, but this is not required. • If the platform can accommodate or is designed for sample types other than purified RNA, that should be noted here as well.
5	**Per sample cost** • Solvers should track and report the per sample cost associated with processing and analysis from the receipt of reference samples to the output of the data submitted for evaluation. • This cost should include the cost of supplies, labor, equipment wear and tear, etc. and should reflect a viable commercial cost per sample charge if these assay were to be conducted as a contract service.
6	**Downstream data analysis requirements (if appropriate)** • Solvers should also document in detail the steps taken to generate the final processed data table from the raw data. • Note any software or internal data that is proprietary (e.g. that the Solver would not include when publishing results from this platform). • If proprietary software is required, the Solver should provide the name of the software and licensing costs (if any). These will be factored into evaluation of the overall cost per sample.

7	**Commercial viability** • Future implementation of an ecological high throughput transcriptomics-based chemical screening program is expected to generate thousands or even tens of thousands of samples per year for transcriptomic analyses. Consequently, the winning Solution will require a commercially viable level of throughput that can meet the potential sample demand, including production of supplies, reagents, and the infrastructure necessary for sample processing and analysis. Solvers must describe a pathway and timeline to a commercially viable level of throughput that can reasonably meet HTP sample demand.

Step 5. Judging and Award Selection

Solver submissions will be evaluated within 120 days of the Challenge end date. Submissions will be judged based on the following criteria:

1. Quality and performance of the platform and associated data (40%)
2. Economic and commercial viability of the platforms (30%)
3. Transcriptome coverage (30%)

Scoring Overview

Scoring will be based on the weighted (% of total score) criteria provided in the tables below. Each criterion is scored based on either a nominal or fractional scale of 0-5 with 0 being the lowest and 5 being the highest, or by pass/fail with 5 being pass and 0 being fail.

1. Quality and performance	Points	Weighting

• Quality control: Does the platform contain a quality control system that addresses consistency within and between samples and is consistent with current standards used within various platforms for transcriptomic analyses?
  ○ E.g., Microarray chip-based platforms should contain hybridization controls, redundant positional controls to evaluate edge effects, etc.
  ○ E.g., RNA-seq platforms provide number of reads per sample, base quality score by cycle, nucleotide distribution by cycle, GC content, etc.
  ○ Note – scoring for this category will require subjectivity from the judging panel
	0 to 5	10%

• Data collection/extraction: Are the data collection/extraction methods and expression normalization/quantification methods described in adequate detail? Are they compatible with the ToxCast high throughput transcriptomics data analysis pipeline?	0 to 5	5%
• Precision: Precision will be determined by evaluating 1) coefficients of variation of gene expression values across unblinded technical duplicates, 2) correlation analysis of fold-change profiles between selected reference samples, resulting in a metric of concordance, and 3) clustering unblinded reference samples when analyzed together with all conditions. Results from these three analyses will be normalized and merged into a multiplier between 0-1 that will be used to determine the total score between 0-5.	multiplier x 5	10%
• Accuracy: Accuracy will be determined by evaluating 1) the percent concordance between fold-change values determined by Solver data compared to the fold-change values determined by pre-qualification. Results to determine the total score between 0-5.	0 to 5	10%
• Quantity of RNA: Was the quantity of reference RNA used per analysis tracked and reported (Y = points awarded; N=0 points)?	(0 to 5)	5%

- Results generated using < 1 µg total RNA = 2 pts
- Results generated using < 0.25 µg total RNA = 1 pt
- Results generated using < 0.1 µg total RNA = 1 pt
- Results generated using < 0.01 µg total RNA = 1 pt

2. Economic and commercial viability	Points	Weighting
• **Economic viability** (i.e., cost per sample, including downstream data analysis cost) o Is the cost of sample preparation and per sample supply and reagent costs for conducting the sample analysis and generating the data provided? If proprietary downstream data analysis software is required, include a per-sample adjustment to overall sample cost based on software license cost. o Total per-sample cost is i. $20 or less = 5 pts ii. >$20-$30 = 4 pts. iii. >$30-$50 = 3 pts. iv. >$50-$75 = 2 pts. v. >$75-$100 = 1 pt. vi. >$100 = 0 pts.	0 to 5	20%
• **Commercial viability and throughput capability** o Is there a reasonable demonstration/description of how and when the Solver would be able to meet the potential throughput requirements of HTP sample generation?	0 to 5	10%

3. Coverage	Points	Weighting
• **Approach and annotation** o Is the approach taken for detection and quantification of transcript expression adequately described? (e.g., whether the platform employs a targeted or non-targeted analysis and the general means by which the platform detects and quantifies transcript presence and abundance) o Are annotation files provided with each platform/species that contain the required information and link to the data files?	0 to 5	10%
• **Transcriptome coverage** o What proportion of transcriptome coverage do the platforms have in relation to the pre-qualification standards? The mean percent coverage will be calculated across the four species' platforms and used as a multiplier to determine point value.	multiplier x 5	20%

• Species coverage o Did the solvers provide a platform and reference sample data for all four species?	Y or N	*100%

*Eligible submissions will include platforms, data and associated required information for all four species.

Notification

Sponsors will notify the Solvers, via the technical point of contact within 120 days of the Challenge end date. Solvers will receive scores for their submission along with a blinded summary of the scores attained for the competing submissions. The winning Solution will be announced at the same time.

After all Solvers have been notified, Sponsors will publicly announce the Award and the winning solution. The announcement will be accompanied by a press release, social media, and other public promotion by the Sponsors. By participating in the Challenge, the Solvers agree to public disclosure of all team members that contributed to the winning Solution, and their affiliations.

Step 6. Reporting

Following selection of the winning Solution, the Sponsors will prepare a written report summarizing the results of the TARGET Challenge. The number of registered Solvers, and Solvers submitting data and documents for the competition will be reported along with summaries of the scores obtained. The rationale for selection of the winning Solution, based on the scoring criteria will be provided. Solvers will have the opportunity to review the report to ensure non-disclosure of confidential intellectual property prior to dissemination or publication. Data for the winning Solution will be made public. Data from non-winning submissions may be made public as required to facilitate peer review and publication of the Challenge report.

Following selection of the winning Solution, reference samples and experimental details associated with the reference samples will be disclosed to all Solvers that participated in the competition. Solvers are welcome to publish results and conclusions related to their analysis of the reference sample set but should include relevant individuals from the Sponsoring organization as co-authors to ensure that experimental details and associated conclusions are presented accurately.