DARTpaths: Mapping developmental and reproductive toxicity (DART) genes and pathways for cross-species comparison of toxic compound effects

Background

Developmental and reproductive toxicity (DART) testing assesses the potential effects of new chemicals on adult fertility and sexual behaviour, embryo implantation and the development of the foetus. These studies use large numbers of animals (primarily rats and rabbits), are expensive and time consuming, and the relevance to humans of effects seen in these studies is not always clear.

In recent years, there has been increasing interest in using cell-based, non-mammalian and computational approaches to improve our ability to predict chemicals with DART potential. This has coincided with the development of the Adverse Outcome Pathway (AOP) concept which links molecular initiating event(s) (MIE) caused by a chemical interaction at a molecular level, with adverse effects in an organism or population through a scientifically proven set of biochemical pathways. Advances in this area could enable the prediction of an adverse outcome within an organism in the absence of animal studies.

However, the current understanding of DART AOPs is incomplete and, potentially, many mechanisms could result in DART effects (see National Research Council, 2000 and Appendices 1 and 2 for a partial list of potential MIEs and the processes they participate in that, if perturbed, may cause DART effects). Integrating DART data generated in human cells (ToxCast program, Lamb J *et al*., 2006) and alternative non-mammalian model organisms (e.g. zebrafish, *C. elegans, D. discoidium*) (van der Laan JW *et al*., 2012, PREDART, 2012) in DART-specific AOPs may improve the translation of data generated in these systems to effects in humans. With many of the related mechanisms and developmental routes assumed to be shared, at least in part, by most eukaryotic species, the expectation is that a common response measured at the molecular level can be translated to a molecular response in human and thus more easily, to a properly predicted effect in human.

The key to unlocking the potential of these alternative test systems is to understand the conservation of key pathways across different species, including non-mammalian models, conventional mammalian test species and humans, and to identify and link genes with the same function and role in driving DART between organisms. The challenge here is to develop an approach to integrate the available information on the relation between particular genes and particular physiology in the various model organisms and the mapping of those relations between model organisms in functional AOPs.

3Rs benefits

DART testing requires the use of large numbers of animals. An OECD 414 prenatal development test requires ~800 rats, and ~900 rabbits for a second species test. An OECD 443 extended one generation test requires ~1400 rats for a mimimal basic design, i.e. no cohorts and extension to F2 generations.

Currently, there is no overall systematic approach (to enable data mining or query ‘what if’ questions) to assess the species relevance, usefulness and applicability domain of the available model systems (conventional *in vivo* mammalian animal models and alternative non-mammalian systems) used to assess DART. Therefore key questions such as those listed in appendix 3 often remain unanswered, requiring companies to use conventional regulatory animal DART tests. Reliably linking data of specific genes with specific responses between organisms is essential to the development of reliable alternative non-mammalian DART test systems, as well as defining the applicability domains of these models.

Addressing this Challenge offers the following 3Rs benefits:

- A reduction in the number of regulatory studies carried out in animals by using alternative non-mammalian test systems to pre-screen compounds for DART, allowing better compound pre-selection and at the same time providing a better insight into the mode of action of tested compounds.

- In the longer term, as more information is gathered about the applicability domain of the alternative non-mammlian test systems, there may be an oppourtunity to replace animal studies with alternative non-mammalian studies.

- Although this Challenge focuses on DART, the work conducted and the strategy developed could impact other areas of toxicology, pharmacology and chemical development, potentially increasing the 3Rs impact substantially.

Need for collaboration

The data and information (pathway developmental biology data, omics and mechanistic data, and *in vivo* phenotypic data) that is needed to address this Challenge has to be collected from both public (for example (Mudunuri U *et al*., 2009, Kanehisa M *et al*., 2016, Fabregat A *et al*., 2016) and appendix 4) and private (e.g. within industry) resources. There may also be insufficient mechanistic data in some chemical classes so the generation of this lacking information from *in vitro* cell models and alternative non-mammalian model organisms may be needed as part of this project.

This can only be done in a collaboration by academia, SMEs and industry with expertise that includes, but is not limited to:

- Toxicology

- Developmental biology

- (Bio-)Informatics

- Chemoinformatics

- Innovative mathematical modelling and algorithms

- Database generation and curation

- Secure IT framework and data handling

Collaboration between the disciplines provides the various perspectives that are needed to define sensible molecular routes that are linked to genes and to identify the commonality between organisms in DART.

Overall aim

- To develop a data strategy and supporting data management structure to properly integrate available information on the relation between specific genes and specific physiology, or specific compounds and specific effects, for model organisms to include human, mouse, rat, rabbit, zebrafish, fruitfly, nematode and *D. discoideum*.

- To map those gene-to-physiology or compound-effect relations properly between these organisms all with a focus on humans.

- To build and use this framework to establish two proof-of-concepts (PoCs):

    1. Identify key data gaps in mechanistic understanding and demonstrate that newly generated mechanistic data for classes of substances with these gaps (e.g. UVCBs) can be integrated;

2. To evaluate whether this framework and mechanistic data could have predicted a developmental toxicity d*e novo* in a novel chemical class (e.g. of a crop protection chemical).

- To employ the strategy and provide a mapping of the DART pathways and genes discussed in the National Research Council report (2000) and appendices 1 and 2 between mammalian and non-mammalian model organisms.

Key deliverables

- Development of a data model that:
    - Links orthologous genes between model organisms (human, mouse, rat, rabbit, zebrafish, fruitfly, nemotode and slime mould), with a focus on genes involved in developmental pathways. Various predictions are available for genes that have the same function between organisms (Kanehisa M *et al*., 2016, Huerta-Cepas *et al*., 2016). Formulation of a method that improves the appropriate linking of genes from large paralog families.
    - Links orthologous pathways between organisms and genes to these pathways. Various public resources have accumulated so-called pathway maps that are linked to genes and organisms, which can be machine-read (Mudunuri U *et* al., 2009, Kanehisa M *et al*., 2016, Huerta-Cepas *et al*., 2016). Methodology to extract relevant data on function from the public domain and to link those to the appropriate genes for the test organisms.
    - Captures the “chemical space” of chemicals that have been evaluted in each model system, thereby defining the chemical applicability domain, and permits querying of this with new chemicals to predict DART.
    - Can be used for cross-organism mapping and validation of selected DART pathways (selected DART related routes (National Research Council, 2000 and molecular initiating events linked to genes) in a set of model organisms to include human, mouse, rat, rabbit, zebrafish, fruitfly, nematode and slime mould, thereby defining the biological applicability domain of the model systems.
- Development of a data management structure that is stable and allows reliable updates of the content.
- Visualisation of the gene-pathway data and assessment of the validity of the routes and links by comparing expression data for various organisms that have been exposed to the same compound, to enable a side-by-side comparison of pathway perturbation in the different species.

<u>Phase 1 deliverables</u>

- Identification of further key competency questions (in addition to those in appendix 3) that the toxicology community has regarding models for DART prediction.
- Identification and prioritisation of key DART pathways & MIEs (based on the chemical spaces of interest) to use in the data model to demonstrate a cross-organism mapping strategy.
- An initial data model (describing the concepts and relationships between them that will be) containing the species of interest and a prioritised pathway as an exemplar.
- A plan to establish which mechanistic data from *in vitro* cell models and alternative non-mammalian model organisms in the chemical spaces being mentioned must be generated

within the project.[1]

- A plan to fully develop and implement this to meet the key deliverables as defined above.
- A proposal for long term accessibility and sustainability of the data model and tools.
- Robust plans to deliver Phase 2 of the Challenge including commercialisation and dissemination.

Phase 2 deliverables

- Develop a cross-organism mapping strategy for multiple DART pathways.
- Generation of mechanistic data from *in vitro* cell models and alternative non-mammalian model organisms for chemical classes where data is missing, and integrate such data into the model.
- Develop a user friendly graphical user interface (GUI) that permits the users to interact with the data to answer the competency questions.
- Demonstration that the framework works for the two PoCs and that additional requirements to improve the model are identified.
- Demonstration that appropriate dissemination activities/plans are in place to maximise uptake of the developed tool.

It is important to note that the CRACK IT Challenges competition is designed to support the development of new 3Rs technologies and approaches, which will improve business processes and/or lead to new marketable products. The application must include a plan to commercialise the results into a product or service. This should be taken into consideration when completing your application.

Sponsor in-kind contributions

- Know-how on DART in the chemical industry.
- Input into prioritisation of DART pathways.
- Software user requirements.
- Where available, information gathered in the PREDART CRACK IT Challenge.
- Where available, *in vitro* and *in vivo* phenotypic and omics data to support the PoCs.

[1] For some classes of chemicals there is a wealth of information on reproductive toxicology endpoints both on the mechanistic and guidance-driven conventional *in vivo* studies. However, for other classes of substances the information available may be less complete or absent. In the context of this project particularly the classes of UVCB (Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials) substances (e.g. petrochemicals/petroleum products), which comprise about a quarter of the chemicals on the European market by volume, limited mechanistic data is available while there is a great amount of guidance-driven conventional *in vivo* data present and could be shared by the sponsors. Whereas for newly invented chemical classes (e.g. in pharmaceuticals, crop protection chemicals and “green” industrial chemicals) there is the possibility to generate a wealth of *in vitro* mechanistic data but only a very limited set of conventional *in vivo* data. Therefore, unless this missing mechanistic information is generated and the ability to credibly evidence why and under what circumstances we should believe a prediction from it is established, the applicability domain (in terms of chemicals from different classes) and 3Rs impact of these approaches will be limited. Therefore, this aspect should be carefully considered in the project.

- Connection to networks of toxicologists and other expertise relevant to the Challenge, and other programs such as ToxCast.

Duration

Phase 1: six months, Phase 2: Up to 3 years

Budget

Phase 1: £100K, Phase 2: £1 million

Sponsors

Shell, Syngenta

References

ECETOC Special Report no.19: Building a Prenatal Developmental Toxicity ontology http://www.ecetoc.org/publication/special-report-no-19-building-prenatal-developmental-toxicity-ontology/.

Fabregat A *et al.* (2016). The Reactome pathway Knowledgebase. *Nucleic Acids Res* 44(D1): D481-487.

Huerta-Cepas J *et al*. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44(D1):D286-293.

Kanehisa M *et al.* (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44(D1): D457-462.

Lamb J *et al.* (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313(5795):1929-35.

Mudunuri U *et al*. (2009). BioDBnet: the biological database network. *Bioinformatics* 25(4):555-556.

National Research Council (2000). *Scientific Frontiers in Developmental Toxicology and Risk Assessment*. Washington DC: The National Academies Press.

PREDART CRACK IT challenge, 2012: http://crackit.org.uk/challenge-10-predart.

ToxCast program https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data.

van der Laan JW *et al*. (2012). Testing strategies for embryo-fetal toxicity of human pharmaceuticals. Animal models vs. in vitro approaches: a workshop report. *Regul Toxicol Pharmacol* 63(1): 115-123.

Appendix 1: List of protein targets or biological pathways that may participate in the molecular iniatiating event of a DART adverse outcome pathway (Syngenta curated).

Name of Target Protein/pathway	Target/Receptor Code
Androgen Receptor	AR
Aryl hydrocarbon	Ah
Bone protein-matrix gla protein	MGP
Cyclooxygenase-1	COX1
Cytochrome P450 (CYP26)	CYP26
Cytochrome P450 aromatase (CYP19)	CYP19
Dihydrofolate reductase	DHFR
FGF signalling pathway	FGFR
Hedgehog signalling pathway	SHH
Hedgehog signalling pathway	PTCH
Hedgehog signalling pathway	SMO
Histone deacetylase	HDAC
N-methyl-D-aspartate-receptors	NMDA
Oestrogen Receptor: alpha	Era
Oestrogen Receptor: beta	Erb
Peroxisome proliferator activated receptor	PPARA
Retinoic acid receptor (alpha)	RARA
Retinoic acid receptor (beta)	RARB
Retinoic acid receptor (gamma)	RARC
Thymidylate synthase inhibition	TYMS
Thyroid hormone receptor (alpha)	TR (alpha)
Thyroid hormone receptor (beta)	TR (beta)
Microtubule depolymerisation	TUB
Microtubule stabilisation	TUB
VEGF signalling pathway	VEGFR2
WNT signalling pathway	WNT
Cereblon	CRBN
Acetyl-CoA carboxylase	ACC1/2
Copper chelation	
dihydroorotate dehydrogenase inhibition	dhod
HPPD inhibition	hpd
orthosteric nAChR agonists	nAChR (embryonic)
5alpha Reductase	SRD5A2, also SRD5A1 & SRD5A3
Acetylcholinesterase Inhibition	AChE
Angiotensin II receptor antagonist	AGTR1, AGTR2
Angiotensin-converting enzyme (ACE)	ACE
Carbonic anhydrase	
DNA polymerases	
GABA-A receptor agonist	GABARA
Glucocorticoid receptor	GR
Lysyl oxidase	
Opiate agonist	ZOR, MOR, other subtypes?
Other enzymes involved in folate production & inhibition	
Phosphodiesterases	
Reductase involved in Vitamin K recycling	
Ribonucleotide diphosphate reductase	
Type III deiodinase	DIO3
Vitamin D receptor	VDR

Farnesyl pyrophosphase synthetase	FPPS
GABA A receptor antagonists	GABARA
mevalonate / cholesterol pathway	CYP51
mevalonate / cholesterol pathway	CYP17
mevalonate / cholesterol pathway	INSIG1, INSIG2
mevalonate / cholesterol pathway	Sc5d
mevalonate / cholesterol pathway	Dhcr24
mevalonate / cholesterol pathway	DHCR7
mevalonate / cholesterol pathway	NSDHL

Appendix 2: List of cellular behaviours and morphogenetic effects whose alteration is believed to be a key event of DART adverse outcome pathways

Cellular effects	Morphogenetic Effects
Apoptosis	Cell recruitment
Cell proliferation	Pattern formation
Motility	Biological clocks (e.g. somite clock)
Differentiation	Angiogenesis / vasculogenesis

Appendix 3: Key competency questions

- For a particular chemical that requires testing on reproductive toxicology endpoints, which alternative species is most representative for man?
- For a particular chemical that requires testing on reproductive toxicology endpoints, for which tests have examples in the same chemical space been tested?
- In cases where there is an understanding of the pathway involved for a particular chemical of interest, which alternative species has this pathway and most resembles relevant pathways in man?
- Which alternative species actively possess known key reproductive toxicology pathways, and how do these compare across the alternative species?
- Are certain species better suited than others to address specific parts of the reproductive cycle e.g. fertility (male/female), prenatal development, multigenerations?
- Is it overly conservative looking at omics patterns and/or biomarkers in comparison with phenotypic effects for the alternative species?

Appendix 4: List of additional sources of data to support the Challenge

- Elements of the embryonic development, regenerative medicine and stem cell data base (http://discovery.lifemapsc.com/in-vivo-development) developed by LifeMap and similar sources in other species.
- Efforts are ongoing to create a Developmental Toxicity Ontology to inform the processes of AOP inference and DART prediction from the perspective of perturbed normal developmental biology, known effects observed in DART studies and features of chemistry that are known to perturb it (http://www.ecetoc.org/publication/special-report-no-19-building-prenatal-developmental-toxicity-ontology/).

- NC3Rs PREDART CRACK IT Challenge: http://crackit.org.uk/challenge-10-predart.