EU-ToxRisk

An Integrated European 'Flagship' Program
Driving Mechanism-based Toxicity Testing and Risk Assessment
for the 21st Century

Deliverable 10.1:

Mapping ToxCast data to AOPs

Grant Agreement number: 681002

H2020-PHC-33-2015-single-stage

Start date of project: 1 January 2016

Duration: 72 months

Lead beneficiary of this deliverable: UCPH, INERIS

Deliverable type: Report

Dissemination Level: PU

Release: FV

Due date of deliverable: 31/12/2016

Actual submission date: 17/08/2017

Description: Report on the mapping of available ToxCast data to the EU-ToxRisk AOPs developed in WP5

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 681002

1 Executive Summary

The EU-ToxRisk project aims to drive the required paradigm shift in toxicological testing away from 'black box' animal testing towards a toxicological assessment based on human cell responses and a comprehensive mechanistic understanding of the complex chain of events that link chemical exposure to toxic outcome. EU-ToxRisk extensively integrates the Adverse Outcome Pathway (AOP)-based toxicity testing concept, by identifying and quantitatively describing the AOPs which link the effects of test compounds to adverse effects in humans. In this context, research has been carried out to link chemicals to Adverse Outcome Pathways (AOPs) through a molecular initiating event (MIE). The results of this work are documented in this deliverable.

By definition, AOPs are considered not chemical specific. In order to link chemicals to AOPs via MIEs, the ToxCast chemical library, established by the U.S. Environmental Protection Agency (EPA), was mined and matched to the 161 AOPs that are currently described and accessible at the AOP wiki website. As a result, 5,290 chemicals were matched to 82 AOPs through 48 ToxCast assays, each assay being considered as being indicative of an MIE. All chemicals identified through this process can potentially activate an AOP through to the adverse outcome (AO) and should be investigated further.

Following the initial mapping exercise the steatosis AOP was more closely examined. 29 compounds were identified in ToxCast that showed activity to at least one of the AOP steatosis assays without triggering cytotoxicity. Individual dose-response analyses were performed for each compound. Preliminary results show that the majority of these compounds tend to have a high number of responses on nuclear receptors studied in ToxCast suggesting them as potential MIEs related to steatosis. Secondly, a gene expression data analysis was performed on a set of 31 compounds associated with steatosis based on a gene set pathway enrichment analysis and a time series analysis. The work described in this deliverable confirmed that several MIEs are linked to steatosis and also identified several novel potential MIEs.

Table of Content

1 Executive Summary 1
Table of Content 2
List of Tables 3
Table of Figures 4
List of Acronyms 5
2 Introduction 6
3 Data Integration and Mapping Approach 7
3.1 ToxCast Data 7
3.2 Adverse Outcomes Pathways Data 8
3.3 Mapping ToxCast to AOP Data 9
4 Mapping ToxCast and Toxicogenomics Data to AOPs: A Case Study with Steatosis 12
4.1 Mapping ToxCast on the AOP Steatosis 12
4.1.1 Cytotoxicity and Steatosis Endpoints 12
4.1.2 Cytotoxicity and Endpoints Relative to the Steatosis AOP Events 15
4.1.3 Relationship between Receptor Dose-Responses and Steatosis 16
4.2 Mapping Toxicogenomics on the AOP steatosis 20
5 Conclusion 24
6 Bibliography 25
Annexes 27

List of Tables

Table 1: List of steatosis endpoint tested in ToxCast....13
Table 2: List of compounds with a hit call on a steatosis endpoint....14
Table 3: List of ToxCast database assays related to the AOP steatosis....15
Table 4: List of compounds associated to steatosis and with genes expression data in TG-GATEs....20
Table 5: List of genes having different expression profiles over time.... 21

Table of Figures

Figure 1: Snapshot of the AOPs table in AOP Wiki 9
Figure 2: Chemicals deregulation on 48 assays associated to AOPs 10
Figure 3: Bioactivity for the ToxCast chemicals on 48 assays associated to AOPs 11
Figure 4: Schematic representation of the steatosis AOP 12
Figure 5: Heatmap of bioactivity in ToxCast for 29 steatotic compounds 17
Figure 6: Heatmap of bioactivity for the 310 compounds tested on steatosis in ToxCast 18
Figure 7: Steatosis response for the 29 compounds as a function of each nuclear receptor response 19
Figure 8: Mapping of the gene deregulation by the 31 compounds to the NAFLD pathway 21
Figure 9: Time series analysis for 30 compounds on the ATP synthase in the mitochondria F0 complex 25

List of Acronyms

Abbreviation / acronym	Description
aeid	assay endpoint identifier
AOP	Adverse Outcomes Pathways
AO	Adverse Outcome
BMAD	Baseline Median Absolute Deviation
CAS	Chemical Abstracts Service
chid	chemical identifier
EAGMST	Extended Advisory Group on Molecular Screening and Toxicogenomics
EPA	U.S. Environmental Protection Agency
GAGE	Generally-Applicable Gene
JRC	Joint Research Centre
KE	Key Events
MIE	Molecular Initiating Events
NAFLD	Nonalcoholic Fatty Liver Disease
OECD	Organisation for Economic Cooperation and Development
PoD	Point of Departure
qAOP	Quantitative Adverse Outcomes Pathways
QSAR	Quantitative Structure Activity Relationships
SAAOP	Society for the Advancement of AOPs
WP	Work Package

2 Introduction

One of the great challenges in toxicology and risk assessment is to understand the underlying mechanisms of complex diseases, adverse drug reactions and chemical toxicity. To expand the use of mechanistic toxicological data, the Organisation for Economic Cooperation and Development (OECD) has launched the adverse outcome pathway (AOP) concept connecting biological perturbations by chemical hazards to adverse outcomes [1]. So far, only a limited number of compounds are linked to AOPs. The U.S. Environmental Protection Agency (EPA) has developed a programme (ToxCast) aimed at testing large libraries of chemicals using a large set of high throughput toxicity assays. Exploiting this data to link chemicals to AOPs through a molecular initiating event (MIE) would be of great interest for risk assessment and regulatory applications. Further, integrating and exploring these connections with computational approaches would help in the development, confirmation or optimisation of AOPs [2].

In this context, this deliverable aims to map the ToxCast data to AOPs that are published on the AOPwiki site. Such mapping will allow for an overview of the biological assays already implemented in ToxCast that are connected to AOPs but also to detect some MIEs where no assays have yet been developed in the ToxCast programme and for which EU-ToxRisk could contribute assays. Furthermore, this process allows collecting chemicals from ToxCast that have been determined to be bioactive on these MIEs (defined as proteins, genes or cells assays in ToxCast). For some assays, many compounds are bioactive whereas for some others assays, only a few compounds have been tested. These results could guide the EU-ToxRisk partners to assess the bioactivity of additional compounds.

Once the mapping was developed, we further investigated the steatosis AOP by analysing the chemicals tested for steatosis in ToxCast. Based on an individual dose-response model on 29 compounds, a majority of nuclear receptors were impacted, assuming their role as MIEs related to steatosis. In a second study, a gene expression data analysis was performed on a set of 31 compounds associated with steatosis. With this study, it is possible to confirm several MIEs linked to steatosis but also to suggest potential new MIEs.

Mapping of the ToxCast data to the AOPs and studies on the steatosis AOP are described further in the next chapters.

3 Data Integration and Mapping Approach

In order to set up the mapping process between the data in ToxCast and a list of AOPs, we needed to collect the data internally and to select the most useful information. We will first describe the ToxCast data collected, then the AOPs data gathered and finally how we connected the ToxCast data to AOPs.

3.1 ToxCast Data

The ToxCast programme within the U.S. EPA aims at testing a large library of chemicals using *in vitro* high-throughput screening (HTS) approaches to support the development of improved toxicity prediction models. It started in 2007 and has since contributed to the Tox21 testing programmes. Currently, the ToxCast chemical library contains 9,076 compounds tested on 1,192 endpoint assays [3]. The full data is accessible at the EPA website, where we downloaded the database (in MySQL format) as well as the R package to facilitate the process and analysis of the chemical screening data.

The assays in ToxCast can be organism-based, cell-based or biochemical and are checked for negative/positive or down/up regulation. For example, a collection of biochemical assays measuring binding constants and enzyme inhibition values of chemicals for CYP450 and nuclear receptors was considered in the study. High-content imaging assays that test a chemical's effects on a range of phenotypes in either the human hepatoma cell line HepG2 or rat primary hepatocytes was reported, including apoptosis and hepatic steatosis. Gene expression assays in human primary hepatocytes were considered, measuring the changes of key nuclear receptor target genes, phase I and II metabolic enzymes and transporters. Chemicals were also tested in transcription reporter assays for 48 transcription factor binding sites in HepG2 human liver hepatoma cell lines. In addition, cytotoxicity assays, genotoxicity assays and real-time cell electronic sensing (RT-CES) to measure the time-dependent response to chemicals were also performed. Finally, complex cell systems based on a biologically multiplexed activity profiling (BioMAP) using primary human cells were developed to characterise effects relevant to human tissue and inflammatory disease following exposure to chemicals. Interestingly, for some studies (cytotoxicitiy, high content imaging assays) chemicals have been tested at different concentrations and different times, which can be used for dose-time response curves.

In addition, chemical information is also reported in ToxCast. Data about the chemical structure (SMILEs code and InChi keys format), substance name, CAS Number as well as Quality Control (QC) level is available. The mass identification, stability and purity (>97%) were assessed using liquid chromatography-mass spectrometry (LC/MS). Such information can be useful, for example, for the development of Quantitative Structure Activity Relationships (QSARs) and cheminformatics approaches intended to be developed in Work Package (WP) 3 of the EU-ToxRisk project on 'QSAR, cheminformatics and bioinformatics appraoches'. Overall, information about the assay and chemicals considered in ToxCast can be found at the EPA website.

3.2 Adverse Outcomes Pathways Data

An AOP is a conceptual framework for organising existing knowledge concerning the predictive and/or causal linkages between measurable/observable biological changes (termed key events; KEs) that are essential to the progression from a molecular initiating event (MIE) to an adverse outcome (AO) considered relevant to regulatory decision making [4]. In 2012, the OECD initiated an international AOP development programme in order to introduce some standardisation and rigour to assure that AOP descriptions included the information required to facilitate assessment of the types of measurements and weight-of-evidence supporting an AOP. To help both sharing of AOP knowledge and development, an AOP-wiki was built as a user-friendly, open source interface.

In our study, the AOP wiki was considered for the mapping of ToxCast chemicals. This AOP wiki represents a joint effort between the Joint Research Centre (JRC) and the Environmental Protection Agency (EPA) to support the OECD AOP knowledgebase effort and represent the central repository for all AOPs developed as part of the OECD AOP development effort by the Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST). The AOP development programme is overseen by the OECD, the EAGMST and the society for the advancement of AOPs (SAAOP), which play an active role in the internal review process for approvals/development of an AOP.

In the AOP wiki, a list of 161 AOPs is accessible. The AOPs are categorised under 6 levels of approvals/development:

- AOP endorsed by OECD (approved)
- EAGMST approved
- EAGMST under review
- EAGMST under development
- SAAOP (Society for the Advancement of AOPs)
- Recently added.

For each AOP a name, an MIE and an Adverse Outcome (AO) is provided. For example, the aromatase inhibition leading to reproductive dysfunction AOP (AOP 25) has been endorsed by the OECD whereas the 5-hydroxytryptamine transporter (5-HTT; SERT) inhibition leading to decreased shelter seeking and increased predation AOP (AOP 98) is under development, meaning that further information is needed to approve this AOP.
We can notice that a single MIE can be a protein, a gene, DNA alkylation and can be linked to several AOPs (for example the Aryl-hydrocarbon receptor (AhR) is linked to AOP21, AOP57, AOP131). In addition, each AOP identifier is linked to only one MIE. However, several MIEs might contribute to the same AO (see Figure 1).

Name	MIE	AO	Status
EGFR Activation Leading to Decreased Lung Function	EGFR		SAAOP AOP Under Development
5-hydroxytryptamine transporter (5-HTT; SERT) inhibition leading to decreased shelter seeking and increased predation	5-HTT	increased predation	EAGMST Under Development
5-hydroxytryptamine transporter (5-HTT; SERT) inhibition leading to population decline	5-HTT	increased predation	EAGMST Under Development
5-hydroxytryptamine transporter (5-HTT) inhibition leading to population increase			

Figure 1: Snapshot of the AOPs table in the AOP Wiki

3.3 Mapping ToxCast to AOP Data

To perform the mapping between both databases, we had to link common information. In our case, we chose the gene symbol ID in ToxCast (corresponding to a gene name symbol in HUGO Gene Nomenclature Committee; HGNC) and the MIE ID in AOP, which is also a gene name symbol in most cases. For some MIEs in the AOP wiki, a gene name synonym was considered. For example, to define the serotonin transporter, the term 5-HTT was used in the AOP wiki and the gene name SLC6A4 in ToxCast. Therefore, we converted the MIE term in the AOP wiki into a gene ID for simplification. A second point was to collect only the type of activity of interest for a defined AOP through the activity profile defined in ToxCast (i.e down/up regulated or agonist/antagonist/inhibitor). In the AOP table, the term “activation” or “inhibition” is largely employed, assuming a direct effect of a chemical to an MIE (protein). Therefore, in ToxCast, we considered only chemicals that have a type of activity (binding or deregulation) that follow the assigned effect in the AOP table. For example, we mapped only chemicals in ToxCast depicting an “inhibition” or “deregulation” of AChE to the AOP “acetylcholinesterase **inhibition** leading to acute mortality” and we did not consider the compounds having the opposite effect for this AOP.

The list of AOPs with the MIE associated to it, the corresponding genes tested in ToxCast and the number of compounds matching the AOP are listed in the table “AOP_toxcast_match_Cpd.xlsx” (see Annexes). Overall, 48 assays in ToxCast have been matched to 82 AOPs in the AOP wiki. The table provides an overview of the MIEs on which compounds have been evaluated as active in the different ToxCast assays. For example, there are more than 1,000 actives compounds on AhR (agonist), Cyp19A1 (inhibitors), ESR1 (inhibitors), PXR (agonist), PPARG (agonist) or AR (antagonist) whereas there are 0 compounds active on ADRB2 (agonist). Also, some proteins/genes listed in the AOP wiki have not been tested in ToxCast (e.g. ABCB11 inhibitors, DOI2 inhibitors, HPPD, etc...). Some of the compounds that trigger the MIE associated with an AO in the AOP wiki are confirmed in

ToxCast. This is the case for example for the compound 2,3,7,8-TCDD, an AhR agonist which is suspected to lead to rodent liver tumours (AOP 41) or roziglitazone, a PPARγ agonist which is involved in the impaired fertility (AOP7). However, many other compounds are also measured in ToxCast for these MIEs and could be susceptible to have an impact on the AO described in the AOP wiki.

In addition, the chemicals tested and active for these 48 assays in ToxCast were integrated in 2 heat maps in order to facilitate the visualisation of the chemical-MIE associations. A first one is based on the binary (up/down regulation) information (file InteractiveHeatmap-UpDownregulated.html, see Annexes) (Figure 2).

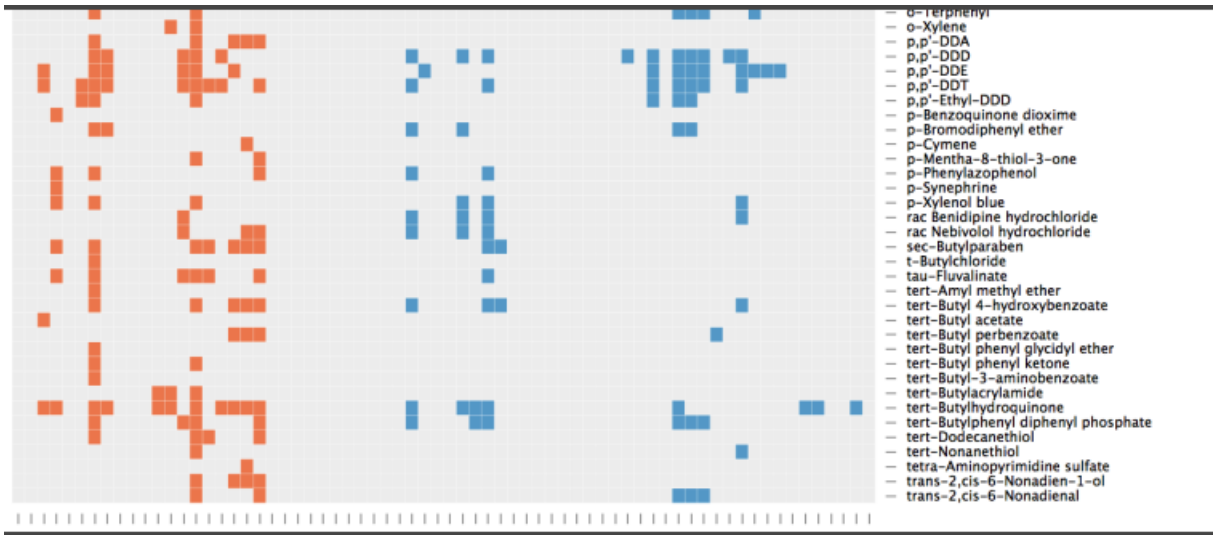

Figure 2: A snapshot of the chemicals deregulation (y axis) on the 48 assays (x axis). Red square means up regulated and blue depicted down regulated

A second heat map depicts the AC50 value (in µM) when data is available (file InteractiveHeatmap.html, see Annexes). In this condition, the range of activity is from 10e-6 µM (yellow square) to 50000 µM (dark red) (Figure 3).

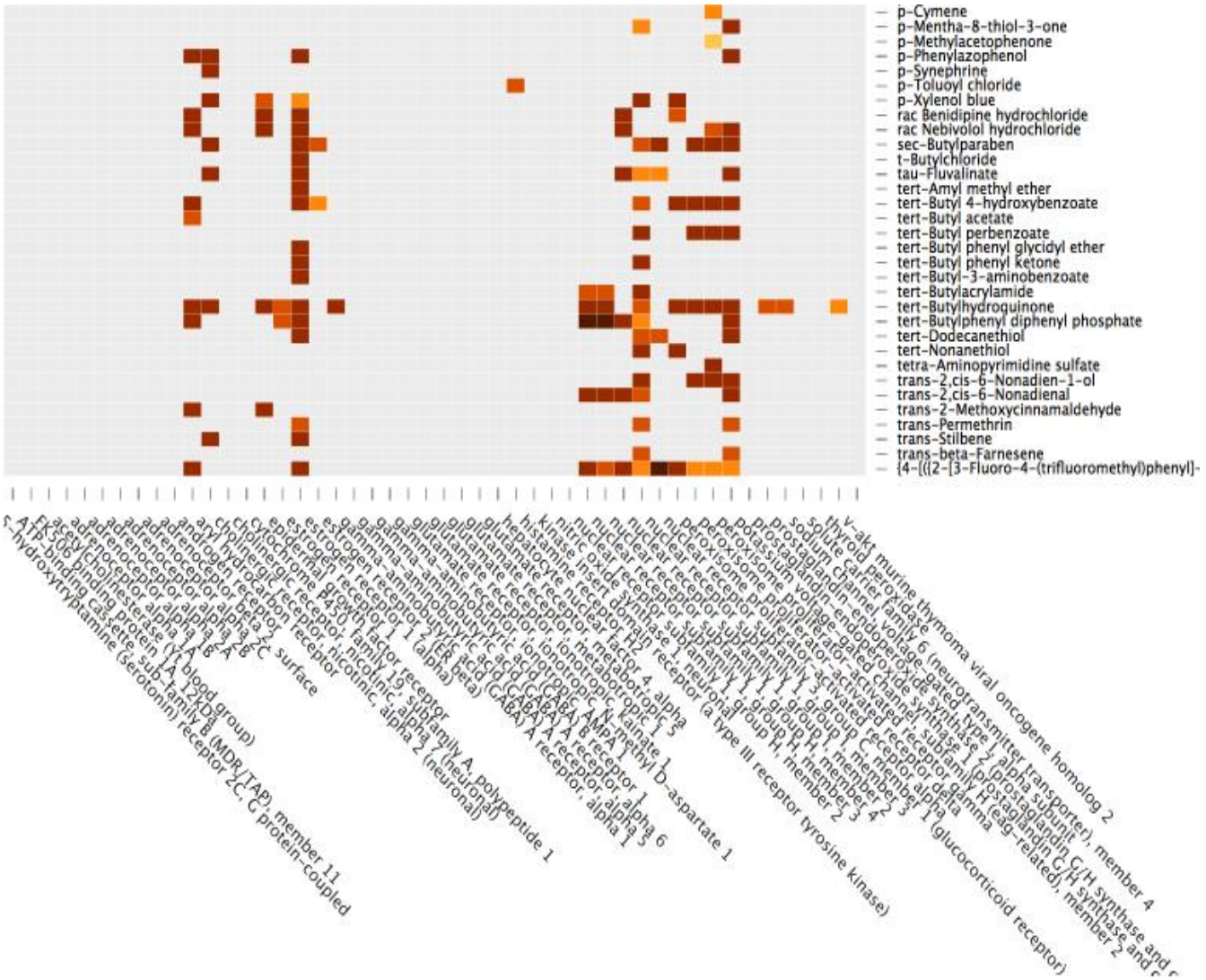

Figure 3: A snapshot of the measured bioactivity (AC50 value in µM) for the ToxCast chemicals (y axis) on the 48 assays (X axis)

Overall, 5,290 chemicals measured on 48 assays in ToxCast were mapped to MIEs for 82 AOPs in the AOP wiki. Interestingly, for some MIEs, no assays have been developed in ToxCast and could guide the development of new assays. Furthermore, for some assays, no chemicals have been measured active and so new chemicals could be tested for these assays. Finally, for some assays, a large set of chemicals were measured active and could be potentially involved as "stressors" of an MIE in an AOP.

4 Mapping ToxCast and Toxicogenomics Data to AOPs: A Case Study with Steatosis

EU-ToxRisk centers its work around case studies, which drive the selection of chemicals, and that are chosen carefully to support development of an AOP-based integrated approach for testing and assessment (IATA). One of the EU-ToxRisk case studies is related to steatosis. We decided to investigate this AOP by mapping the ToxCast data and by developing a toxicogenomics data analysis to validate and to identify potential new gene signatures associated to the steatosis AOP.

4.1 Mapping ToxCast on the AOP Steatosis

4.1.1 Cytotoxicity and Steatosis Endpoints

Steatosis is an accumulation of triglycerides within the cytosol of hepatocytes and is associated with extra-hepatic diseases. A variety of drugs and environmental chemicals can induce steatosis leading to non-alcoholic fatty liver disease (NAFLD) [5]. In the AOP wiki, an AOP on steatosis is under development by EAGMST and SAAOP for which several MIEs are considered. A general schematic representation of the steatosis AOP is shown in Figure 4.

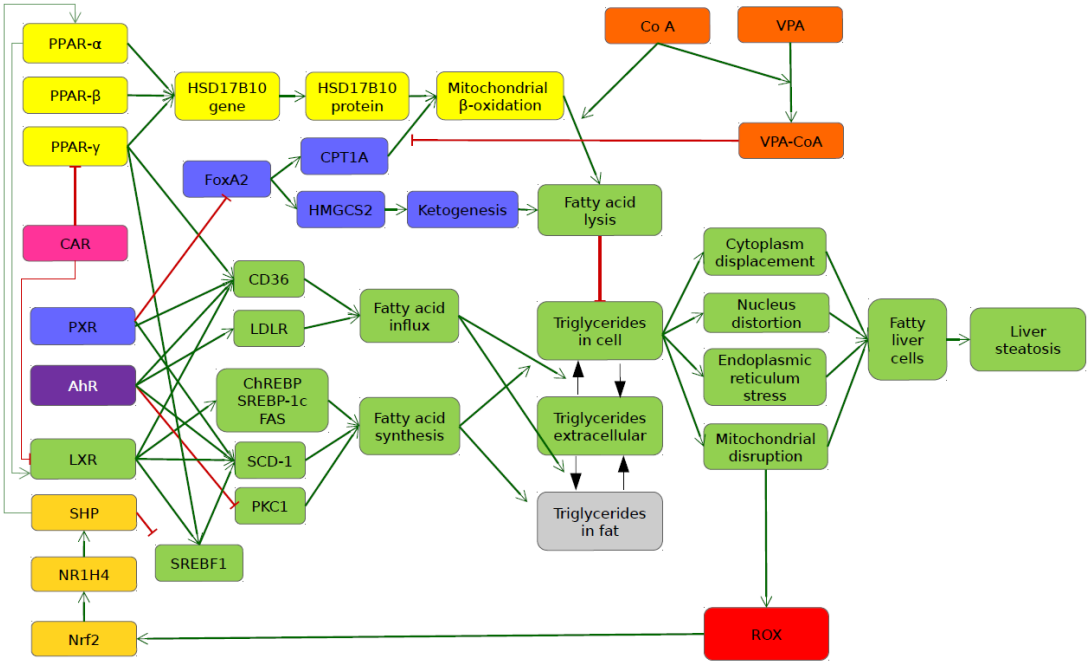

Figure 4: Schematic representation of the steatosis AOP.

In the ToxCast database, compounds were tested for steatosis with 1 assay on rat liver primary cells at 3 timepoints, i.e. 1 hour, 24 hours, and 48 hours. Both up and down regulation were considered as endpoints in the ToxCast analysis (Table 1). 310 compounds were tested on at least one of these 6 endpoints.

Active compounds on each endpoint are identified in the ToxCast high-throughput screening data by "hit-calls". In the ToxCast analysis pipeline [6], all sample dose-responses are analysed and the best dose-response model, in terms of Akaike information criteria (AIC) is selected. If the modelled and observed response are greater than a cut-off value that is defined according to the baseline variability for the endpoint (often 6 times the baseline median absolute deviation) then the compound is considered to be active and the hit-call is set to 1.
Initially, out of the 310 compounds tested for steatosis, 40 were identified as producing a hit call on a steatosis endpoint (Table 2).

aid (assay identifier)	aeid (assay endpoint identifier)	assay endpoint component name
383	1151	APR_Hepat_Steatosis_1hr_dn
383	1152	APR_Hepat_Steatosis_1hr_up
384	1167	APR_Hepat_Steatosis_24hr_dn
384	1168	APR_Hepat_Steatosis_24hr_up
385	1183	APR_Hepat_Steatosis_48hr_dn
385	1184	APR_Hepat_Steatosis_48hr_up

Table 1: List of the 6 steatosis endpoints tested in ToxCast and consiedered for the analysis**.**

In the ToxCast data, a large number of hit-calls are likely related to general cytotoxicity, which produces a "burst effect" at high concentrations [7,8]. Judson et al. [8] defined a z-score in a way that for chemicals with 2 or more hits in cytotoxicity assays, the AC50 values on each endpoint were compared to the AC50 on cytotoxicity assays and to the distribution of variances between AC50 values across the cytotoxicity assays. A large z-score (> at 3) indicates that activity occurred at concentrations far below the cytotoxicity threshold. These hits are hypothesised to be more likely associated with specific biomolecular interactions with the assay target.

The 310 compounds were therefore considered to be steatotic if their response was a hit-call on at least one of the 6 steatosis endpoints, at concentrations where cytotoxicity was not observed. On the basis of 33 cytotoxicity-related assays and a calculated z-score > 3, only one compound in ToxCast was considered to be steatotic at non-cytotoxic concentrations. In order to preserve a sufficient amount of data, we decided to consider cytotoxicity in a more specific way.

Steatosis was tested on hepatocytes. Amongst the cytotoxicity endpoints, the Tox21 MMP viability assay appeared to be the most relevant because it was also based on HepG2 cells. Individual dose-response models on endpoint Tox21 MMP viability (Assay Endpoint ID (aeid) 799) were collected from the SQL ToxCast database. This endpoint had been tested for all 40 compounds (15 of them had been tested twice). A total of 30 of the compounds tested had at least one dose-response modelled.

If the maximum-modelled response was smaller than 6 Baseline Median Absolute Deviation (BMAD) for a compound, we considered there was no cytotoxicity. According to this criterion, cytotoxicity occurred for 14 compounds but not necessarily at the concentrations where steatosis occurred: Chemical ID (Chid) "20154", "20319", "21409", "28038", "32329", "32372", "32500", "32573", "32580", "34492", "34566", "34609", "34956", "40362".

For each of these compounds, if the Activity Concentration at 10% (AC10) on the cytotoxicity endpoint was smaller than the AC90 on the steatosis endpoint, then the dose-response curves overlapped and cytotoxicity was considered to interfere with the measurement of steatosis. As a result, for 9 compounds, there was no steatosis response at non-cytotoxic concentrations.

They were removed from the analysis (they are greyed in**Error! Reference source not found.** Table 2 nd their plots are greyed) as they are considered non-steatotic compounds: chemical identifier (chid) "20154", "28038", "32329", "32372", "32500", "34492", "34609", "34956", "40362".
For two of the remaining cytotoxic compounds, the steatosis responses that occurred at non-cytotoxic concentrations were not hit-calls (Chid "32573" and "32580"). These were also removed from the analysis (they are grey in Table 2).

Chid	Chemical name	CAS	Steatotic only at cytotoxic concentrations
20154	Clorophene	120-32-1	X
20182	Bisphenol A	80-05-7	
20319	Chlorothalonil	1897-45-6	
20794	Maneb	12427-38-2	
21332	Thiram	137-26-8	
21409	Triphenyltin hydroxide	76-87-9	
22265	Alachlor	15972-60-8	
22325	2,2-Bis(4-hydroxyphenyl)-1,1,1-trichloroethane	2971-36-0	
23848	Acetochlor	34256-82-1	
23869	Ametryn	834-12-8	
23892	Abamectin	71751-41-2	
24048	Difenzoquat metilsulfate	43222-48-6	
24235	Flusilazole	85509-19-9	
24270	Prochloraz	67747-09-5	
24276	Propargite	2312-35-8	
24337	Thiobencarb	28249-77-6	
27204	Methyl isothiocyanate	556-61-6	
28038	3-Iodo-2-propynyl-N-butylcarbamate	55406-53-6	X
32329	Bensulide	741-58-2	X
32372	Difenoconazole	119446-68-3	X
32376	Dimethenamid	87674-68-8	
32464	Profenofos	41198-08-7	
32498	Triclosan	3380-34-5	
32500	Triflumizole	68694-11-1	X
32550	Fenpyroximate (Z,E)	111812-58-9	
32573	Pyridaben	96489-71-3	X
32577	Tefluthrin	79538-32-2	
32578	Thiodicarb	59669-26-0	
32580	Trifloxystrobin	141517-21-7	X
32638	Pyraclostrobin	175013-18-0	
32647	2-(Thiocyanomethylthio)benzothiazole	21564-17-0	
34492	Cyazofamid	120116-88-3	X
34566	Emamectin benzoate	155569-91-8	
34588	Famoxadone	131807-57-3	
34590	Fenamidone	161326-34-7	
34609	Fipronil	120068-37-3	X
34625	Fluoxastrobin	361377-29-9	
34818	Azamethiphos	35575-96-3	
34956	Tetraconazole	112281-77-3	X
40362	Niclosamide	50-65-7	X

Table 2: List of compounds with a hit call on a steatosis endpoint**.** Chid is a chemical identifier. Chemical Abstracts Service (CAS) number is a Chemical Substances identifier. Cytotoxic response at a lower dose than steatotic response is marked with an X.

Overall, 29 compounds are considered as steatotic and investigated further.

4.1.2 Cytotoxicity and Endpoints Relative to the Steatosis AOP Events

34 endpoints from the ToxCast database were related to the steatosis AOP (Table 3). They belong to several assay sources, namely Attagene (ATG), Bioseek (BSK), Novascreen (NVS), Odyssey Thera (OT), and Tox21. Several endpoints have the same targets, for example, PPARg is targeted by 4 different endpoints. Endpoints identified by aeid 517, 518, 713, 715, 716, 718, 719, 721 and 754 were not measured with all the steatotic compounds. For the dose-responses on receptors, even small magnitude responses were of interest for the AOP, so the information about hit calls was discarded: all dose-responses modelled in ToxCast were kept.

In the previous section, 3 compounds were identified that were cytotoxic at concentrations higher than the range where steatosis occurred. For these compounds cytotoxicity was considered to interfere with the receptor response, if for a receptor dose-response from cell-based assays, the AC10 on the cytotoxicity endpoint was smaller than the AC90 on the receptor endpoint. If a compound had been tested several times on a receptor, the test was considered acceptable only if all the dose-response curves met that criterion.

As a result, among the 18 cell-based assays, the only two dose-responses that were removed from the analysis due to cytotoxicity were from aeid 85 and 126 for compound “34566” (Emamectin benzoate). All dose-responses for the compounds “20319”, “21409” and for the 26 compounds without any cytotoxicity hit-call can be used (see table 3):

assay component endpoint name	official symbol	Receptor	aeid
ATG_Ahr_CIS_up	AHR	AhR	63
ATG_DR4_LXR_CIS_up	NR1H3	LXR	70
ATG_IR1_CIS_up	NR1H4	FXR	85
ATG_NRF2_ARE_CIS_up	NFE2L2	NRF2	97
ATG_PBREM_CIS_up	NR1I3	CAR	101
ATG_PPRE_CIS_up	PPARA	PPARa	102
ATG_PXRE_CIS_up	NR1I2	PXR	103
ATG_SREBP_CIS_up	SREBF1	SREBF1	107
ATG_CAR_TRANS_up	NR1I3	CAR	116
ATG_FXR_TRANS_up	NR1H4	FXR	120
ATG_LXRa_TRANS_up	NR1H3	LXR	125
ATG_LXRb_TRANS_up	NR1H2	LXR	126
ATG_PPARa_TRANS_up	PPARA	PPARa	132
ATG_PPARd_TRANS_up	PPARD	PPARd	133
ATG_PPARg_TRANS_up	PPARG	PPARg	134
ATG_PXR_TRANS_up	NR1I2	PXR	135
BSK_CASM3C_LDLR_down	LDLR	LDLR	213
BSK_CASM3C_LDLR_up	LDLR	LDLR	214
NVS_ENZ_hPKCz	PRKCZ	PKC1	517
NVS_ENZ_hPKCz_Activator	PRKCZ	PKC1	518
NVS_NR_hCAR_Agonist	NR1I3	CAR	712
NVS_NR_hCAR_Antagonist	NR1I3	CAR	713
NVS_NR_hFXR_Agonist	NR1H4	FXR	715
NVS_NR_hFXR_Antagonist	NR1H4	FXR	716

NVS_NR_hPPARa	PPARA	PPARa	718
NVS_NR_hPPARg	PPARG	PPARg	719
NVS_NR_hPXR	NR1I2	PXR	721
OT_FXR_FXRSRC1_0480	NR1H4	FXR	753
OT_FXR_FXRSRC1_1440	NR1H4	FXR	754
OT_PPARg_PPARgSRC1_0480	PPARG	PPARg	757
OT_PPARg_PPARgSRC1_1440	PPARG	PPARg	758
TOX21_PPARg_BLA_Agonist_ratio	PPARG	PPARg	802
TOX21_AhR_LUC_Agonist	AHR	AhR	806
TOX21_ARE_BLA_agonist_ratio	NFE2L2	NRF2	1110
TOX21_FXR_BLA_agonist_ratio	NR1H4	FXR	1119
TOX21_FXR_BLA_antagonist_ratio	NR1H4	FXR	1120
TOX21_PPARd_BLA_agonist_ratio	PPARD	PPARd	1124
TOX21_PPARd_BLA_antagonist_ratio	PPARD	PPARd	1125
TOX21_PPARg_BLA_antagonist_ratio	PPARG	PPARg	1127
ATG_SHP_TRANS2_up	NR0B2	SHP	1366

Table 3: List of ToxCast database assays related to the steatosis AOP**.** The colours refer to the event in the AOP schema in Figure 4

4.1.3 Relationship between Receptor Dose-Responses and Steatosis

A first graphical overview of whether the nuclear receptor responses were linked to steatosis can be provided by heat maps of all responses on the 29 steatotic compounds (Figure 5 and Figure 6). The heat map in Figure 5 suggests that one group of endpoints is more related to the 24 hours and 48 hours steatosis measurements (the 13 endpoints to the left of the heat maps). The heat map in Figure 6 suggests that the 29 steatotic compounds tend to have more responses on the nuclear receptors studied in ToxCast.

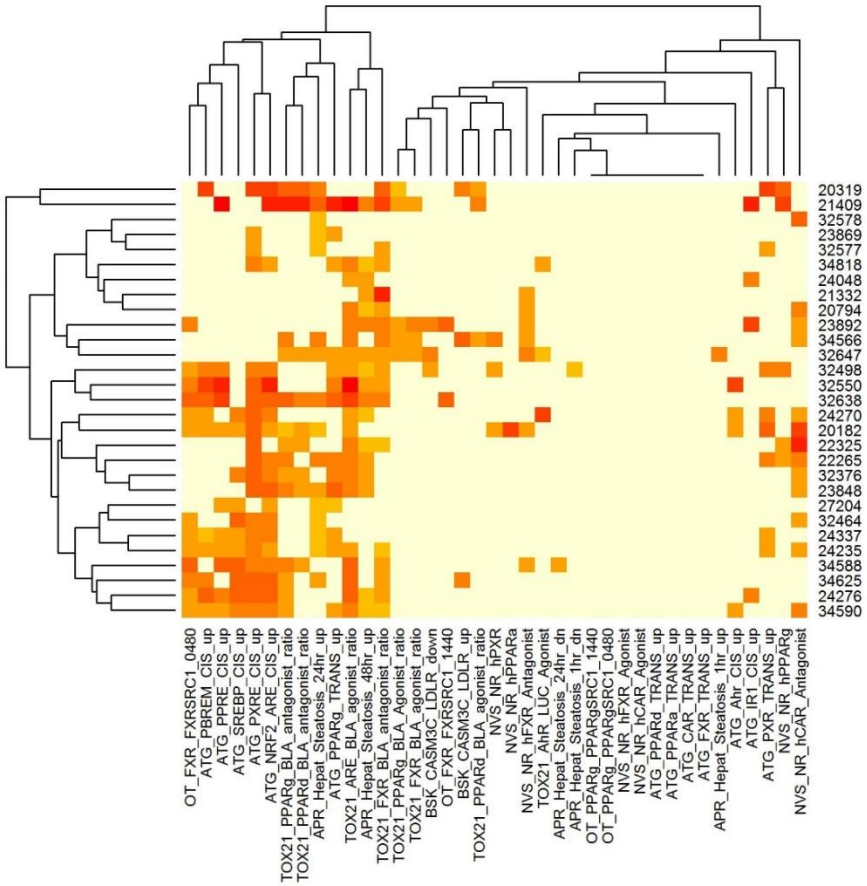

Figure 5: Heat map of -logAC50s on all endpoints for the 29 steatotic compounds. Red indicates a low AC50, pale yellow indicates no response (AC50=0). These are not necessarily hit calls

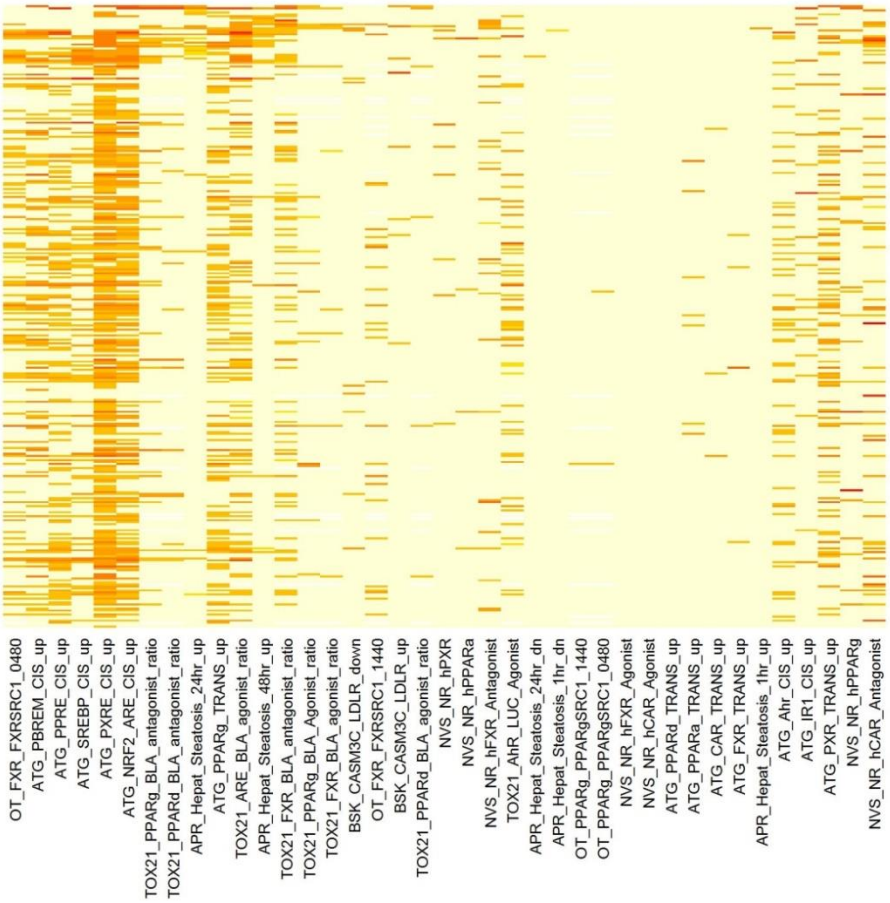

Figure 6: Heat map of -logAC50s on all endpoints for the 29 steatotic compounds (at the top) and the rest of the 310 compounds tested for steatosis. Red indicates a low AC50, pale yellow indicates no response (AC50=0). White indicates missing data. The columns are in the same order as the previous figure.

As a preliminary step towards using the ToxCast data to feed a quantitative steatosis AOP, the steatosis responses were plotted as a function of the nuclear receptor responses. This was performed using the individual dose-response models that were collected from the ToxCast database and that were not affected by cytotoxicity. Figure 7 shows the results for steatosis at 48 hours. This figure does not highlight any clear pattern in the relationship between steatosis and nuclear receptor responses.

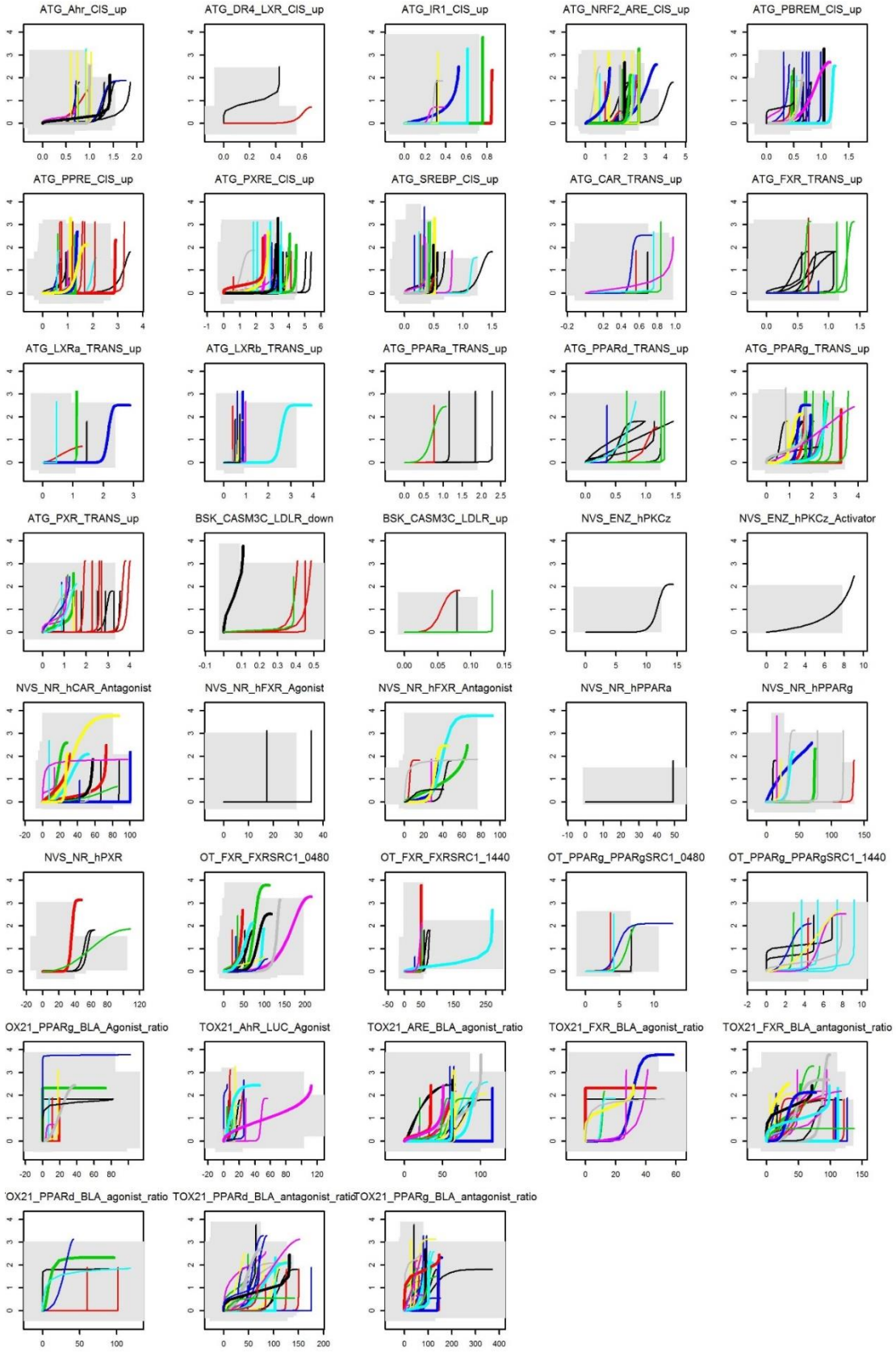

Figure 7: Steatosis response (APR_HepG2_Steatosis_48hr_up) as a function of each nuclear receptor response. In each graph, the colour indicates the compound (only 8 colours, not consistent over all nuclear receptors). When both steatosis response and nuclear receptor response are hit calls, the curve is in bold. The grey area represents the range of responses observed in the data (union over all compounds in each graph).

4.2 Mapping Toxicogenomics on the AOP steatosis

In a second phase, we were interested in using toxicogenomics data to understand the underlying mechanisms of action associated with steatosis.

To start, we collected compounds suspected of inducing steatosis. The selection was done by journal text mining, looking to the terms “steatosis”, “fatty change”, “lipid accumulation” and “lipidosis”.

Then, we selected compounds for which microarray data could be recovered in TG-GATEs, one of the most extensive sources of public toxicogenomics data for cross-species analysis established by the Japanese Toxicogenomics Project [9]. In total, 31 compounds were considered in this study (Table 4).

Acetamide	Cyclosporine A	Methyltestoterone
Acetaminophen	Diltiazem	Phenylbutazone
Allyl Alcohol	Disulfiram	Puromycin aminonucleoside
Amiodarone	Ethanol	Rifampicin
Amitryptyline	Ethinylestradiol	Simvastatin
Aspirin	Ethionamine	Terbinafine
Carbon tetrachloride	Hydroxyzine	Tetracycline
Clomipramine	Imipramine	Valproic acid
Clozapine	Lomustine	Vitamin A
Colchicine	Metformin	
Coumarin	Methapyrilene	

Table 4: List of compounds associated to steatosis and with gene expression data in TG-GATEs

For the analysis, the set of 31 compounds was analysed in different conditions (rat *in vivo*, rat *in vitro*, human *in vitro*, low/middle/high doses, 2/8/24 hours). In order to uncover the common toxicity producing mechanisms of very different compounds, the normalisation was carried out altogether using Bioconductor, Robust-multi-array Average (RMA) and the limma package implemented in R project [10]. Instead of an individual gene/molecule-based analysis, we applied a gene set pathway enrichment analysis, using a method called GAGE (Generally-Applicable Gene set enrichment) [11]. Such a study allows us to observe which pathways are mostly affected and by which compounds from the gene expression profile. Here, the GAGE method was combined to the KEGG database, a biological pathway database [ref]. In KEGG, there is no specific pathway developed for steatosis. However, a non-alcoholic fatty liver disease (NAFLD) pathway is proposed. Although, the NAFLD pathway represents general mechanisms (ER stress, cell death pathway, survival pathway...), steatosis has been reported to be the hallmark of NAFLD and deregulation of genes in this pathway might be related to steatosis [12].

Using GAGE, all the compounds were mapped to the non-alcoholic fatty liver disease (NAFLD) pathway from KEGG (Figure 8). For each gene, a green bar corresponds to a compound that down-regulates the gene and a red bar corresponds to a compound that up-regulates the gene. A grey bar means no change on the gene expression. So, we can see that some compounds – tested in primary human hepatocytes – have an up-regulated effect on PPARα gene or a down-regulated effect on LXR, confirming their impact in steatosis through these MIEs. But some compounds have an inverse effect on these genes and have a steatotic effect through others MIEs. We can notice for example the genes PI3K, L-PK, AMPK, XBP1, AP-1, IL-1. Interestingly, the majority of the compounds deregulate the β-oxidation in mitochondria, which is strongly associated to steatosis.

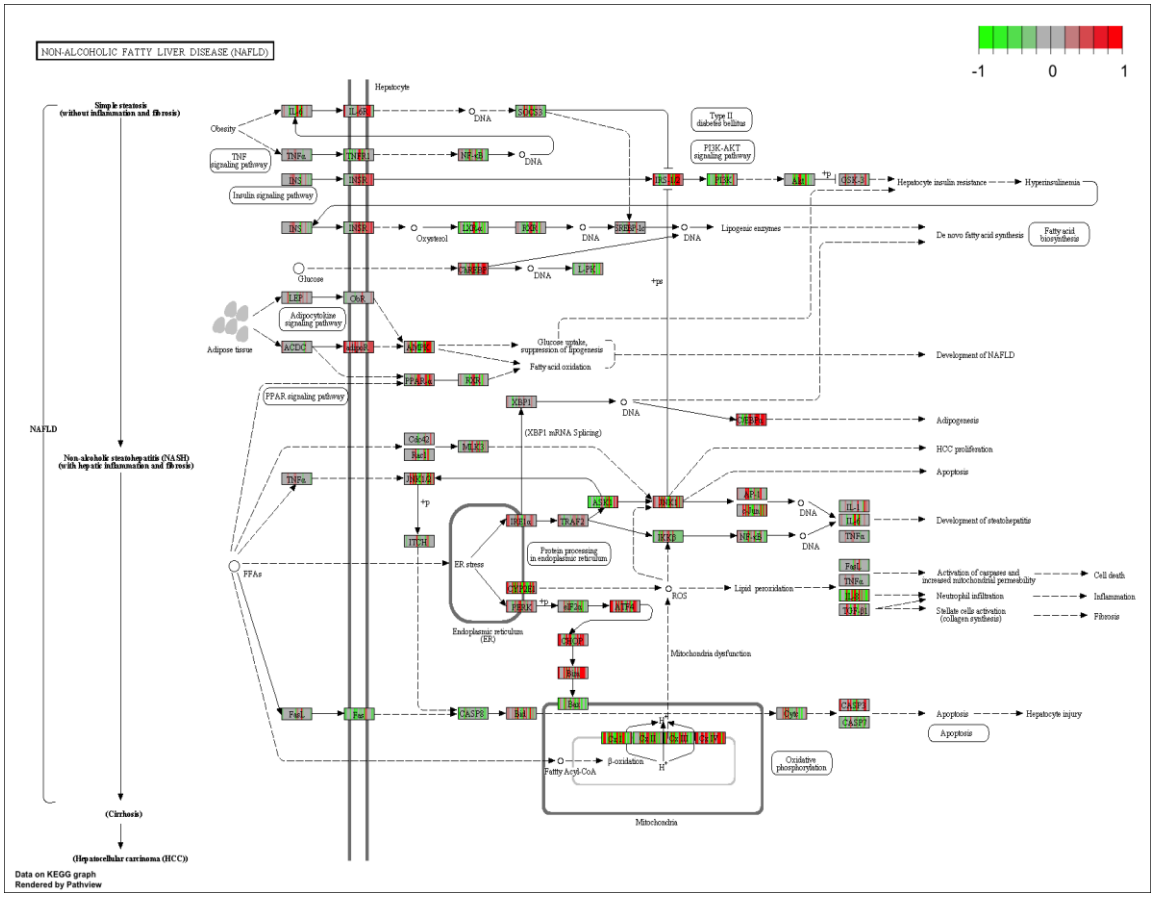

Figure 8: Mapping of the gene deregulation in primary human hepatocytes study by the 31 compounds to the NAFLD pathway. A green bar corresponds to a compound that down-regulates the gene and a red bar corresponds to a compound that up-regulates the gene. A grey bar means no change on the gene expression.

We went a step further, performing a time series analysis using the R package maSigPro [13]. This method consists of a statistical process able to identify genes that have different expression profiles over time. Such analysis allowed identifying 48 specific genes for which their expression changed over time following the same pattern in many compounds (Table 5).

ID	Gene Symbol	Gene Name
91663	MYADM	myeloid associated differentiation marker [Source:HGNC Symbol;Acc:HGNC:7544]
27286	SRPX2	sushi repeat containing protein, X-linked 2 [Source:HGNC Symbol;Acc:HGNC:30668]
10491	CRTAP	cartilage associated protein [Source:HGNC Symbol;Acc:HGNC:2379]
84803	GPAT3	glycerol-3-phosphate acyltransferase 3 [Source:HGNC Symbol;Acc:HGNC:28157]
9787	DLGAP5	DLG associated protein 5 [Source:HGNC Symbol;Acc:HGNC:16864]
83875	BCO2	beta-carotene oxygenase 2 [Source:HGNC Symbol;Acc:HGNC:18503]
3161	HMMR	hyaluronan mediated motility receptor [Source:HGNC Symbol;Acc:HGNC:5012]
6790	AURKA	aurora kinase A [Source:HGNC Symbol;Acc:HGNC:11393]
2633	IDH3G	isocitrate dehydrogenase 3 (NAD(+)) gamma [Source:HGNC Symbol;Acc:HGNC:5386]
28998	MRPL13	mitochondrial ribosomal protein L13 [Source:HGNC Symbol;Acc:HGNC:14278]

	BP1	guanylate binding protein 1 [Source:HGNC Symbol;Acc:HGNC:4182]
54492	NEURL1B	neuralized E3 ubiquitin protein ligase 1B [Source:HGNC Symbol;Acc:HGNC:35422]
10112	KIF20A	kinesin family member 20A [Source:HGNC Symbol;Acc:HGNC:9787]
80114	BICC1	BicC family RNA binding protein 1 [Source:HGNC Symbol;Acc:HGNC:19351]
65084	TMEM135	transmembrane protein 135 [Source:HGNC Symbol;Acc:HGNC:26167]
7105	TSPAN6	tetraspanin 6 [Source:HGNC Symbol;Acc:HGNC:11858]
9615	GDA	guanine deaminase [Source:HGNC Symbol;Acc:HGNC:4212]
114883	OSBPL9	oxysterol binding protein like 9 [Source:HGNC Symbol;Acc:HGNC:16386]
9488	PIGB	phosphatidylinositol glycan anchor biosynthesis class B [Source:HGNC Symbol;Acc:HGNC:8959]
55771	PRR11	proline rich 11 [Source:HGNC Symbol;Acc:HGNC:25619]
11167	FSTL1	follistatin like 1 [Source:HGNC Symbol;Acc:HGNC:3972]
2519	FUCA2	fucosidase, alpha-L- 2, plasma [Source:HGNC Symbol;Acc:HGNC:4008]
9588	PRDX6	peroxiredoxin 6 [Source:HGNC Symbol;Acc:HGNC:16753]
79594	MUL1	mitochondrial E3 ubiquitin protein ligase 1 [Source:HGNC Symbol;Acc:HGNC:25762]
51292	GMPR2	guanosine monophosphate reductase 2 [Source:HGNC Symbol;Acc:HGNC:4377]
81610	FAM83D	family with sequence similarity 83 member D [Source:HGNC Symbol;Acc:HGNC:16122]
55872	PBK	PDZ binding kinase [Source:HGNC Symbol;Acc:HGNC:18282]
59	ACTA2	actin, alpha 2, smooth muscle, aorta [Source:HGNC Symbol;Acc:HGNC:130]
7802	DNALI1	dynein axonemal light intermediate chain 1 [Source:HGNC Symbol;Acc:HGNC:14353]
5445	PON2	paraoxonase 2 [Source:HGNC Symbol;Acc:HGNC:9205]
3242	HPD	4-hydroxyphenylpyruvate dioxygenase [Source:HGNC Symbol;Acc:HGNC:5147]
28998	MRPL13	mitochondrial ribosomal protein L13 [Source:HGNC Symbol;Acc:HGNC:14278]
11004	KIF2C	kinesin family member 2C [Source:HGNC Symbol;Acc:HGNC:6393]
1606	DGKA	diacylglycerol kinase alpha [Source:HGNC Symbol;Acc:HGNC:2849]
10158	PDZK1IP1	PDZK1 interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:16887]
9122	SLC16A4	solute carrier family 16 member 4 [Source:HGNC Symbol;Acc:HGNC:10925]
23082	PPRC1	peroxisome proliferator-activated receptor gamma, coactivator-related 1 [Source:HGNC Symbol;Acc:HGNC:30025]
123264	SLC51B	solute carrier family 51 beta subunit [Source:HGNC Symbol;Acc:HGNC:29956]
6372	CXCL6	C-X-C motif chemokine ligand 6 [Source:HGNC Symbol;Acc:HGNC:10643]
79053	ALG8	ALG8, alpha-1,3-glucosyltransferase [Source:HGNC Symbol;Acc:HGNC:23161]
9928	KIF14	kinesin family member 14 [Source:HGNC Symbol;Acc:HGNC:19181]
788	SLC25A20	solute carrier family 25 member 20 [Source:HGNC Symbol;Acc:HGNC:1421]
114883	OSBPL9	oxysterol binding protein like 9 [Source:HGNC Symbol;Acc:HGNC:16386]
55526	DHTKD1	dehydrogenase E1 and transketolase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:23537]
56922	MCCC1	methylcrotonoyl-CoA carboxylase 1 [Source:HGNC Symbol;Acc:HGNC:6936]
84803	GPAT3	glycerol-3-phosphate acyltransferase 3 [Source:HGNC Symbol;Acc:HGNC:28157]
9488	PIGB	phosphatidylinositol glycan anchor biosynthesis class B [Source:HGNC Symbol;Acc:HGNC:8959]
516		ATP synthase
9588	PRDX6	peroxiredoxin 6 [Source:HGNC Symbol;Acc:HGNC:16753]

Table 5: List of genes that are highly deregulated after exposure of the 31 chemicals at time 2h, 8h and 24h.

Among the list of genes, we notice that MYADM has been previously described as hepatocellular carcinoma sign [14], SLC51B maintains the enterohepatic cycling and distribution that could affect glucose lipid homeostasis [15], PRDX6, an antioxidant enzyme involved in the reduction of

phospholipids hydroperoxides, has been previously proven related to NAFLD in rats administered ethanol, GPAT3 (glycerol-3-phosphate acyltransferase 3) has been suggested to have a role in lipid absorption and it has also been parallelised with PPARα pathway, showing de-regulation of fatty acid b-oxidation at the same time than GPAT3 [16].

The variability of a gene can be visualised for all compounds at any time and any dose. For example, the ATP synthase in the mitochondria F0 complex shows an increase of the expression along the time for the majority of the compounds, except for the compound allyl alcohol (AAA). Interestingly, if we see a time-gene expression dependency, the relationship dose-gene expression is less obvious. For some compounds (CPM, DSF, ETH, HYZ, IMI, TBF, WY), there is no variability of the gene expression dependent to the dose (Figure 9).

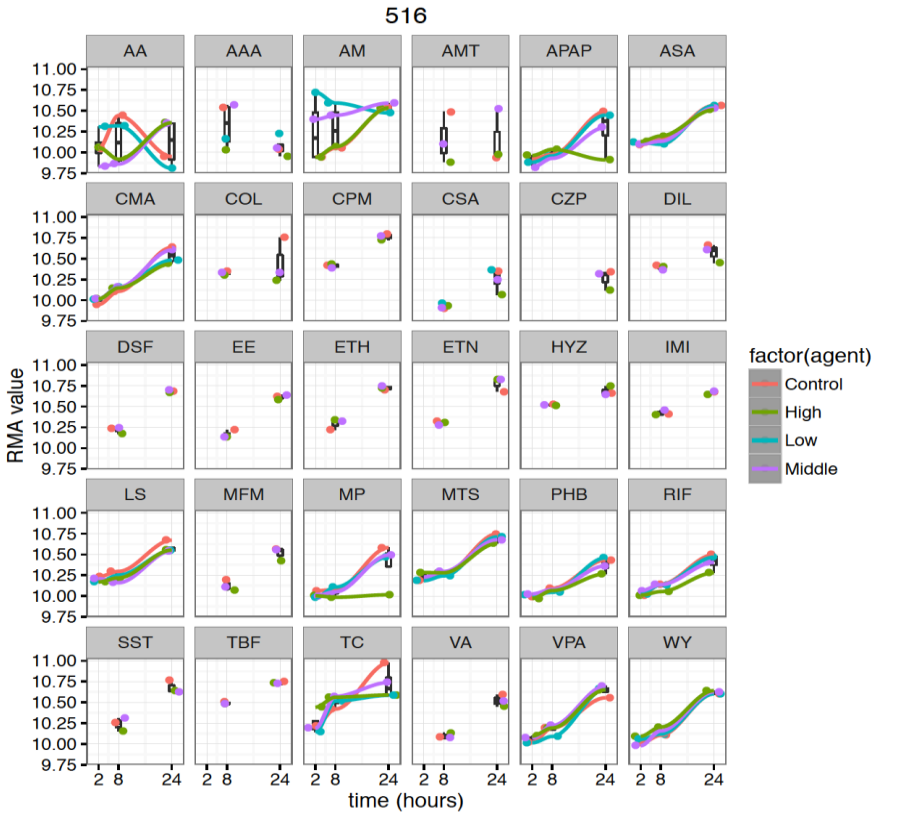

Figure 9: Time series analysis for 30 compounds studied in human *in vitro* hepatocytes for the gene 516 (ATP synthase in the mitochondria F0 complex). Each colour corresponds to the chemical exposure concentration. The X-axis is the time and the Y-axis the gene expression.

Overall, these gene expression studies allow us to validate some MIE described on the AOP steatosis but they can also suggest potentially new MIEs to investigate for this AOP. Discussions with partners from the case study on steatosis within the EU-ToxRisk project are ongoing to test some of the compounds and hypotheses in order to confirm these new findings.

It is planned to continue this work with the analysis of the “repeat dose” studies that have been performed in TG-GATEs and not considered yet in the analysis. Protein-protein interaction network analysis will be also integrated in the analysis in a similar way to the gene set pathway enrichment analysis in order to enrich the AOP on steatosis.

5 Conclusion

In this deliverable, ToxCast data has been mapped to the AOP concept developed by the OECD and available on the AOP wiki. So far, 5,290 chemicals have been mapped to 82 AOPs through 48 assays performed in ToxCast. For some MIEs, such as ADRB2, IYD, HPDD, GRIK1, GRIA1 no assays have been developed in ToxCast that could guide the development of new assays. Furthermore, for some assays, no chemicals have been measured active, so new chemicals could be tested for these assays. This is the case for HMGCR, EcR, Gox, DIO1, DIO2, DIO3, ABCB11, DBH, DUOX1, DUOX2, SLC5A5. Finally, for some assays (AhR, AR, ESR1, CYP19A1) a large set of chemicals were measured active and could be potentially involved as "stressors" of an MIE in an AOP. Further tests could be of interest for these sets of compounds. In addition, the ToxCast-AOP mapping study presented here could support modellers and computational toxicologists for the development of computational approaches and improved toxicity prediction models for MIEs on which a large set of compounds have been tested. It could be envisioned for example to develop QSARs models for some specific AOPs.

Taking advantage of the fact that hepatic steatosis was assessed for a large set of chemicals on rat primary hepatocytes in ToxCast, a dose-response analysis has been performed on the steatosis AOP. The steatosis responses were plotted as a function of the nuclear receptor responses for a set of 29 chemicals. NRF2, PXR, PPARγ and AhR are the most frequently targeted proteins confirming the impact of these proteins on steatosis. However individual dose-response models do not highlight any clear dose-response pattern in the relationship between steatosis and nuclear receptor responses. During the analysis, we have noticed that the cytotoxicity of steatotic compounds could interfere with the steatosis effect in ToxCast and should be considered carefully on the step of the analysis (quantitative steatosis AOP).

Finally, a toxicogenomics data analysis was realised on a set of 31 compounds associated with steatosis. Instead of an individual gene/molecule-based analysis, we applied a gene set pathway enrichment analysis and a time series analysis. Performing the analysis with a set of compounds (instead of one compound at the time) has emphasized genes involved in steatosis. This analysis allowed to validate genes suspected to act as MIEs on steatosis (PPARα, LXR, etc.) but also to identify potential new genes as MIEs for steatosis. It will be of interest to further discuss with partners of the consortium to test some of the compounds on known proteins related to steatosis or to validate genes as new MIE for steatosis.

As a perspective, a similar approach to the one developed for steatosis could be performed on other AOPs. Still, we will need to have transcriptomic data or/and phenotypic screens for a set of compounds and tested for a specific phenotype or AOP. So far, only steatosis and apoptosis have been tested in ToxCast. Similarly, gene expression profiling is essentially available to assess liver toxicity.

Overall, such information will be updated according to the modification and the release of new AOPs in the AOP wiki and integration of new compounds known to be active on these assays. *In vitro* data obtained in the EU-ToxRisk project will be also integrated when available.

6 Bibliography

[1] Ankley GT; Bencic D; Breen M *et al*. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem*. 2010, 29, 730-741.

[2] Oki NO; Edwards SW. An integrative data mining approach to identifying adverse outcome pathway signatures. *Toxicology*. 2016, 350-352, 49-61.

[3] Richard AM; Judson RS; Houck KA *et al*. ToxCast chemical landscape: paving the road to the 21st century toxicology. *Chem Res Toxicol*. 2016, 29, 1225-1251.

[4] Villeuneuve DL; Crump D; Garcia-Reyero N; Hecker M *et al*. Adverse Outcome Pathway (AOP) development I: Strategies and principles. *Toxicol. Sciences*, 2014, 142, 312-320

[5] Angrish MM; Kaiser JP; McQueen C; Chorley B. Tipping the balance: Hepatotoxicity and the 4 apical key events of hepatic steatosis. *Toxicol. Sci*. 2016, 150, 261-268.

[6] Filer DL; Kothiya P; Setzer WR; Judson RS; Martin MT (2015). The ToxCast™ Analysis Pipeline: An R Package for Processing and Modeling Chemical Screening Data. https://www.epa.gov/sites/production/files/2015-08/documents/pipeline_overview.pdf

[7] Karmaus AL; Filer DL; Martin MT; Houck KA (2016). Evaluation of food-relevant chemicals in the ToxCast high-throughput screening program. *Food and Chemical Toxicology* 92, 188-196.

[8] Judson R; Houck K; Martin M *et al*. Editor's Highlight: Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol. Sci*. 2016 152, 323-339.

[9] Uehara T; Ono A; Maruyama T et al. The Japanese toxicogenomics project: application of toxicogenomics *Mol. Nutr. Food Res.* 2010, 54, 218-227.

[10] Ritchie ME; Phipson B; Wu D *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015, 43, e47.

[11] Luo W; Friedman MS; Shedden K *et al.* GAGE: Generally applicable gene set enrichment for pathway analysis*. BMC Bioinformatics* 2009, 10, 161.

[12] Takahashi Y. and Fukusato T. Histopathology of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World J. Gastroenterol*. 2014, 20, 15539-15548.

[13] Conesa A; Nueda MJ; Ferrer A; Talon M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformartics* 2006, 22, 1096-102.

[14] Megger, D.A., et al., Proteome Analyses of Hepatocellular Carcinoma*. J Clin Transl Hepatol*, 2014. 2, 23-30.

[15] Arab, J.P., et al., Bile acids and nonalcoholic fatty liver disease: Molecular insights and therapeutic perspectives*. Hepatology*, 2017. 65, 350-362.

[16] Newton, B.W., et al., Liver proteome analysis in a rodent model of alcoholic steatosis*.* J *Proteome Res*, 2009. 8, 1663-71.

EU-ToxRisk

An Integrated European 'Flagship' Program
Driving Mechanism-based Toxicity Testing and Risk Assessment
for the 21st Century

Deliverable 10.1 (Annexes):

Mapping ToxCast data to AOPs

Grant Agreement number: 681002

H2020-PHC-33-2015-single-stage

Start date of project: 1 January 2016

Duration: 72 months

Lead beneficiary of this deliverable: UCPH, INERIS

Deliverable type: Report

Dissemination Level: PU

Release: FV

Due date of deliverable: 31/12/2016

Actual submission date: 17/08/2017

Description: Report on the mapping of available ToxCast data to the EU-ToxRisk21 AOPs developed in WP5

Table of Contents

AOP_toxcast_match_Cpd.xlsx file 29

InteractiveHeatmap-UpDownregulated.html 30

InteractiveHeatmap.html file link 31

AOP_toxcast_match_Cpd.xlsx file

The "AOP_toxcast_match_Cpd.xlsx" file contains the number of compounds bioactive for an assay related to an AOP:

AOP_toxcast_match
_Cpd.xlsx

InteractiveHeatmap-UpDownregulated.html

The InteractiveHeatmap-UpDownregulated.html link to a heat map describing all the compounds from ToxCast active to at least 1 of the 48 assays associated to an AOP:

InteractiveHeatmap_
UpDownregulated.htr

InteractiveHeatmap.html file link

The InteractiveHeatmap.html file link to a heatmap showing the Activity Concentration value at 50% (AC50) for all the compounds from ToxCast active to at least 1 of the 48 assays associated to an AOP:

InteractiveHeatmap.html