The following handle holds various files of this Leiden University dissertation:
http://hdl.handle.net/1887/59476

Author: Huppelschoten, S.
Title: Dynamics of TNFalpha signaling and drug-related toxicity
Issue Date: 2017-12-06
Chapter 6

High content live cell imaging of the dynamics of multiple adaptive stress response pathways enables assessment of drug-induced liver injury liabilities

**Highlights**

- DILI compound exposure induces specific adaptive stress reporter activation with distinct time dynamics and magnitude
- Quantitative data analysis and hierarchical clustering reveals both mechanistic and predictive insight into mechanisms of toxicity
- Feature selection and SVM classification show predictive power of a diverse and complex set of biologically interpretable features

This chapter is based on a manuscript in preparation:

Steven Wink\(^1\)\(^2\), Steven Hiemstra\(^1\)\(^2\), Suzanna Huppelschoten\(^1\)\(^2\), Janna Klip\(^1\), Bob van de Water\(^1\)

**High content live cell imaging of the dynamics of multiple adaptive stress response pathways enables assessment of drug-induced liver injury (DILI) liabilities**

\(^1\)Division of Toxicology, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

\(^2\)These authors contributed equally
Drug-induced liver injury remains a major concern during drug development. There is an urgent need for improved mechanistic understanding and prediction of DILI liabilities from \textit{in vitro} approaches. Previously, we have reported a high content live cell microscopy platform containing mechanism-based fluorescent protein toxicity pathway reporters to detect cellular stress responses reflecting primary human hepatocyte drug responses. Here, we have systematically evaluated the application of integrated stress pathway target gene activation of four key adaptive stress pathways: oxidative stress, ER stress, DNA damage stress and inflammatory stress. More than 118 FDA-labeled drugs in 5 concentrations based on human plasma maximum concentration levels were screened for reporter activation, using live cell confocal imaging. Quantitative data analysis revealed activation of single or multiple reporters by most drugs in a concentration and time dependent manner. Hierarchical clustering of time course dynamics and refined single cell analysis allowed the allusion of key events in toxicity. Concentration response modeling was performed to calculate benchmark concentrations (BMCs). Extracted temporal dynamic parameters and BMCs were used to assess the predictive power of sublethal adaptive stress pathway activation. Although cellular adaptive responses were activated by non-DILI and severe-DILI compounds alike, dynamic behavior and lower BMCs of pathway activation were sufficiently distinct between these compound classes as shown by unsupervised hierarchical clustering and their combined high correct classification. The high level detailed temporal and concentration-dependent dynamics of adaptive stress pathway activation by exposure with compounds with DILI liabilities add to the overall understanding and prediction of drug-induced liver toxicities

\textbf{Keywords:} Drug-induced liver injury, live cell imaging, adaptive stress response pathways, machine learning

\textbf{Introduction}

Despite major efforts to understand and predict drug-induced liver injury (DILI), unpredicted liver failure upon drug use remains an important problem both in the clinic and during drug development\textsuperscript{313}. In general, various chemical, genetic, and lifestyle factors can contribute to the development of DILI. To be able to improve prediction of DILI liabilities of new molecular entities it is essential to understand how currently known compounds induce injury. Of most interest are the idiosyncratic DILI compounds, inducing unpredicted results at a normal dose, and only in a very small group of patients.

Gene expression analysis has contributed significantly to our understanding of DILI\textsuperscript{314,315}. This has led to the identification of specific signaling pathways that are activated during DILI and are possibly predictive for chemical-induced liver injury. Key among these are classic stress responses activated to maintain cellular homeostasis, including the oxidative stress response, the ER stress response and the DNA damage response\textsuperscript{314}, together with inflammatory stress\textsuperscript{208}. We have established fluorescent protein reporter
 Dynamic stress response pathway evaluation for drug safety assessment

hepatoma cell lines using bacterial artificial chromosome (BAC) GFP tagging, that capture each of these four pathways using SRXN1, CHOP, p21 and ICAM1 as biomarkers. For the oxidative stress pathway we have established a SRXN1-GFP reporter. Parent compounds or their metabolites can oxidize, reduce or alkylate cellular components, thereby directly inducing oxidative stress or indirectly for example via disruption of essential mitochondrial functioning. The activation of transcription factor Nrf2 together with AP1 is dependent on the redox sensor Kelch-like ECH-associated protein 1 (Keap1), and induces expression of antioxidants among which the small redox protein (SRXN1). SRXN1, conserved in all eukaryotes, is best characterized for its ATP-dependent reduction of the hyperoxidized form of peroxiredoxin. Peroxiredoxin (Prx) is an enigmatic protein, its reversible inactivation caused by its own substrate. However, it is hypothesized that SRXN1 activity is the essential switch between Prx either acting as chaperone of the intracellular messenger peroxides and its usual peroxidase function, protecting the cell against oxidative damage. Furthermore, deglutathionylation of s-glutathionylated cysteins by SRXN1 has been shown to be essential for the functioning of phosphatases. In vivo studies show that in the liver, upregulation of SRXN1 via Nrf2 activation, upon either cadmium-induced heavy metal stress and Pyrazole exposure or alcohol-induced CYP2E1 related toxicity, is vital for protection against fulminant oxidative stress and subsequent organ failure.

To monitor ER stress we have established a CHOP-GFP reporter. ER stress is a protective response upon the accumulation of untranslated proteins in the endoplasmic reticulum (ER). This can be induced by compound or metabolite-induced disruption of for example N-linked glycosylation, disulfide bond formation, transport of folded proteins from the ER, protein folding, calcium homeostasis, or ER stress response components itself. The outcome of the compound-induced ER stress response, cell survival or cell death, is dependent both on the strength and duration of the response. Upon ER stress, three classical signaling pathways are activated via the sensor proteins PKR-like ER kinase (PERK), activating transcription factor 6 (Atf6), and inositol-requiring enzyme 1 (Ire1). Activation of PERK leads to arrested protein translation by phosphorylation of the eukaryotic transcription factor 2, in order to diminish the flux of unfolded proteins into the ER. Permitted translation of activating transcription factor 4 (Atf4) and results in the specific expression of ER function-related proteins and the transcription factor C/EBP homologous protein (CHOP). Activation of the Ire1 and Atf6 pathways can also induce CHOP expression, although with lower efficiency. CHOP activity induces or represses the expression and function of many proteins, including proapoptotic proteins like Bcl-X, BAX and caspase3, and the suppression of antiapoptotic proteins as Bcl2. In the liver, CHOP expression is not conclusively linked to protection or increased drug-induced injury. In vitro as well as in vivo studies have detected CHOP induction upon drug exposure, however, at this moment a causal relationship with cell death or protection against cell death cannot be drawn.

DNA damage can be monitored by a fluorescent protein reporter for the p53 downstream target gene p21, as shown previously. The cellular protective response upon DNA damage induces cell cycle arrest and subsequent senescence, to prevent genomic
instability during replication. DNA damage can be induced either directly by chemical mutagens, usually electrophiles that directly interact with DNA and form covalent bonds, or indirectly by the formation of ROS that can interact with DNA\textsuperscript{135}. The key transcription factor p53 is activated upon DNA damage and induces expression of its many target genes, among which is p21\textsuperscript{325}. The best-characterized function of p21 is its effective inhibition of cyclin-dependent kinases (CDK), which halts the progression of the cell cycle in the G1 or G2/M phase. Localization of p21 has been found both in the nucleus and the cytoplasm, reflecting different roles in both cell cycle arrest and antiapoptosis signaling respectively\textsuperscript{326}. In the liver, \textit{in vivo} studies show upregulated p21 nuclear expression upon drug exposure, mostly via p53 activation\textsuperscript{327,328}.

Finally, a fluorescent reporter for ICAM1 allows the monitoring of the cytokine-mediated activation of NF-κB signaling\textsuperscript{28}. Besides intrinsic drug qualities as dose and lipophilicity, several host factors have been described to be involved in the development of idiosyncratic DILI. Among these, inflammatory processes have been known to be detrimental in the process of liver injury, but are also known to be essential for recovery after injury\textsuperscript{208}. Inflammatory cytokine excretion firstly activates and recruits innate as well as adaptive immune cells, and secondly activates the hepatocytes and non-parenchymal cells themselves. In \textit{in vivo} studies, the production of proinflammatory cytokine TNFα upon lipopolysaccharide exposure in combination with drugs is strongly suggested to increase toxicity\textsuperscript{10}. TNFα stimulation of endothelial cells as hepatocytes has been known to induce both adaptive, pro-survival, and adverse, pro-death, cellular signaling. The activation of the key transcription factor nuclear factor κB (NF-κB) ensures the expression of several pro-survival genes, but also enhances the proinflammatory gene expression\textsuperscript{20}. One of the proinflammatory genes induced by NF-κB activation is intercellular adhesion molecule 1 (ICAM1). ICAM1 is expressed at the membrane of TNFα-activated hepatocytes, aiding in the adherence and transendothelial migration of leukocytes from the bloodstream\textsuperscript{329}. ICAM1 is widely used as a marker for inflammation and ICAM1 expression is also increased upon inflammation in the liver\textsuperscript{330}.

Given the central role of the above pathways in liver injury and DILI specifically, our objective was to evaluate the application of our panel of target gene BAC-GFP reporter cell lines that represent these four major adaptive stress response pathways to predict DILI liability. Previously, we demonstrated that the BAC-GFP reporters allow the quantification of the chemical-induced stress responses similar to primary human hepatocytes\textsuperscript{44}. Here we systematically determined the application of the BAC GFP-SRXN1, GFP-CHOP, GFP-p21 and GFP-ICAM1 reporters for the assessment of DILI using a set of 118 FDA-labeled drugs with DILI drug label classification. The concentration- and time-dependent GFP responses were determined in association with several cytotoxicity parameters. In this study, we provide quantitative information of the dynamic adaptive stress response activation for all 118 drugs allowing detailed mode-of-action assessment. Mechanism-based temporal dynamic data together with concentration response modeling was used for prediction of DILI outcome.
Materials and methods

Cell culture
Human hepatoma HepG2 cells were acquired from ATCC (clone HB8065). HepG2 SRXN1, DDIT3 (CHOP), CDKN1A (p21) and ICAM1 BAC GFP reporter cell lines were generated and characterized as described previously. HepG2 BAC-GFP reporters were maintained and exposed to drugs in DMEM high glucose supplemented with 10% (v/v) FBS, 25 U/mL penicillin and 25 µg/mL streptomycin. The cell lines were used between passage 5 and 25. For live cell imaging, the cells were seeded in Greiner black µ-clear 384 wells plates, at 8,000 cells per well.

Reagents
All reference compound chemicals were acquired from Sigma-Aldrich and freshly dissolved in DMSO; except for metformin, fluphenazine, buthionine sulfoxamine, bromoethylamine (all PBS), acetaminophen and phenobarbital (all DMEM). TNFα was acquired from R&D Systems (Abingdon, UK). DILI compounds were a kind gift from the Dr. Weida Tong, NCTR-FDA. All compounds were maintained as 500-fold stock such that final treatments did not exceed 0.2 % v/v DMSO.

Microscopy
Accumulation of GFP levels, propidium iodide (PI) and Hoechst staining was monitored using a Nikon TiE2000 confocal laser microscope (lasers: 540nm, 488nm and 408nm), equipped with an automated stage and perfect focus system at 37 degrees Celsius with humidified atmosphere and 5% CO₂/air mixture. Prior to imaging at 20x magnification, HepG2 cells were loaded for 45 minutes with 100 ng/mL Hoechst33342 to visualize the nuclei, upon which the Hoechst-containing medium was washed away to avoid Hoechst phototoxicity and replaced with medium containing PI to monitor cell death. Each 384-well plate contained one reporter cell line, which was exposed to all the compounds used in the screen at one certain concentration (1, 5, 10, 50 or 100 C-max); for each concentration at least two replicates were imaged per reporter cell line. For the ICAM1-GFP reporter experiments, cells were first exposed for 8 hours to compound only. Next, TNFα was added to all wells, up to a final concentration of 10 ng/mL, which diluted the compound concentration ten times. Upon TNFα exposure, imaging was started.

Reporter response quantification
Quantitative image analysis was performed with CellProfiler version 2.1.1 with an in house developed CellProfiler module implementing the watershed masked algorithm for segmentation. Image analysis results were stored as HDF5 files. Data analysis, quality control and graphics was performed using the in house developed R package h5CellProfiler (manuscript in preparation). For each reporter hourly intensity levels of the GFP signal, the nuclear Hoechst33342 intensity levels and at 24 hours the PI staining were measured at the single cell level. In addition cell numbers, nuclei size and cell speed were measured.
**Data analysis**

GFP intensity cell population means were calculated. In addition, for each plate the cell population mean GFP intensity of the DMSO treated cells was calculated to determine background control values. Per plate, the single cells that had values above the 2X mean, 3X mean were counted. For ICAM1, the background control values consisted of DMSO conditions treated with TNFα, and the single cells with values above, as well as below background values were counted. Due to the non-symmetric distribution of ICAM1 cell population GFP intensities, the interquartile range (IQR) was used to count the number of cells 1.5X, 2X and 3X above and below the TNFα IQR control values (Suppl. Fig. 1). To account for PI background staining noise the PI segmentations were masked by a 2 pixel dilated nuclei. The area of these nuclei and the PI objects were divided to obtain a PI/nuclei ratio. These ratios were filtered to be at least 10 % of the cell size and following this procedure each cell was either flagged as alive or dead in the final time point of the 24 live imaging session. PI positive fraction were normalized to DMSO (or TNFα for ICAM1) by subtracting the control PI positive fractions. Linear regression was applied with time as independent variable to quantify treatment effects on a plate to plate basis of cell speed, nuclear size, Hoechst33342 nuclear intensity cell numbers. The slope coefficient mean over all plates was used to obtain a compound-concentration specific summary feature. All summary features were scaled between 0 and 1 with the formula \((x-x_{\text{min replicate}})/(x_{\text{max replicate}} - x_{\text{min replicate}})\), with the exception of 1) the cell count features which were scaled between 0 and 1 by calculating cell fractions and 2) the ICAM1-GFP intensity features which were scaled between -1 and 1 to account for up or down regulation of the TNFα-induced ICAM1 expression (Suppl. Fig. 1).

**Concentration response data transformation and Benchmark Concentration (BMC) modeling**

The maximum values over time of the scaled intensity levels and positive GFP fractions were selected for the concentration response curves. These values were fit to a 4 parameter log-logistic model using the drc package\(^{333}\). BMC values were calculated as the concentration at which +0.25 (and – 0.25 for ICAM1) absolute increase from the initial response values occurred (Fig. 4A). The replicate means of the maximum over-time features were calculated for each compound-concentration preceding unsupervised hierarchical clustering. For the time courses (Fig. 3) natural cubic splines with 8 degrees of freedom were fit after which 24 discrete equidistant time points were selected to calculate per-time point replicate means. The time course hierarchical clustering was performed by first calculating Manhattan-based distances between all time course vectors. The mean Manhattan-based distances over all reporters were used as inputs for the Ward-based clustering. This ensured the temporal dynamics were also clustered appropriately.

**Data representation**

All HCI data representations were generated or modified with Illustrator CS6, Fiji, ggplot2\(^{219}\), the `aheatmap` function of the NMF package\(^{334}\).
Severe vs non-Severe DILI prediction with Support vector machine
FDA DILI-annotation was used as ‘ground truth’ with non-DILI (n=16), less-severe DILI (n=36) and ambiguous DILI (n=12) grouped as ‘nonSevere DILI’ and severe-DILI (n=54) as ‘severe DILI’ resulting in a two-classification problem. Features were obtained by time dynamic feature extraction of time courses using functional data analysis using the in house developed R-package ‘celloscillate’ and the BMC and cmax normalized BMC values (Fig. 6). Feature selection and SVM model tuning was performed in an 200 times iterative process with randomly selected 80/20 – equal class distributed training/test set procedure. The training phase included a first feature selection step using a Kolmogorov-Smirnov test for equal distributions between the two classes followed by pair-wise correlation filter step (>0.8 or <-0.8). After the feature selection step the training phase. The second step in the training phase consisted of the SVM model tuning with 10 repeats of 10-fold cross-validation. The test phase on 20% of the compounds was performed using the selected features and tuned SVM model. Reported prediction results are the average of the 200 test-set runs (Fig. 6), the ROC distribution of the test-runs are displayed in Fig. 6A. Hierarchical clustering of the 20 selected features (Fig. 6) correspond to the features selected >150 times through the 200 iterations.

Gene expression analysis
CEL files were downloaded from the Open TG-GATEs database: “Toxicogenomics Project and Toxicogenomics Informatics Project under CC Attribution-Share Alike 2.1 Japan” http://dbarchive.biosciencedbc.jp/en/open-tggates/desc.html and analyzed as described previously

Statistics
For statistical significance of all time courses first linear interpolation was applied for each separate time course using the `approx` function from the R-stats package to obtain 100 equal discretized time points for each replicate. The high number of linear interpolations was required to retain the original noise in the time course data. Following this step, a one-way ANOVA for functional data method was applied using the `anova.onefactor` function of the R-package fda.usc to determine significant difference in time-curves compared to DMSO for SRXN1/CHOP/p21 or TNFα for ICAM1. Multiple testing correction was applied using the fdr-method (Benjamini & Hochberg). SRXN1/CHOP/p21 were assessed for significant upregulation and ICAM1 for significant down- or up-regulation. For the log-BMC values a linear model with the BMC as explanatory and C-max as explanatory variable was fit as null-model. The null-model was compared in an anova to a model containing DILI-class as additional additive explanatory variable. The models were compared in an anova for significant effect of DILI-class. For the C-max normalized BMC a Welch two-sample t-test was performed between the severe and non-severe DILI groups.
Results

*High content adaptive stress response screen with DILI compounds*

To assess the application of adaptive stress response pathway activation for the assessment of DILI liabilities we screened 123 compounds, of which 118 with known DILI liabilities (Fig. 1 and Table 1). As an adaptive stress response read-out, HepG2 BAC-GFP reporter cell lines for oxidative stress (SRXN1-GFP), DNA damage (p21-GFP), ER-stress (CHOP-GFP) and inflammatory cytokine signaling stress (ICAM1-GFP) were used (28 and see Suppl. Fig. 2). Stress response activation following drug exposure was monitored with live cell confocal microscopy for a period of 24 hours. The time-resolved single cell data was quantified using an established image analysis pipeline44. For labeling DILI compounds we used the FDA DILI labeling, which labels drugs either as no-DILI-concern, ambiguous DILI-concern, less-DILI concern or most-DILI-concern335. Most-DILI-concern drugs are highly associated with DILI and represent multiple specialist verified cases of DILI. Less-DILI-concern drugs represent few verified cases of DILI. If drugs are suspected to cause most- or less-DILI-concern, but the presented cases cannot be conclusively validated by experts, drugs received the ambiguous DILI-concern label. No-DILI-concern drugs are on the market for decades and are never associated with DILI. In order to separate out clear examples of DILI, we made two classes: non-severe and severe DILI, where the most-DILI-concern drugs are in the severe DILI group and all others are in the non-severe DILI group. In addition, we included FDA labeling in eight separate classes of hepatotoxicity ranging from no hepatotoxicity to fatal hepatotoxicity (Table 1). The screen also included control reference compounds included negative controls (DMSO and medium) and positive controls (i.e. DNA damage inducers, alkylating agents, ER stress inducers etc.) (Table 1 and Fig. 2).

*Single cell analysis allows fine tuning of sensitivity versus dynamic range of BAC-GFP reporter responses*

All reporters were exposed to five concentrations: 1, 5, 10, 50 and 100 C-max followed by automated live cell imaging and multiparametric image analysis (Fig. 1). C-max values and drug metabolite formation were obtained either from FDA or from literature (see Table 1). For all images single cell analysis was performed to extract a diverse set of quantitative data, including GFP reporter activity, cell number and cytotoxicity (Suppl. Fig. 1). SRXN1-GFP, p21-GFP and CHOP-GFP reporter single cell data was used to derive quantitative data for four different determinants of reporter activity: intensity, fraction of cells with GFP intensity levels above control values. All ICAM1-GFP reporter drug exposures were primed with TNFα exposure; likewise, ICAM1-GFP shows a gradual increase over 24 hour time period in the vehicle control. Therefore drug treatment can lead to an up- or downregulation of the ICAM1-GFP response. Systematic evaluation of these descriptors for the least and strongest responding compound for each individual reporter allowed fine tuning of the sensitivity versus the dynamic range (Fig. 2).
For example, based on the SRXN1-GFP intensity over the single cell population chlorpromazine would not have been defined as positive in the SRXN1-GFP reporter cell line, because only in a small proportion of cells that contain a higher level of SRXN1-GFP the signal was detected. Yet, the GFP_pos.2m and GFP_pos.3m responses of chlorpromazine were more sensitive descriptors that also allowed evaluation of the time course dynamics. Similar observations were made for nitrofurantoin and clozapine for the CHOP-GFP and p21-GFP reporters, respectively. However for strong inducers of oxidative stress (diethylmaleate; DEM), UPR (thapsigargin) and DNA damage (etoposide), GFP mean intensity already allowed detection of the reporter responses, while GFP_pos.2m caused an early saturation, thereby lowering the information value of the temporal dynamics. Further, well-known inflammatory stress modulating compounds diclofenac (DCLF) and azathioprine (AZA) showed inhibiting and enhancing modulatory effects on ICAM1-GFP, respectively.

![Diagram](image_url)

**Fig. 1:** BAC cloning, BAC reporter DILI screen and analysis pipeline Left panel) BAC cloning technology is used to maintain endogenously regulated reporter protein levels and regulation. Monoclonal reporter selection from a high number of clones to ensure endogenous response to positive control stimuli and suitability of reporter for imaging. Middle panel) High content live cell screen of 123 compound at 1, 5, 10, 50 and 100 C-max at 2 or 3 replicates. Right panel) Image and data analysis is performed with CellProfiler/Fiji and R, respectively. Some in-house tools were developed in CellProfiler and R to assist in the quality and analysis of the large data output.
**DILI compounds show specific reporter activation with distinct time dynamics and magnitude**

For evaluation of the reporter activation for the entire compound screen GFP_pos.2m was selected as the most sensitive initial readout. The GFP_pos.2m time courses were used to calculate the mean of the replicates for SRXN1-GFP, CHOP-GFP, p21-GFP and ICAM1-GFP reporter responses for all compounds (Fig. 3A and Suppl. Fig. 3). Some compounds showed a response in all four reporters, where ICAM1-GFP can be increased or decreased due to compound exposure. Methyldopa (MD) for example shows an increase in SRXN1-GFP, CHOP-GFP and p21-GFP and a clear decrease in ICAM1-GFP. Mercaptopenurine (6MP) induced all four reporters. The data also allowed discrimination of specific time dynamics of stress pathway activation. Thus, for nimesulide (NMS), rifampicin (RFP) and oxytetracycline (OXY) an initial CHOP-GFP response at 100x C_max and a delayed SRXN1-GFP response was observed. In contrast, for the azathioprine (AZA), colchicine (CLC) and dacarbazine (DTIC) an SRXN1-GFP response was followed by CHOP-GFP, eluding to the primary mode of action or type of cellular stress. As a next step we performed hierarchical clustering of the time courses from all 118 compounds representing the reporter activities from all BAC-GFP reporter cell lines (Suppl. Fig. 4). We observed a cluster with strong modulation of stress responses containing mostly the higher C_max treatment conditions (Fig. 3B). This highly active cluster showed an overrepresentation of severe DILI drugs as well as more severe classes of hepatotoxicity (liver necrosis, acute liver failure and fatal hepatotoxicity). Clusters with activation of all four reporters were present, with ICAM1-GFP either up- or downregulated. p21-GFP did show few responses and did not contribute much to the DILI compound clustering. The time response clearly demonstrated the dynamics of the various stress response programs and allowed discrimination between primary stress type and subsequent secondary responses. Strikingly, suppression by DILI compounds of the cytokine-induced ICAM1-GFP expression was highly correlated with activation of the CHOP-GFP reporter, which in a few cases was co-occurring with SRXN1-GFP activation. In conclusion, time resolved clustering results in marked clustering of time dynamics and contributes to mode of action identification.
**Fig. 3:** Time dynamics of a subset of the screened drugs

A) GFP responses over time of GFP_pos.2m of SRXN1-GFP, CHOP-GFP and p21-GFP and of GFP_dif.2m of ICAM1-GFP. Statistics are performed as described in the material and methods section and represent *p < 0.01 with the corresponding color to dissect between the different reporter lines. B) Zoom of hierarchical clustering of time dynamic responses. Red is an upregulation and blue is a downregulation. On the left the severe/non-severe (purple) and the hepatotoxicity class (grey) labeling are indicated as well as the C-max values (green).
Clustering of compounds concentration relationships reveals strong clustering towards severe DILI

Next, we summarized time course data by extracting the time point at which the reporter expression reached a peak response using the various quantitative GFP reporter activity descriptors as well as cytotoxicity measurements, including PI and cell number. Hierarchical clustering revealed one large non activate cluster (compound names in red) and one activated group divided over three clusters (compound names in green, blue and purple) (Fig.4). The purple cluster is marked by an increase in SRXN1-GFP, CHOP-GFP and (for some compounds) p21-GFP in combination with a decrease in ICAM1-GFP. The blue cluster is characterized by a strong increase in CHOP-GFP and cytotoxicity and a strong decrease in ICAM1-GFP, also in lower concentrations. In the green cluster SRXN1-GFP and ICAM1-GFP both show a clear increase. Most severe drugs are in one of the activated clusters, as well as the black and dark grey hepatotoxicity classes. The C-max value for magnitude of activation could be discerned through the column wise increasing C-max level and corresponding maximum response level and increased sensitivity features within each reporter, respectively. For example, the fatal hepatotoxicity compound mercaptopurine showed a relatively strong activation of both the oxidative and inflammatory stress pathways, yet, the point-of-departure for inflammatory stress was at a lower Cmax than for the onset of oxidative stress. In contrast, the fatal hepatotoxicity compound ketoconazole showed no SRXN1-GFP response at the intensity feature level but only at the more sensitive GFP positive cell counts starting earliest at 50 C-max and as the primary and only stress-type. Thus, the current high content data analysis revealed the value of measuring quantitative adaptive stress responses for the different DILI classes with a clear distinction in primary stress responses for individual DILI compounds.

Benchmark concentrations reveal low point of departure for SRXN1-GFP and CHOP-GFP in the severe DILI group

Based on the concentration response curves extracted from the peak response in GFP_pos.2m (SRXN1-GFP, CHOP-GFP and p21-GFP) and GFP_dif.m2 (ICAM1-GFP) we were able to calculate the benchmark concentration (BMC) (Fig. 5A). We defined the BMC as the concentration where at least 25% of the cells reach the two times average threshold of the control values.
**Fig. 4:** Hierarchical clustering of peak GFP response in tims  
Hierarchical clustering of responses of intensity and count related GFP responses in dose response fashion. Cell death measurements cell number and PI staining are in the right bar. On the left side DILI labeling is depicted in three bars: severe/non-severe, DILI-concern labeling and hepatotoxicity class labeling. Compound names are colored based on the clustering.

**Fig. 5:** Benchmark concentration (BMC) versus the absolute C-max values per reporter  
A) Explanation of how we extract BMC from the fitted dose ranges from different GFP reporters. B) For each reporter the absolute BMC (y-axis) is plotted against the absolute C-max (x-axis), each dot represents a compound which reached the 0.25 threshold. Purple indicates a severe DILI drug, Light blue indicates a non-severe drug; * = p<0.05 and ** = p<0.01. C) The BMC is divided by the absolute C-max value per compound and represented in the severe/non-severe DILI classes and per reporter; * = p<0.05 and ** = p<0.01.
A) Obtaining time dynamic features with functional data analysis

Time dynamic feature extraction

B) Feature selection and Unsupervised Clustering

Unsupervised hierarchical clustering of selected features

C) Assessing predictive power severe vs non-severe DILI with Support Vector Machine classification

ROC curves 200 test-set runs: median RDC = 0.73

Compoundcorrected prediction rates

<table>
<thead>
<tr>
<th>Severity Label</th>
<th>No match</th>
<th>Cholestasis/steatohepatitis</th>
<th>Liver amino transferase increased</th>
<th>Hyperbilirubinemia</th>
<th>Jaundice</th>
<th>Liver necrosis</th>
<th>Acute liver failure</th>
<th>Fatal hepatotoxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DILI Concern</td>
<td>No-DILI-C investigator</td>
<td>Ambiguous DILI-&gt;cc</td>
<td>Less-DILI-C investigator</td>
<td>Most-DILI-C investigator</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

80/20 train/test set prediction validation

Median values after 200 runs:

- Balanced Accuracy: 0.67
- Sensitivity: 0.60
- Specificity: 0.75

Fraction of times predicted correctly in test runs
This BMC can function as an indicator for the point of departure for further risk assessment modeling. The C-max values for the screened compounds show a large concentration range which ranges from 1.7 nM to 0.94 mM. To correct for this difference we plotted the BMC against the absolute C-max value (Fig. 5B) and we divided the BMC by the absolute C-max (Fig. 5C). After this correction, we are able to observe a lower BMC in SRXN1-GFP, CHOP-GFP and ICAM1-GFP for severe DILI drugs, compared to non-severe drugs. In SRXN1-GFP and CHOP-GFP this difference is statistically significant (p < 0.001 for CHOP & p < 0.05 for SRXN1 for both methods), indicating severe DILI drugs have a lower point of departure and are therefore more potent to induce adaptive stress responses.

Feature selection and SVM classification show predictive power of a diverse and complex set of biologically interpretable features

Time dynamic features and BMC values were extracted as described in the material and methods section (Fig. 6). The resultant table of 273 variables corresponding to the total set of features and 118 observations corresponding to the compounds was subjected to machine learning to assess temporal stress pathway activation and concentration–response relations for predictive power and feature importance.

Feature selection and support vector machine (SVM) model tuning was performed over 200 iterations of random training/test dataset sampling (Suppl. Fig. 1), as described in detail in the material and methods. The features selected more than 150 times in the 200 iterations were subjected to unsupervised hierarchical clustering (Fig. 6B). Interestingly a diverse set of features, encompassing all reporters, BMC, C-max and also toxicity features including cell death with TNFα at 10 C-max, cell death at 100 C-max and cell speed at 50 C-max. Early and late slope features from the reporters seem to be preferred over the max magnitude values. The resulting clustering with these features shows 3 dominant clusters with enriched severe DILI compounds (purple, blue and green clusters, Fig. 6B).

The 200 independent test-set prediction validations with the tuned SVM model resulted in an average ROC of 0.73 (Fig. 6C, left panel) and a sensitivity of 0.60 and specificity of 0.75 with ‘positive’ being the severe-DILI group. Over the 200 runs the correct prediction rates for each compound was calculated (Fig. 6C, right panel), clearly a subset of 30 compounds have some uncertainty as to being predicted correctly. The remaining compounds are all either predicted correctly, or predicted falsely. No enrichment for DILI-class can be seen for these prediction rates.
Discussion
Here we investigated the integrated application of a panel of four adaptive stress response BAC reporters in high content high throughput screening as a method for DILI liability assessment. We focused on adaptive stress pathway signaling as these defense programs are a universal theme in all life forms and respond before the onset of overt toxicity. We monitored four downstream target genes for oxidative stress (Srxn1), ER-stress/UPR (CHOP), DNA damage (p21) and inflammation (ICAM1); these are selective targets for these pathways. Using CellProfiler and R-package h5CellProfiler we quantified all images and summarized the responses in a time- and concentration dependent relationship. Concentration information was used to extract the benchmark concentration for each compound-reporter combination. This revealed significant changes between the severe and non-severe DILI classes for Srxn1-GFP and p21-GFP. Furthermore, a non-significant yet similar trend was observed for ICAM1-GFP. To perform detailed analysis on the contribution of time and concentration based features to DILI classification, we extracted 20 features and assed these data for predictive power using support vector machine approaches. Using these features we are able to classify severe versus non-severe DILI classes with a sensitivity of 60% and a specificity of 75%. We demonstrate how advanced analysis of live-single cell data can provide key information on the concentration-time course reporter responses that can be applied for DILI liability evaluation. We showed that integration of such mode-of-action assessment using different reporters increases our mechanistic understanding of severe DILI concern compounds.

By using adaptive stress response measurements we are able to dissect detailed information on the mode-of-action of different drugs (Fig. 4). Three differently activated clusters are clearly visible. Strikingly, Srxn1-GFP is often not activated without the activation of other responses (CHOP-GFP, ICAM1-GFP). This indicates oxidative stress is either occurring as a secondary effect after cellular stress induction or as a primary cause which also disturbs other systems as protein folding or inflammation. Since we also capture the time dependent adaptive stress activation dynamics, we are able to make a distinction between these two possibilities (Fig. 3, Suppl. Fig. 2 and 3). Interestingly, the blue colored compound cluster with strong decrease in ICAM1-GFP show also strong increase in CHOP-GFP, but no or minor activation of Srxn1-GFP, meaning that a specific mode-of-action regulates both the inflammation and the ER-stress pathways. Previously, we demonstrated this dual role for diclofenac. Furthermore, for two other members of this cluster, nefazodone and clozapine, this dual activation has been illustrated before. An increase in ICAM1-GFP is often accompanied with an increase in Srxn1-GFP (green cluster). Interestingly, also a decrease in cell number is observed in these drugs, indicating a role for ICAM1-GFP in cell cycle arrest. No major increase of p21-GFP is seen for most compounds, this is probably due to the fact that genotoxicity is thoroughly checked during drug development. The minor role of p21-GFP is especially visible when looking at the benchmark concentration. No difference is observed between the severe and non-severe DILI groups, but this is mainly due to a low amount of compounds (19) that reach the 25% positively
activated cell threshold in \( \text{p}21\text{-GFP} \). In contrast, SRXN1-GFP, CHOP-GFP and ICAM1-GFP show lower benchmark concentrations at the \( \text{C-max} \) value, indicating that the severe DILI group are already activating stress responses at lower concentrations. This is especially pronounced in SRXN1-GFP and CHOP-GFP, where the BMCs of the severe DILI group are significantly lower than the BMCs of the non-severe group after \( \text{C-max} \) correction. For CHOP-GFP this is expected, since CHOP is an important adaptation-adversity-switch in ER stress signaling and therefore activated when the system is more perturbed and closer to cytotoxicity\(^{339} \). For SRXN1-GFP this is less obvious since multiple drugs cause oxidative stress, often followed by adaptation\(^{340} \). Yet, we are able to distinguish the severe from the non-severe DILI group.

The screen was performed in a time-resolved live single cell setting. To date, toxicity screening efforts using high content imaging have mostly focused on single time point fluorescent dyes or anti-bodies\(^{341} \) with several real-time based toxicity screening efforts\(^{342} \). However, the use of dyes and anti-bodies brings additional noise to already very noisy systems as fixation and anti-body binding are likely additional sources of variability; this is not an issue using our reporter models. With the use of our reporter cell lines biological signaling can be visualized with a high time resolution to more accurately pinpoint the primary mode-of-action in relation to cellular stress. Time course signaling data also greatly benefits computational modeling efforts as these require detailed time and dose response dynamics, this is only feasible using live cell imaging data. Furthermore, we were able to extract features based on this time and dose response dynamics. These features were used in support vector machine approach to assess possible differences in cellular adaptive signaling between less- severe and severe DILI. Due to the limited number of no-DILI compounds (\( n = 16 \)) and the total of 118 compounds tested we had obtained a limited set of observations for building predictive models. However by combining the less-Severe DILI and ambiguous DILI cases together with the no-DILI compounds we were still able to show significant predictivity with an independent subset of our data not used in the SVM tuning process. This promising result indicates the importance of adaptive signaling dynamics in DILI. However exact pinpointing of key-features remains difficult without a significantly larger dataset containing a more balanced safe versus unsafe labeled compound annotation. We anticipate the development of an ever increasing database containing such detailed signaling based features linked to chemical exposure, i.e. compound specific biological fingerprints, will ultimately aid in the safety evaluation and early (DILI) prediction of new drugs and chemicals.

In addition to screening more DILI related drugs, more adaptive stress response pathways (heat shock response, mitochondrial toxicity, hypoxia) or reporters (Hmox1-GFP, A20-GFP, Btg2-GFP, BiP-GFP) can be screened. Another issue which affected the result of the prediction is labeling of the drugs in different DILI-concern categories. The labeling of these drugs changes over time, making it difficult to establish a balanced set of most-DILI-concern drugs versus no-DILI-concern drugs\(^{331,335} \). Separating the most-DILI-concern drugs from all other drugs allows classification, however at a cost of including drugs with ambiguous nature related to DILI-severity. This led to poor prediction of several less-DILI
drugs (glimepiride, cimetidine, metformin, omeprazole, tacrolimus, clofibrate, entacapone, phenobarbital, pioglitazone and verapamil) (Fig. 6C, right panel). Some of them are known to induce adaptive stress responses (verapamil and metformin)\textsuperscript{343,344}. In addition, all most-DILI-concern drugs are or have been on the market, meaning they passed toxicity testing stages of drug development. Therefore, when these reporters are to be used in drug development with new chemicals prediction is expected to perform better.

We performed the screen using the HepG2 cell line. Although HepG2 has several advantages for \textit{in vitro} screening (unlimited lifespan, cheap, easy to culture), the major setback is their lack of metabolic capacity. However, several compounds that involve biotransformation-dependent toxicity do show a SRXN1-GFP oxidative stress response (e.g. acetaminophen and sulindac). To test whether there is a concordance between the HepG2 BAC-GFP reporters and the transcript levels in primary human hepatocytes we used the TG-GATES dataset to calculate the correlation of the activity of the reporters genes between TG-GATES and the current BAC-GFP HepG2 DILI screen (Suppl. Fig. 5). This indicates a significant overlap for SRXN1 responses of transcript levels in primary human hepatocyte and BAC-GFP HepG2, indicating a minor role for drug metabolism in the DILI prediction. Still, some of the drugs in the poorly predicted set are supposed to be metabolized (zafirlukast, troglitazone, methimazole, mexilitine). It is noteworthy to mention a large cloud of points exists in the top left (boxed area, Suppl. Fig. 5) demonstrating mainly enhanced SRXN1 protein activation in comparison to PHH transcript levels indicating increased oxidative stress detection sensitivity. Previously, we optimized a HepG2 3D spheroid protocol to enhance liver like properties and to enable chronic exposures\textsuperscript{129}. In future research we can test whether this would increase prediction of DILI drugs.

In conclusion, we have shown that BAC-GFP reporter cell lines are a sensitive tool to provide detailed mechanistic information regarding the adaptive stress response activation in a broad compound screening setting using high-content live single cell imaging. Such detailed insights in the perturbations of signaling pathways after chemical exposure provides key information for safety assessment and possibly predictive purposes. We anticipate that our BAC-GFP reporter platform will contribute to the early pre-clinical screening for DILI liabilities and possibly also other chemical safety assessment paradigms.

**Acknowledgements**

We would to thank Dr. Weida Tong (NCTR, FDA) for providing the DILI compounds and his contribution to this manuscript. This work was financially supported by the Innovative Medicine Initiative MIP-DILI project (Grant agreement No. 115336), the FP7 Seurat-1 program DETECTIVE project (Grant agreement No. 266838), and the European Union’s Horizon 2020 EU-ToxRisk project (Grant agreement No. 681002).
Table 1: Test compound set
Alphabetically sorted list of screened compounds in this study including their c-max values, abbreviations, DILI-concern labeling, severity class, hepatotoxicity class (1 = no hepatotoxicity, 2 = cholestasis/steatohepatitis, 3 = liver aminotransferases increase, 4 = hyperbilirubinemia, 5 = jaundice, 6 = liver necrosis, 7 = acute liver failure, 8 = fatal hepatotoxicity), C-max reference and metabolic potential based on the livertox.nih.gov database (YES = compound is metabolized, NM = compound is not metabolized, NA = not available in the database).
Supplementary Figures

**Image analysis measurements**

**Time course features**
- For cytosol (Supp1, lcam1):
  - Integrated intensity (cell population mean of sum pixel values single cell cytoplasmas)
  - Mean intensity (cell population mean of mean pixel values single cell nuclei)
- For nuclei (Chap. & p21):
  - GFP_pos_2m: GFP positive (fraction cells > 2 X DMSO)
  - GFP_pos_3m: GFP positive (fraction cells > 3 X DMSO)
  - GFP_pos_m3sd: GFP positive (fraction cells > mean+3SD DMSO)
  - Nuclei Area (mean of area of nuclei)
  - Cell number (mean of cell number over duplicate images)
  - Hoechst intensity (cell population mean of mean pixel values nuclei)
  - Cell Speed (cell population mean of cell pixel distances over 2 consecutive frames)
- For lcam1:
  - GFP above 1m/2m/3m (fraction cells > 0.5/1/2 * IQR DMSO first 2 time points)
  - GFP below 1m/2m/3m (fraction cells < 0.5/1/2 * IQR DMSO first 2 time points)

**Normalization & transformation**
- Int: GFP intensity levels normalized by subtracting DMSO per time point on a per plate basis followed by min-max scaling each replicate set between 0 and 1. For lcam1 between -1 and 1.
- Necrosis/ Necrosis_TNFα: necrosis positive fraction without/ with TNFα, normalized by subtracting DMSO/ TNFα levels.
- Slope_CellNumber/ Slope_Nuclei_Area/ Slope_Hoechst_int/ Slope_Cell_Speed: linear regression slopes over time.
- GFP_dif_1m/ GFP_dif_2m/ GFP_dif_3m: GFP above – GFP below, normalized by subtracting TNFα levels.

**Fitting & statistics**

- Natural cubic spline fits time courses
- Time-resampling
- One-way ANOVA bootstrapping functional data
- p-values time courses
- Mean resampled time points
- Time course heatmap

**BMC/ PoD**

- Max over time of 2m GFP pos. fractions
- Fit 4 parameter log logistic model
- Determine BMC at response +0.25 from initial value

**Machine learning**

- Full dataset: 118 compounds
- Test dataset: 12 compounds
- Training dataset: 106 compounds
- repeat 200X
- Feature selection; KS test + pair-wise correlation filter
- SVM tuning with CV
- Prediction assessment

**Suppl. Fig. 1: Data analysis workflow** The features in red are displayed in the figures of the results section.
Suppl. Fig. 2: ICAM1-GFP characterization A) Time lapse images with and without TNFα. B) Comparison BAC-reporter ICAM1-GFP and wild type HepG2 with western blotting after induction with TNFα and stained with ICAM1, GFP and tubulin antibodies. C) Time lapse images of ICAM1-GFP after knock down of caspase 8 and NF-κB subunit RelA and TNFα exposure.
**Suppl. Fig. 3:** Time course graphs of all treatments and concentrations. For all compounds the reporter activity is shown for GFP_pos.2m (SRXN1-GFP, CHOP-GFP and p21-GFP) and GFP_dif.m2 (ICAM1-GFP). Statistics were performed as described in the material and methods section: * = p<0.01
Suppl. Fig. 4: Hierarchical clustering of time course summarized in a heatmap. Red depicts an upregulation and blue a downregulation. On the left the severe/non-severe (purple) and the hepatotoxicity class (grey) labeling are indicated as well as the C-max values (green).
Suppl. Fig. 5: Correlation of TG GATES primary human hepatocytes transcript levels versus BAC-HepG2 SRXN1-GFP (blue), CHOP-GFP (green) and p21-GFP (red) BAC-GFP reporter values are plotted against the TG-GATES fold changes, correlation has been depicted next to the graph. Boxed cloud depicts BAC-GFP reporter activated compounds only. Correlation has been calculated using 61 degrees of freedom, *** = p-value < 0.005.