SUPPLEMENTARY DATA

Multi-omic analysis of zebrafish models of acute organophosphorus poisoning with different severity

Benjamin Piña, Tamar Ziv, Melissa Faria, Shani Ben-Lulu, Eva Prats, Marck A. Arick II, Cristian Gómez-Canela, Natàlia García-Reyero, Arie Admon, Demetrio Raldúa[1]

[1]To whom correspondence should be addressed.

Email: drpqam@cid.csic.es

Title of file for HTML: Supplementary Data
Description: Supplementary Methods and Supplementary Figure 1.

Title of file for HTML: Supplementary Table S1
Description: **David functional analysis of DEG genes from RNAseq and Proteome data (Significantly enriched categories, FDR<5%)**

Supplementary Data

Supplementary Methods

Proteomic analysis

For the four selected phenotypes, protein fractions were extracted from five individual larvae using Trizol Reagent (Invitrogen Life Technologies, Carlsbad, CA), precipitated by 80% acetone and dried for the proteomics study. Proteins were extracted from the dry pellets with 9 M urea, 400 mM ammonium bicarbonate and 10 mM DTT and sonicated. 20 μg protein from each sample were warmed up to 60°C for 30 min and modified with 40 mM iodoacetamide in 100 mM ammonium bicarbonate (in the dark, at room temperature for 30 min). Solutions were diluted with water to 2 M urea, 25 mM ammonium bicarbonate and digested with modified trypsin (Promega) at a 1:50 enzyme-to-substrate ratio, overnight at 37°C, followed by an additional trypsinization for 4 hours. The resulting tryptic peptides were desalted using disposable C18 tips (Harvard Apparatus), dried and resuspended in 0.1% formic acid. Peptides were analyzed by LC-MS/MS using a Q-Exactive-Plus mass spectrometer (Thermo-Fisher Scientific) fitted with a capillary HPLC (Easy nLC 1000, Thermo-Fisher Scientific). They were loaded onto a homemade capillary column (about 25 cm long and 75 micron ID) packed with 3.5 m silica ReproSil-Pur C18-AQ resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) in solvent A (0.1% formic acid in water) (Ishihama et al. 2002). Peptides were resolved with a linear gradient from 5% to 28% of solvent B (95% acetonitrile with 0.1% formic acid) during 105 minutes, followed by gradient of 15 minutes from 28 to 95% B and 15 minutes at 95% B at flow rates of 0.15 μl/minute. Mass spectrometry was performed in the positive ion mode (m/z 350-1800, resolution 70,000) using repetitively full MS scans, followed by higher energy collision dissociation (HCD, at 35 normalized collision energy) of the 10 most dominant ions (>1 charges) selected from the full MS scan. The AGC settings were 3×10^6 for the full MS and 1×10^5 for the MS/MS scans. The intensity threshold for triggering MS/MS analysis was 1×10^4. A dynamic exclusion list was enabled with exclusion duration of 20 sec.

Proteomic data Analysis

Mass spectrometry data were analyzed using the MaxQuant software 1.5.2.8 (www.maxquant.org) (Cox and Mann 2008) fitted with the Andromeda search engine (Cox et al. 2011), searching against the *Danio rerio* Uniprot database (of March 2017 containing 59,064 entries) with mass tolerance of 20 ppm for the precursor masses and the fragment ions. Oxidation on methionine and acetylation on the protein N-terminus were accepted as variable modifications and carbamidomethyl on cysteine was accepted as fixed modification. Minimal peptide length was set to six amino acids and a maximum of two miscleavages was allowed. Peptide and protein level false discovery rates (FDRs) were filtered to 1% using the target-decoy strategy. The identified protein table was filtered to remove the identifications from the reverse database, the common contaminants and single peptide identifications.

Data were quantified by normalized label free analysis using the same MaxQuant software (LFQ intensities), based on extracted ion currents (XICs) of peptides enabling quantitation from each LC/MS run for each peptide identified in any experiment.

Statistical analysis of the identification and quantitation results was performed using Perseus software 1.6 (Tyanova et al. 2016). Student's T-test was done with 0.05 FDR and 250 randomizations. Proteins with P value less than 0.05 and a difference of at least 2 fold between groups were labeled as differential.

Data analysis

Hierarchical clustering and mediod pam clustering analysis were performed using the packages *gplots*, *fpc*, and *cluster* in R for the DEG and DEP sets separately. Significant differences between scaled values of all genes included in each cluster were assessed by one-way ANOVA followed by post hoc Tukey's tests ($p<0.05$) in R (*multcomp* package); further graphs, including heatmaps, were performed with the *gplots* package, also in R.

Functional analysis of a combined DEG+DEP dataset was performed using DAVID Bioinformatic Resources 6.8. Gene enrichment analysis was estimated in DAVID using the default zebrafish (*Danio rerio*) background; enrichment significances were set to a false discovery ratio, FDR, below 5%. Identified modules with at least ten hits were included in the network analysis, using the *reshape2* and *igraph* packages in R

(RCoreTeam 2014). Bipartite graphs were drawn from an incidence table of genes (represented by their official gene names, ZFIN.org). Any two given genes were considered linked if they share at least one common KEGG or GO (Gene Ontology) module.

Metabolomics variations between the three acute OPP models and the control larvae were obtained from a previous publication (Gómez-Canela et al. 2018). Pathway-based integration of this metabolomics data with DEG and DEP datasets was performed through metabolic pathways' knowledge defined by KEGG database. The joint list of identified markers from transcriptomic, proteomic and metabolomics data was introduced as input data into the KEGG database (*D. rerio* pathway dataset) to investigate common metabolic pathways and mechanisms affected by the CPO treatment. Identified pathways with at least two hits were included in the network analysis, using the *reshape2* and *igraph* packages in R. Any given metabolite, DEP or DEG was considered linked to others if they share at least one common KEGG pathway or DAVID functional class.

References

Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26**(12):1367-1372.

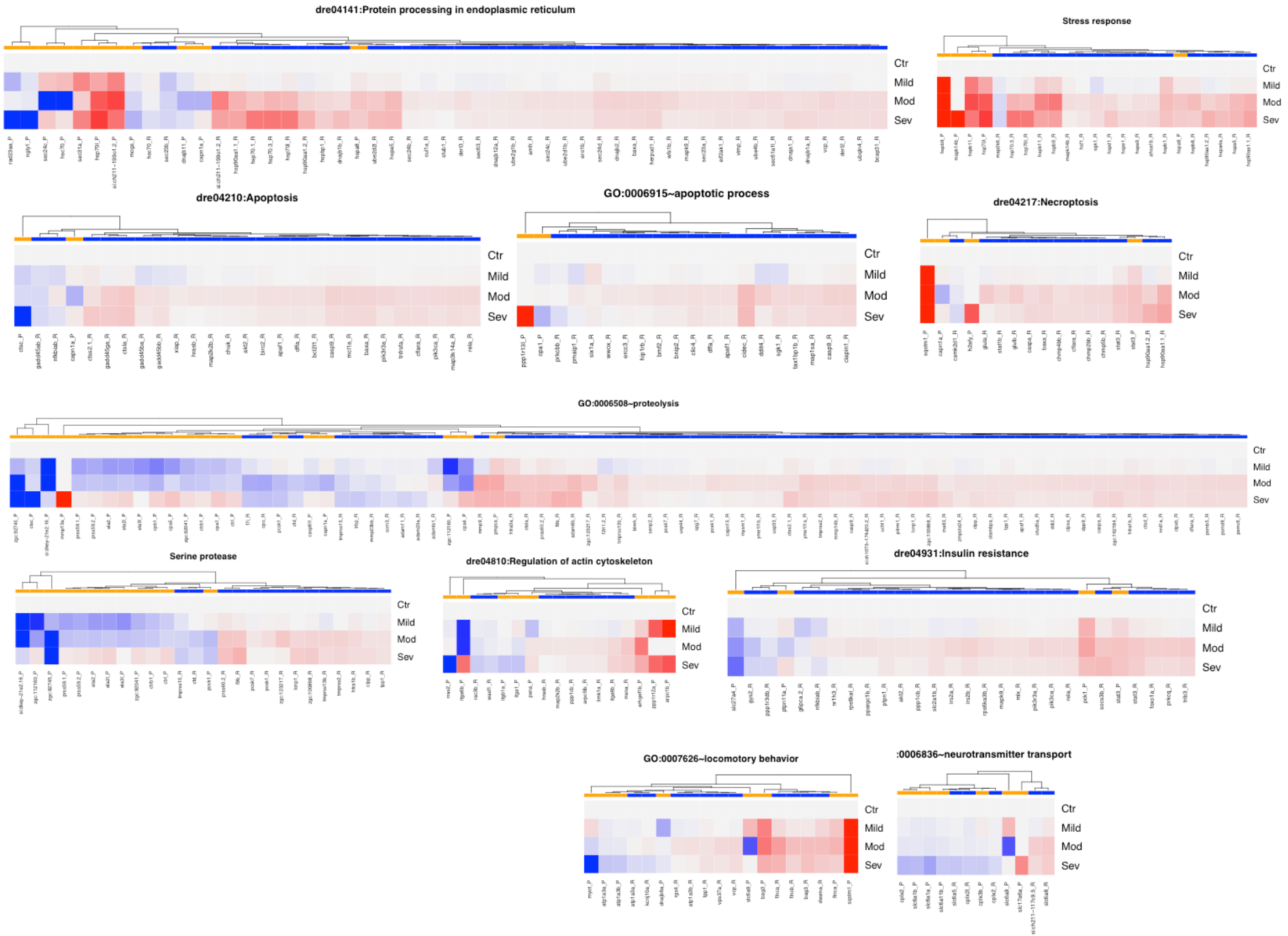
Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research* **10**(4):1794-1805.

Gómez-Canela, C., Tornero-Cañadas, D., Prats, E., Piña, B., Tauler, R., and Raldúa, D. (2018). Comprehensive characterization of neurochemicals in three zebrafish chemical models of human acute organophosphorus poisoning using liquid chromatography-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*, **410**(6), 1-14.

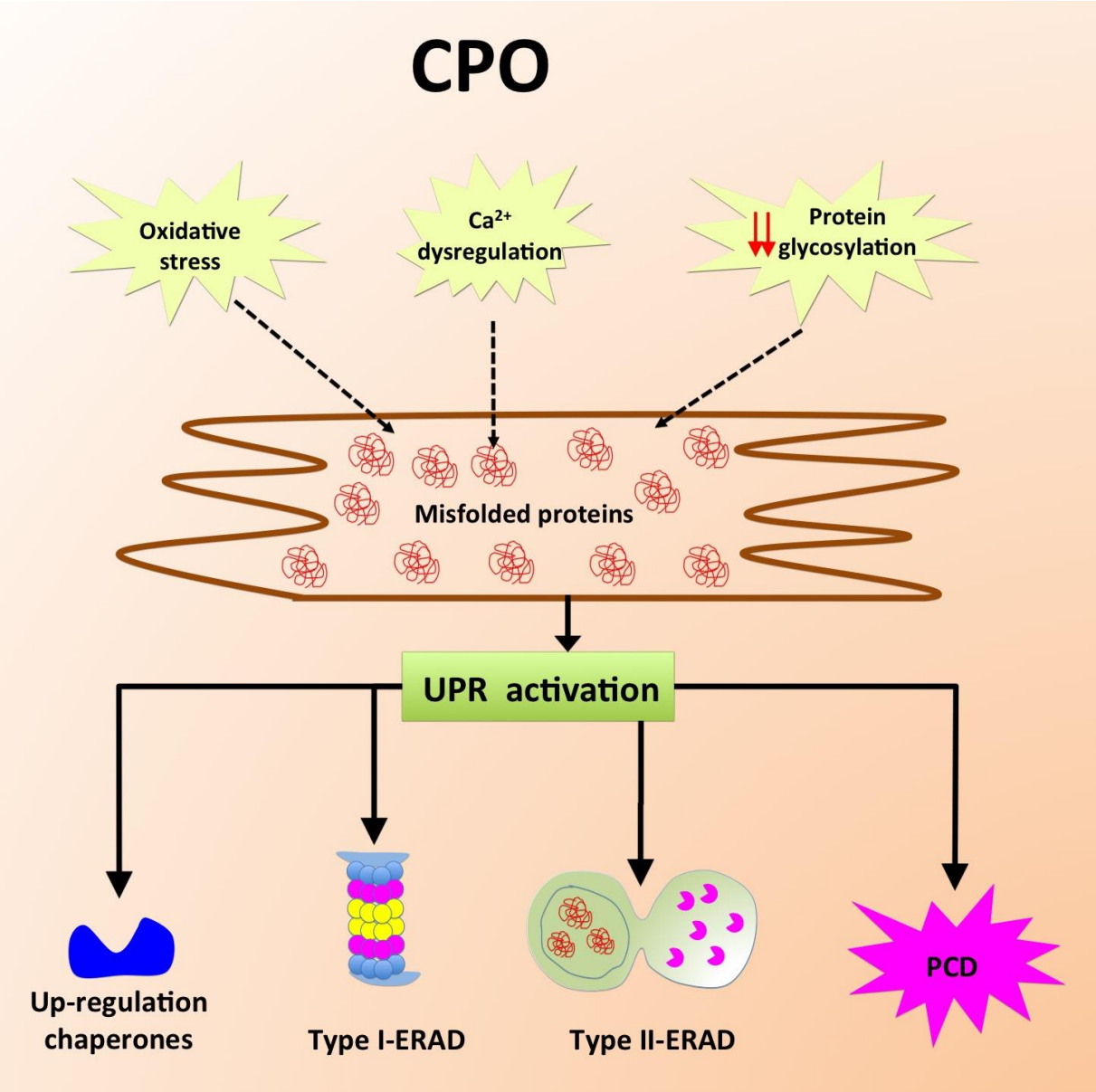
Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002). Microcolumns with self-assembled particle frits for proteomics. *Journal of Chromatography A* **979**(1-2), 233-239.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**(9):731-740.

Supplementary Figures


dre04141:Protein processing in endoplasmic reticulum
Stress response
dre04210:Apoptosis
GO:0006915~apoptotic process
dre04217:Necroptosis
GO:0006508~proteolysis
Serine protease
dre04810:Regulation of actin cytoskeleton
dre04931:Insulin resistance
GO:0007626~locomotory behavior
:0006836~neurotransmitter transport
Ctr
Mild
Mod
Sev

Supplementary Figure S1. Heatmap corresponding to transcriptomic (blue horizontal sectors at the top of the graph) and proteomic data (orange sectors) for all DEGs and DEPs included in different functional categories in the three studied phenotypes. Values correspond to averages from the biological replicates for each transcript/peptide under each condition. Note that control values are set to 0 (white sectors) for each transcript/peptide. Sectors are color coded as in Figure 1.



Supplementary Figure S2. Severe grade of acute OPP model exhibits a molecular phenotype consistent with endoplasmic reticulum (ER) stress. Oxidative stress, Ca^{2+} dysregulation and reduced protein glycosylation induced by CPO may result in the accumulation of unfolded/ misfolded proteins in the ER. The ER stress generated by de accumulation of aberrant proteins is detected by specific sensors triggering the unfolded protein response (UPR). Some of the UPR mechanisms identified in the severe grade larvae include the up-regulation of chaperones, the type I- and II- ER-associated degradation (ERAD) and the programmed cell death (PCD).