Supplemental Material

Transcriptome-Wide Prediction and Measurement of Combined Effects Induced by Chemical Mixture Exposure in Zebrafish Embryos

A. Schüttler, G. Jakobs, J.M. Fix, M. Krauss, J. Krüger, D. Leuthold, R. Altenburger, and W. Busch

Table of Contents

Figure S1. Demonstration of sum(CI) calculation: Sum(CI) is calculated by summing up the differences between the 95% confidence interval for the CA model fit versus the 97.5% quantile of all single substance control measurements for each time point and concentration at which the mixture exposure was investigated.

Figure S2. Mixture design: The mixture proportions for the mixture investigated in this study were optimized for a maximum number of nodes showing a combined effect level > 1.5. A) Line graph exemplarily showing the predicted logarithmic fold-change (log₂FC) of a node of transcripts under mixture exposure for an arbitrarily chosen mixture (in comparison to control conditions). Dashed lines show the expected effects of the mixture components alone, the solid red line shows the predicted mixture effect assuming concentration addition (CA). The combined effect level was retrieved by summing up the distance between predicted effect (assuming CA) and the effect of the mixture component with the strongest effect. This was added up for all time points (3, 6, 12, 24, 48, and 72 hours post exposure) and 5 designed mixture concentrations (from LC0.5 to LC25, which were derived from a predicted lethality curve assuming CA for each mixture proportion), and all nodes of the toxicogenomic universe to arrive at the expected combined effect level for a specific mixture proportion.; B) The count of nodes with a combined effect level > 1.5 is shown for an array of different mixture proportions (total proportion including the naproxen proportion sums up to 1). The black dot illustrates the selected mixture proportion (diuron 11%, diclofenac 2.6%, and naproxen 86.4%) for this study. Numerical data is available in Table S6.
**Figure S3. Lethality curves:** Concentration-response curves for lethality in zebrafish embryos, obtained for the mixture investigated in this study. Dots represent the measured data (provided in Table S7), coloured lines are the respective fitted models (Table S8), and dashed lines indicate the selected exposure concentrations for the toxicogenomic experiment according to equations 1 and 2.

**Figure S4. Concept and results of mixture effect prediction with the concept of independent action (IA):** The calculations and results are exemplarily shown for node node #1119. Single compound log₂FC data retrieved in a previous study were transformed to log₁₀(unaffected fraction) for the concentrations of the substances in the mixture. Mixture prediction based on IA for log₁₀(unaffected fraction) was calculated by summing the individual log₁₀(unaffected fraction) values and the CTR-model was fitted to these values. Predicted log₂FCs and Sensitivity curves as well as predicted component contributions were derived from these values. Respective predictions were plotted on the toxicogenomic universe for the 20 nodes predicted to be significantly affected with IA (Table S9, cf. also Fig. 1 for more details).

**Figure S5. Concept and results of mixture effect prediction with the concept of effect addition (EA):** The calculations and results are exemplarily shown for node node #1118. Single compound log₂FC data retrieved in a previous study were extrapolated to the concentrations of the components in the mixture. Log₂FCs were then summed up and the CTR-model was fitted to derive predictions. Respective predictions on Sₘₐₓ and the component contributions were plotted on the toxicogenomic universe for the 13 nodes predicted to be significantly affected with EA (Table S9, cf. also Fig. 1 for more details).

**Figure S6. Concept and results of mixture effect prediction with the concept of Boolean assumption (BA):** The calculations and results are exemplarily shown for node node #1118. Single compound qualitative expression data retrieved in a previous study were used to derive qualitative expectations regarding the responses to the mixture exposure. If any of the mixture components affected a certain node of the toxicogenomic universe, it was hypothesized that the mixture induces an effect on that node, too. This qualitative concept resulted in predicting 436 nodes of the toxicogenomic universe to be affected due to the mixture exposure (Table S9, cf. also Fig. 1 for more details).

**Figure S7. Experimental mixture effect observation:** Exemplary data analysis workflow and observed whole transcriptome sensitivity towards mixture exposure. A) Measured response of example node (#1061) with three dimensional representation of CTR-model fit as black plane; B) Measured response (black dots) for four of six measured time points with CTR-model fit as black line, half-maximum log₂FC is shown as horizontal red line, EC₅₀s as vertical red lines; C) Fitted sensitivity trend for node #1061; D) Projection of Sₘₐₓ values for all significantly affected nodes, observed under mixture exposure, on the toxicogenomic universe. Dot size represents sum of significant effect (sum(CI)), color represents Sₘₐₓ. All data are available in Table S9.

**Figure S8. Distribution of fitted parameter values of CTR-model:** A) tₘₐₓ (predicted); B) 1/Sₘₐₓ (predicted); C) tₘₐₓ (observed); and D) 1/Sₘₐₓ (observed); CA: concentration addition; CTR-model: concentration and time dependent response model. Data are available in Table S9.
**Figure S9. AIC-weights for CTR-model:** Distribution of AIC-weights when comparing the fit of the concentration and time-dependent regression (CTR) model to a A) null-model (predicted data); B) spline fit (predicted data); C) null model (observed data); and D) spline fit (observed data); CA: concentration addition; CTR-model: concentration and time-dependent response model; AIC: Akaike Information Criterion. Data are available in Table S9.

**Figure S10. Quantitative predictivity of different mixture concepts:** Prediction deviation ratios (PDRs) of A) $S_{\text{max}}$ and B) $t_{\text{max}}$ for all 204 nodes, either predicted with CA or observed in the experiment to be affected due to the mixture exposure, projected on the toxicogenomic universe; Distributions of PDRs of C) $S_{\text{max}}$, and D) $t_{\text{max}}$ values; CA: concentration addition, IA: independent action, EA: effect addition. Data provided in Table S9.

**Figure S11. Example node (#876) showing gene expression in two directions:** Node #876 showing regulation in different directions with different exposure time, concentration, or substance (black dots); A) single compound exposures; B) mixture exposure; black dots: individual measurements; black lines: fitted CTR-model, red lines: CA predictions.

**Figure S12. Example node (#34) for undue significance threshold estimation:** In node #34 the significance threshold based on single substance control data was much higher than the significance threshold based on the control data from the mixture exposure experiment. This was leading to a false-negative prediction for this node.

**Figure S13. Example node (#818) showing high variation between log$_2$FCs of assigned genes:** Node #818 contains four genes; Cyp1c1 shows higher log$_2$FCs due to A) single substance exposures, and B) mixture exposure; dots: individual measurements; black and red lines: observed and predicted CTR-model fits, respectively, for this node.

**Figure S14. Example node (#20) showing unexpected mixture effect:** A) single substance exposures caused no down-regulation at 12 hours post exposure (hpe) in node #20; B) a clear down-regulation at 12 hpe was observed due to mixture exposure, black dots: individual measurements; black lines: fitted CTR model, red lines: CA predictions.

**Figure S15. Qualitative predictivity of different mixture concepts:** Confusion matrices for numbers of expected and observed nodes based on A) Boolean assumption (BA), B) effect addition (EA), and C) independent action (IA); D) F1-score for the different mixture concepts.

**Figure S16. Qualitative predictivity of the dose-scaled Boolean assumption (BA):** A) Confusion matrices for Boolean assumption when single compound effects were extrapolated to the concentrations and time points applied in the mixture exposure experiment; B) F1-score for each concentration and two selected time points (hpe: hours post exposure).
Figure S17. Prediction accuracy and compound contributions: A) Demonstration of overlap calculation. The overlapping area between predicted and observed sensitivity curves (dark red) was set in relation to the total area under both curves (red, dark red, and grey). B) Histogram of overlap between the areas of predicted and observed sensitivity curves for all nodes either predicted (with CA) or observed to be affected under mixture exposure; C) Node-wise overlap in relation to the predicted maximum effect unit (max(EU_i)), colors according to predicted component contributions and subtractive CMYK color code (cyan: diuron, magenta: diclofenac, yellow: naproxen), opacity indicates overlap. Data available in Table S9.

Additional File- Excel Document and HTML file