West Virginia Chemical Spill: 5-Day Rat Toxicogenomic Studies

Materials and Methods

Experimental Design

At each of six dosages (0.1, 1, 10, 100, 300, and 500 mg/kg body weight (bw) for MCHM and crude MCHM; or 1, 10, 100, 500, 1000, and 2000 mg/kg bw for PPH), together with vehicle control (MCHM, crude MCHM, PPH 0 mg/kg bw; corn oil), 6-8 week old, male Harlan Sprague Dawley rats received the test article via gavage on each of 5 consecutive days (Table 1). The study groups treated with test article consisted of six animals each; eight animals were in the corresponding vehicle control, 0 mg/kg bw groups. Dosage volume was 5 mL/kg bw and was based on most recently measured bw. Dosing was carried out in the morning with completion by approximately noon each day. Dosing of a treatment group was completed in entirety before moving to the next group. Euthanasia and tissue sample collection were carried out on the day following the final administration of the test articles.

Table1. Chemicals Evaluated in the 5-Day Toxicogenomics Study[a]

CASRN*	Chemical Name	Dose Levels (mg/kg bw/day)	Notes
34885-03-5	4-Methylcyclohexanemethanol (MCHM)	0, 0.1, 1, 10, 100, 300, 500	b
NA	Crude 4-Methylcyclohexanemethanol (crude MCHM)	0, 0.1, 1, 10, 100, 300, 500	c
770-35-4	Propylene glycol phenyl ether (PPH)	0, 1, 10, 100, 500, 1000, 2000	d

*CASRN = Chemical Abstract Registry Number; [a]The original 5-day toxicogenomic study protocol contained an evaluation of dipropylene glycol phenyl ether (DiPPH) in addition to MCHM, crude MCHM, and PPH. NTP was not able to acquire the DiPPH test article at the same time as the other chemicals that were evaluated in the 5-day studies. Following review of the initial data and consideration of the low abundance of DiPPH in the spill, the 5-day study of DiPPH was de-prioritized and not performed; [b]Major constituent of the spilled liquid; [c]Commercial mixture containing >70% MCHM along with lesser amounts of other chemicals; [d]Minor constituent of the spilled liquid.

Moribundity/Mortality

All animals were observed twice daily, morning and afternoon, at least 6 hours apart (no later than 10 am, and no earlier than 2 pm), for signs of moribundity/mortality. Any animal showing signs of moribundity was humanely terminated.

Body Weights

On Study Day 1, prior to administration of test article, body weights of study animals were recorded. Thereafter, body weights were recorded daily and prior to euthanasia on the final day of the in-life portion of the study.

Clinical Observations

Clinical observations were recorded daily for all animals and prior to euthanasia on the final day of the in-life portion of the study. Any observations, which were thought related to exposure to the test article, that were made at times other than during the scheduled observations were also recorded.

Blood Collection and Prioritization

Blood samples at study termination were taken in random order from the abdominal aorta of all surviving rats within each treatment group while under CO_2/O_2 (70/30) anesthesia. In the case where insufficient blood volume was available for all study endpoints, the order of priority was micronuclei determination, clinical pathology, followed by serum metabolomics.

Erythrocyte Micronuclei Determination

Whole blood samples (~200 µL) were collected into tubes containing anticoagulant (tripotassium ethylene diamine tetraacetic acid; K_3EDTA) and refrigerated immediately following collection. Samples were shipped on ice on the day of collection to the NIEHS-designated contract laboratory for

micronuclei determination. Samples were immediately refrigerated, and then shipped with cold packs by overnight courier to the analytical laboratory where they were immediately fixed in ultracold methanol (MicroFlow® Basic Kits, Litron Laboratories, Rochester NY) and stored in a -80° freezer until analysis. Flow cytometric analysis was conducted using a FACSCalibur flow cytometer (Becton Dickinson, San Jose, CA). Reticulocytes (RET; immature red blood cells) were identified by the presence of an active transferrin receptor (CD71+) on the cell surface; mature erythrocytes were identified as CD71-negative.

The analysis of micronuclei (MN) was restricted to the youngest RET (i.e., the subpopulation of erythrocytes with the highest CD71 expression), in order to focus on the population of RET that were least altered by the action of the rat spleen in sequestering and destroying micronucleated red blood cells. Using flow cytometry, MN were detected using the DNA staining dye propidium iodide (PI) in conjunction with RNase treatment. MN-RETs express high levels of CD71 (CD71+) and PI-associated fluorescence, while MN-erythrocytes are negative for CD71 (CD71−) and show PI-associated fluorescence. Twenty thousand CD71+ RET were scored per animal for presence of MN, and approximately 1 x 10^6 total erythrocytes were counted for presence of MN and to determine percentage of RET (%RET) as a measure of chemical induced bone marrow toxicity.

In the erythrocyte micronucleus assay, the animal was considered the experimental unit and approximately 20,000 reticulocytes and/or 1 x 10^6 erythrocytes were evaluated per animal for presence of micronuclei. In addition, the %RET was determined in approximately 1 x 10^6 erythrocytes. The optimum number of cells to score for micronuclei using flow cytometric approaches was determined in earlier studies.[1] Data from each treatment group were summarized as the mean frequency of MN-RET per 1000 RET, plus or minus the standard error of the mean.

With the large number of cells counted by flow cytometry, an assumption is that the number of micronucleated cells is normally distributed. Levene's test was used to determine whether variances among treatment groups were equal. When variances were equal, linear regression analysis was used to test for linear trend. Pairwise differences with the control group were evaluated using Williams' test, after linearizing the data by averaging data points that violate a linear trend. When variances were unequal, nonparametric methods were used to analyze the data: Jonckheere's test[2] to evaluate linear trend and Dunn's test[3] to assess the significance of pairwise differences with the control group. To maintain the overall significance level at 0.05, the trend as well as the pairwise differences from the control group was declared statistically significant if $P < 0.025$. The final (positive or negative) call was determined after considering the results of statistical analyses, reproducibility of any effects observed, and the magnitudes of those effects.

Hematology

Red blood cell count, hemoglobin, hematocrit, manual hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, white blood cell count, differential leukocyte count, reticulocyte count, and platelet count were determined on whole blood samples. In addition, morphological features in all cellular components were qualitatively evaluated.

Clinical Chemistry

Total protein, albumin, urea nitrogen, creatinine, alanine aminotransferase, sorbitol dehydrogenase, alkaline phosphatase, total bile acids, glucose, creatine kinase, cholesterol, triglycerides, bilirubin (total

[1] Kissling GE, Dertinger SD, Hayashi M, MacGregor JT. Sensitivity of the erythrocyte micronucleus assay: dependence on number of cells scored and inter-animal variability. Mutat Res. 2007 Dec 1;634(1-2):235-40.
[2] Jonckheere, A.R. (1954). A distribution-free k-sample test against ordered alternatives. Biometrika 41: 133-145.
[3] Dunn, O.J. (1964). Multiple comparisons using rank sums. Technometrics 6, 241-252.

and direct), sodium, potassium, and chloride were determined on whole blood samples, collected without anticoagulant.

Necropsy

Animals surviving at the end of the study were subjected to complete necropsy. Feed and water were provided until termination. Termination was performed by CO_2 inhalation overdose. Necropsies were performed on control group animals first, followed by dose groups in descending order of dose. All body orifices and tissues were examined. Gross lesions were recorded. Organ weights for liver, thymus, kidneys, testes, epididymis, heart, and lungs were recorded for animals surviving until scheduled termination.

Statistical Analysis of Non-Microarray Data

Two approaches were employed to assess the significance of pairwise comparisons between dosed and control groups in the analysis of continuous variables. Organ and body weight data, which have approximately normal distributions, were analyzed using the parametric multiple comparison procedures of Williams[4,5] and Dunnett.[6] Hematology, hormone data, clinical chemistry, and urine chemistry data, which typically have skewed distributions, were analyzed using the nonparametric multiple comparison methods of Shirley[7] and Dunn.[8] Jonckheere's test [9] was used to assess the significance of dose-response trends and determine whether a trend-sensitive test (Williams' or Shirley's test) was more appropriate for pairwise comparisons than a test that does not assume a monotonic dose-response (Dunnett's or Dunn's test). Trend-sensitive tests were used when Jonckheere's test was significant at $p<0.01$.

Prior to analysis, values identified by the outlier test of Dixon and Massey,[10] were examined by NTP staff. Values from animals that were suspected of being sick due to causes other than experimental treatment, and values that the laboratory indicated as being inadequate due to measurement problems were eliminated from the analysis.

Liver and Kidney Processing at Necropsy

At termination, within 5 minutes subsequent to blood collection, liver and kidneys were removed and weighed prior to other tissues. The left lobe of the liver was divided in half longitudinally and a section fixed. Kidneys were dissected free of fat and a longitudinal section was prepared and fixed. Tissues were fixed for ~24h in 10% neutral buffered formalin (NBF), transferred to 70% ethanol, and refrigerated until completion of processing, within 5 days. Tissues were trimmed and paraffin embedded. The remaining left liver lobe and right kidney were processed for RNA isolation. Tissues were quickly placed into RNA*later*™. Tissue samples were stored at 2 to 8°C overnight, RNA*later*™ removed, and stored at or below -70°C until processed for RNA isolation and microarray analysis. The left kidney and the remainder of the left liver lobe were plunged into liquid nitrogen and stored at -70°C.

---

[4] Williams, D.A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics 27: 103-117.

[5] Williams, D.A. (1972). The comparison of several dose levels with a zero dose control. Biometrics 28: 519-531.

[6] Dunnett, W. (1955). A multiple comparison procedure for comparing several treatments with a control. JASA 50: 1095-1121.

[7] Shirley, E. (1977). A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. Biometrics 33: 386-389.

[8] Dunn, O.J. (1964). Multiple comparisons using rank sums. Technometrics 6, 241-252.

[9] Jonckheere, A.R. (1954). A distribution-free k-sample test against ordered alternatives. Biometrika 41: 133-145.

[10] Dixon, W. and Massey, F. (1951). Introduction to Statistical Analysis. McGraw Hill, New York, pp. 145-147.

RNA Isolation

RNA was isolated from tissue samples preserved in RNA*later*®, using the Qiagen RNeasy Mini Kit (Qiagen Inc., Valencia, CA) with a deoxyribonucleic acid (DNA) digestion step. Concentration and purity was determined from absorbance readings taken at 260 and 280 nm using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE). Purity was deemed acceptable at A_{260}/A_{280} ratios between 1.80 and 2.20. In addition, samples were processed with the RNA 6000 Nano kit and analyzed with a 2100 Bioanalyzer (Agilent Technology, Foster City, CA). With purity and RNA integrity number (RIN) determined, samples were divided into two aliquots; one aliquot was used for Affymetrix microarray expression analysis and the other submitted frozen to NTP Archives. Liver tissue samples remaining after RNA isolation were stored at ≤ -70°C and submitted frozen to the NTP Archives.

Microarray Expression Analysis

Microarray expression analysis was performed using total RNA (target mass 50 to 500 ng) isolated from 87 liver and 87 kidney samples selected at random; 29 samples were obtained for each of three test articles. Only RNA samples with a concentration of ≥ 35 ng/µL, purity between 1.80 and 2.20, and RIN ≥ 8.0 were evaluated. RNA was amplified through cDNA synthesis, *in vitro* transcription, and biotin labeling using the GeneChip® 3' IVT Express Kit (Affymetrix, Santa Clara, CA). The amplified RNA (aRNA) was fragmented and hybridized to Affymetrix Rat Genome 230 2.0 Array. Each array was scanned using an Affymetrix GeneChip® Scanner, model 3000 7G to generate microarray image data (.DAT files) and raw expression level data (.CEL) files. A file ("Annotation and processing details file for Elk River microarrays_16-12-2016"), containing detailed annotation of the processed batches for each of the steps, is included with the accompanying data set.

Microarray Data Analysis

The raw microarray data (.CEL files) was normalized by applying the Robust Multi-array Average (RMA) algorithm[11] from the genomics analysis tool, GeneSpring GX 12.6 (Agilent Technology, Foster City, CA). The normalized data were examined by visual inspection, using a Principal Component Analysis (PCA) plot (Figure 1) and standard quality control metrics describing the technical accuracy of the sample processing. Quality control analysis led to removal of a subset of samples from subsequent analysis, specifically because of poor internal control results (hybridization or cDNA synthesis) or because the samples exhibited clear batch related effects that were observed with principle component analysis (Table 2). Importantly, due to randomization during processing of samples, there was limited impact on the study. With the exception of four dose groups in the crude MCHM kidney studies, gene expression data from a minimum of 3 animals in each dose group from kidney and liver passed quality control and were used in the gene expression benchmark dose analysis. In the 0.1, 1, 10, and 100 mg/kg bw dose groups from the crude MCHM kidney studies, gene expression data from two animals were used in the gene expression benchmark dose analysis. A detailed microarray processing annotation is available for download through file transfer (ftp://157.98.192.110/ntp-cebs/datatype/microarray/ElkRiver/).

[11] Irizarry,R.A. et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249-264.

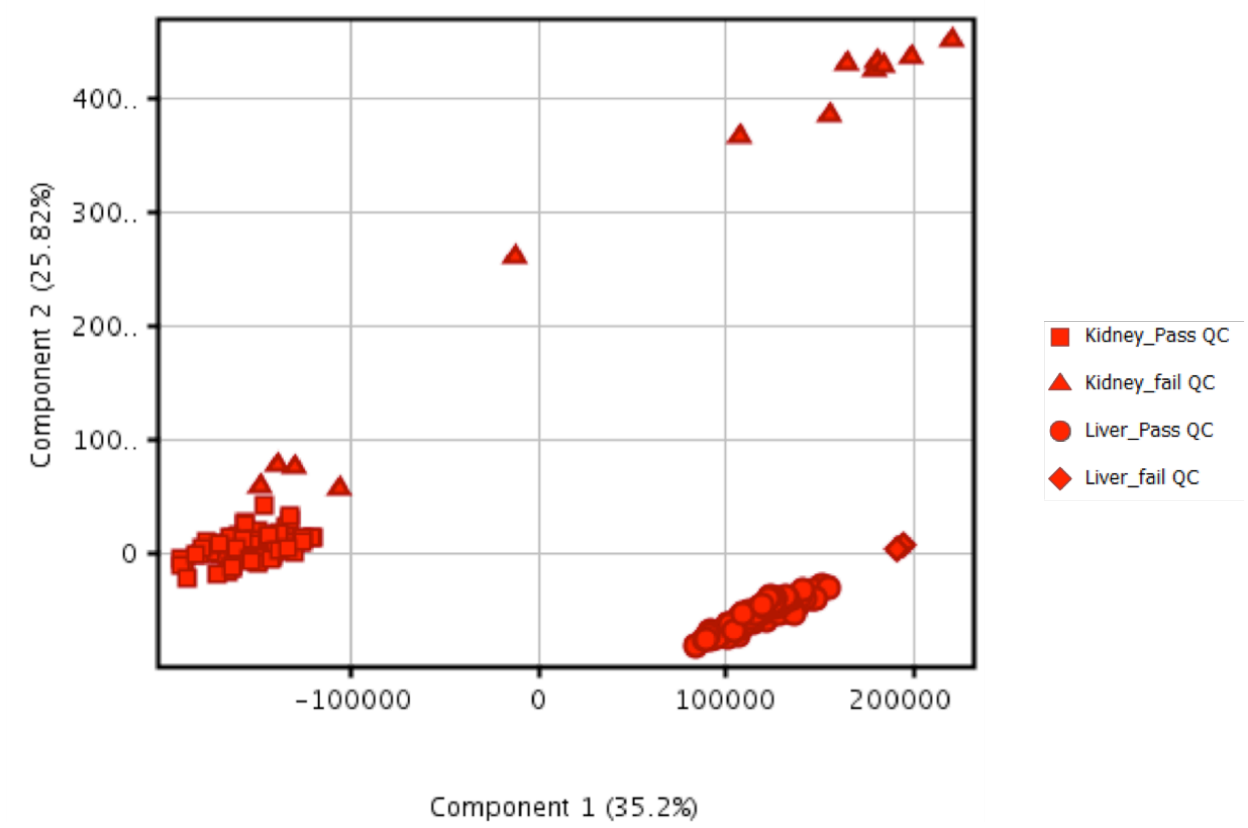

Figure 1. Principal Component Analysis of Microarray Samples Showing Samples that Failed Quality Control (QC = quality control)

Table 2. Microarray Files Removed from Analysis Due to Quality Control Failure

GeneChip File Name	Dose Group	Chemical	Tissue
102K-120814-JAP_(Rat230_2).CEL	Group 1: 0 mg/kg bw	Crude MCHM	Kidney
102L-112014-MW_(Rat230_2).CEL	Group 1: 0 mg/kg bw	Crude MCHM	Liver
108K-120814-JAP_(Rat230_2).CEL	Group 1: 0 mg/kg bw	Crude MCHM	Kidney
112K-120814-JAP_(Rat230_2).CEL	Group 2: 0.1 mg/kg bw	Crude MCHM	Kidney
112L-112014-MW_(Rat230_2).CEL	Group 2: 0.1 mg/kg bw	Crude MCHM	Liver
113K-120814-JAP_(Rat230_2).CEL	Group 2: 0.1 mg/kg bw	Crude MCHM	Kidney
116K-120814-JAP_(Rat230_2).CEL	Group 3: 1 mg/kg bw	Crude MCHM	Kidney
117K-120814-JAP_(Rat230_2).CEL	Group 3: 1 mg/kg bw	Crude MCHM	Kidney
122K-120814-JAP_(Rat230_2).CEL	Group 4: 10 mg/kg bw	Crude MCHM	Kidney
125K-120814-JAP_(Rat230_2).CEL	Group 4: 10 mg/kg bw	Crude MCHM	Kidney
130K-120814-JAP_(Rat230_2).CEL	Group 5: 100 mg/kg bw	Crude MCHM	Kidney
131K-120814-JAP_(Rat230_2).CEL	Group 5: 100 mg/kg bw	Crude MCHM	Kidney
138-120814-JAP_(Rat230_2).CEL	Group 6: 300 mg/kg bw	Crude MCHM	Kidney
140K-120814-JAP_(Rat230_2).CEL	Group 7: 500 mg/kg bw	Crude MCHM	Kidney
231K-112414-JAP_(Rat230_2).CEL	Group 5: 500 mg/kg bw	PPH	Kidney

Probe Set Benchmark Dose (BMD) Analysis

Dose-response analyses of RMA normalized probe sets intensities from Affymetrix Rat Genome 230 2.0 Arrays were performed using BMDExpress 2.00.225 beta.[12] First, the data were divided into six studies (MCHM liver, MCHM kidney, crude MCHM liver, crude MCHM kidney, PPH liver, PPH kidney). Control genes (AFFX-) were removed from each dataset. A statistical (ANOVA, $P<0.05$) and fold change filter (1.5 fold change up or down relative to control for probe sets that were passed into the MBP analysis and 2-fold change for genes passed in the individual gene analysis) was applied to each data set to remove probe sets from subsequent analysis that did not demonstrate a response to chemical treatment. Hill, Power, Linear, Polynomial 2°, Polynomial 3°, and Polynomial 4° dose-response models were fit to the probe sets that passed the statistical and fold change filter in each of the six data sets. After all models were fit to a probe set, a "best" fit was chosen for each probe set based on the prescribed evaluation of fit's p-value, log (likelihood) fit, and Aikake information criterion (AIC).[12] The specific parameter settings, selected from the BMDExpress software when performing probe set-level BMD analysis, were as follows: **maximum iterations**: 250, **confidence level**: 0.95, **BMR factor**: 1.349 (the multiplier of the standard deviation that defined the BMD), **restrict power**: no restriction, and **Constant variance**: selected. The specific model selection setting in the BMDExpress software when performing probe set level BMD analysis was as follows: **best poly model test**: Nested Chi Square, **p-value cutoff**: 0.05, **flag hill model with 'k' parameters**: < 1/3 the lowest positive dose, and **best model selection with flagged hill model**: select Next Best Model with P-Value > 0.05.

[12] Yang L., Allen B. C., Thomas R. S. (2007). BMDExpress: A software tool for the benchmark dose analyses of genomic data. BMC Genomics. 8, 387.

Molecular Biological Process Benchmark Dose Analysis

BMDExpress was also used to identify dose-related changes in molecular biological processes (groups of genes that work together to produce a biological effect). Two sets of molecular biological processes were used in the analysis: KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways[13] and GO (Gene Ontology) biological processes[14]. Probe sets exhibiting significant differential expression (ANOVA, $P<0.05$) and a ≥ 1.5-fold change up or down were considered for MBP BMD Analysis. First best-fit models for each probe set were filtered to remove those that target the same gene, but demonstrated a correlation less than 0.5. Probe sets were then sorted into molecular biological processes (KEGG pathways and GO biological processes). The populated MBPs were then filtered to remove those that were < 5% populated, contained <5 genes with curve fits, and were not significantly populated (i.e., those with a Fischer Exact test p-value >0.05). MBPs that passed these filters were considered "active." BMD values for the active molecular biological processes were then determined by identifying the median and mean BMD and BMD_L for the probe sets in that MBP. The extra layers of filtering MBPs were performed to increase the stringency of the analysis and reduce the contribution of noise to the final results. The BMD values for the active molecular biological processes are reported in the accompanying files (i.e., "MCHM Active Molecular Biological Process Median BMD values," "Crude MCHM Active Molecular Biological Process Median BMD values," and "PPH Active Molecular Biological Process Median BMD values").

Individual Gene Benchmark Dose Analysis

BMDExpress was also used to identify dose-related changes in individual genes. Probe sets exhibiting significant differential expression (ANOVA, $P<0.05$) and a ≥ 2.0-fold change up or down were considered for individual gene BMD Analysis. First best-fit models for each probe set were filtered to remove those with a fit p-value <0.1 and BMD/BMD_L ratio > 20. These probe set filters were used to remove poorly fit models and to remove probe sets with high degree of uncertainty. Probe sets were then filtered to remove those that target the same gene, but demonstrated a correlation less than 0.5. Probe sets that passed these filters were then associated with their genes; it is often the case that multiple probe sets target the same gene. BMD values for the individual genes were then determined by identifying the BMD and BMD_L for the genes. When there were multiple probe sets per gene, the median BMD and BMD_L of those probe sets were reported.

A comprehensive set of results is available through a file transfer site (ftp://157.98.192.110/ntp-cebs/datatype/microarray/ElkRiver/).

[13] http://www.genome.jp/kegg/kegg1.html
[14] http://geneontology.org/