



Metallomics

Identification of Novel Activators of the Metal Responsive Transcription Factor (MTF-1) Using a Gene Expression Biomarker in a Microarray Compendium

Journal:	<i>Metallomics</i>
Manuscript ID	MT-ART-03-2020-000071.R2
Article Type:	Paper
Date Submitted by the Author:	30-Jun-2020
Complete List of Authors:	Jackson, Abigail; Duke University Liu, Jerry; EPA Vallanat, Beena; EPA Jones, Carlton; US EPA, Nelms, Mark; EPA Patlewicz, Grace; EPA Corton, J.; US EPA,

SCHOLARONE™
Manuscripts

ARTICLE

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Identification of Novel Activators of the Metal Responsive Transcription Factor (MTF-1) Using a Gene Expression Biomarker in a Microarray Compendium

Abigail C. Jackson^{1,2}, Jie Liu¹, Beena Vallanat¹, Carlton Jones¹, Mark D. Nelms^{1,3}, Grace Patlewicz¹ and J. Christopher Corton^{1*}

Environmental exposure to metals is known to cause a number of human toxicities including cancer. Metal-responsive transcription factor 1 (MTF-1) is an important component of metal regulation systems in mammalian cells. Here, we describe a novel method to identify chemicals that activate MTF-1 based on microarray profiling data. MTF-1 biomarker genes were identified that exhibited consistent, robust expression across ten microarray comparisons examining the effects of metals (zinc, nickel, lead, arsenic, mercury, and silver) on gene expression in human cells. A subset of the resulting 81 biomarker genes was shown to be altered by knockdown of the MTF1 gene including metallothionein family members and a zinc transporter. The ability to correctly identify treatment conditions that activate MTF-1 was determined by comparing the biomarker to microarray comparisons from cells exposed to reference metal activators of MTF-1 using the rank-based Running Fisher algorithm. The balanced accuracy for prediction was 93%. The biomarker was then used to identify organic chemicals that activate MTF-1 from a compendium of 11,725 human gene expression comparisons representing 2582 chemicals. There were 700 chemicals identified that included those known to interact with cellular metals, such as clioquinol and disulfiram, as well as a set of novel chemicals. All nine of the novel chemicals selected for validation were confirmed to activate MTF-1 biomarker genes in MCF-7 cells and to lesser extents in MTF1-null cells by qPCR and targeted RNA-Seq. Overall, our work demonstrates that the biomarker for MTF-1 coupled with the Running Fisher test is a reliable strategy to identify novel chemical modulators of metal homeostasis using gene expression profiling.

Introduction

Homeostasis of metal ions is tightly regulated in biological systems due to the toxicity of many metals 1-3. At high concentrations, even essential metals like copper, zinc, iron, and manganese, which have important structural and catalytic functions in proteins, can cause toxicity. Human exposures to metals have been increasing as a result of increased use in a number of applications 4. Sources of heavy metals in the environment include geogenic, industrial, agricultural, pharmaceutical, domestic effluents, and atmospheric sources 5. Environmental pollution is significant from point sources involving mining and in foundries, smelters, and other metal-based industrial operations 5, 6. Heavy metals, especially nonessential metals such as arsenic, cadmium, chromium, lead, and mercury are of public health significance due to their

¹Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711

²Department of Chemistry, Duke University, Durham, NC 27708

³Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA.

*Corresponding author

Chris Corton

Center for Computational Toxicology and Exposure

US Environmental Protection Agency

109 T.W. Alexander Dr.

MD-B143-06

Research Triangle Park, NC 27711

corton.chris@epa.gov

919-541-0092 (office)

919-541-0694 (fax)

ability to act as systemic toxicants that induce damage in multiple organs, even at low levels of exposure. Many of these metals are classified as known or probable human carcinogens according to the US Environmental Protection Agency (EPA) and the International Agency for Research on Cancer.

One of the most important biological mechanisms used by cells to regulate and protect against toxic metal exposure is sequestration by metallothioneins (MTs) which are small, cysteine-rich proteins that bind excess metal ions under conditions of metal overload 7. During metal overload, MT-bound zinc ions are displaced, increasing levels of cellular zinc ions that activate the metal responsive transcription factor 1 (MTF-1), the major transcriptional regulator of toxic responses to metal overload. MTF-1 contains six highly conserved DNA binding zinc finger domains, which are sensitive to changes in zinc concentration. Excess zinc binds to the zinc fingers increasing the binding affinity of MTF-1 with metal responsive elements (MREs) in the regulatory regions of target genes including MT family members allowing zinc-inducible gene regulation. The MTF-1-MT regulon forms a negative feedback loop in which activation of MTF-1 leads to increases in the expression of MT family members, sequestration of excess zinc, and subsequent decreases in MTF-1 activation 8-10. MTF-1 also increases the expression of a zinc transporter (ZnT1) which reduces intracellular zinc availability by promoting zinc efflux from the cytoplasm or into intracellular vesicles 11.

There is evidence that increases in oxidative stress can activate MTF-1 through effects on MT family members. During conditions of oxidative stress or hypoxia, MTs become oxidized resulting in structural alterations that lead to release of zinc 12-14. Like metal overload, the increase in free zinc in the cell leads to activation of MTF-1. The universality of oxidative stress-induced MTF-1 under different chemical exposure conditions has not been previously explored. Nuclear factor erythroid 2-related factor 2 (Nrf2) is another transcription factor that responds to cellular oxidative stress; however, the relationship between the activation of MTF-1 and Nrf2 pathways is not fully defined but may include inactivation of the negative regulator of Nrf2 called Keap1 which has a phylogenetically conserved Zn binding site 15.

Because of the relevance of metal exposure to human health and disease, high-throughput screening (HTS) methods to identify conditions in which metal homeostasis is disturbed would be useful to predict cellular responses to toxic metal overload and oxidative stress. The only known HTS assay that potentially measures alterations in metal homeostasis through MTF-1 activation is part of the EPA ToxCast screening program (<http://epa.gov/ncct/toxcast/>; Attagene, "ATG_MRE_CIS_up") which measures the ability of endogenously expressed MTF-1 to activate a reporter gene in HepG2 cells. There are ~700 HTS assays representing ~350 molecular targets that have been used to screen more than 1800 chemicals 16. High-throughput transcriptomic (HTTr) technologies have now been added to the ToxCast screening battery 17 in which targeted sequencing

techniques are used to assess the expression of the human genome in chemically-treated cells 18. HTTr is being used as a global assay to potentially identify any pathway that is perturbed upon chemical exposure, not just the ones that examined by HTS assays. Targeted HTS assays could then be used to confirm the putative modulators 17. Approaches to assess MTF-1 activation in HTTr data have not been described but would be potentially useful to identify environmentally-relevant chemicals that perturb metal homeostasis. In the present study, we describe procedures for assessing MTF-1 activation using computational analysis of gene expression data. The large quantity of microarray data that already exists in public repositories and in commercial databases allows for in silico HTS prediction of chemical agents that activate or suppress a wide range of human molecular targets including MTF-1. As proof of principle, our lab has built and characterized gene expression biomarkers that predict modulation of the estrogen receptor and androgen receptor in human cell lines 19, 20. Using similar approaches, we constructed a gene expression biomarker that accurately detects MTF-1 activation by different metals. We used the biomarker to screen a library of microarray profiles from cells treated with ~2600 organic chemicals to identify novel activators of MTF-1. All chemicals selected for validation were found to exhibit the properties of MTF-1 activators in MCF-7 cells.

Experimental

Use of a gene expression microarray experiment compendium. As described previously 21, data from BaseSpace Correlation Engine (BSCE) (<https://www.illumina.com/products/by-type/informatics-products/basespace-correlationengine.html>; formally NextBio) were used to build an annotated spreadsheet of statistically filtered gene expression comparisons (biosets) from experiments carried out using human cell lines and tissues. This compendium included bioset name, cell line, tissue, chemical name, chemical concentration, treatment time, and study ID. Of the biosets included, 11,725 biosets were derived from experiments in which cells were treated with one or more chemicals. Methods used to derive the statistically filtered gene lists are described in detail in Kuperschmidt et al. 22.

Construction of the MTF-1 biomarker. Biosets from the BSCE database were selected to represent a variety of human cell lines and metal treatment conditions expected to activate MTF-1, including treatments with silver, arsenic, mercury, zinc, and nickel. From these biosets, ten were selected whose fold-changes were consistent across treatments and included genes which exhibited opposite regulation to those in which MTF-1 was knocked down using siRNA 23. Raw expression microarray data for each of the ten biosets was downloaded from the NCBI GEO database and processed using Partek Genomics Suite to generate gene lists with expression fold-change values for each bioset. Gene lists were filtered to include only genes

which had an absolute fold change magnitude of ≥ 1.2 and $p < 0.05$ derived using a one-way ANOVA. The gene lists were uploaded to BSCE and the Meta-Analysis function was used to identify genes with the greatest degree of overlap between the 10 biosets. To construct the biomarker, genes were selected that were differentially expressed in the same direction in at least 5 of the 10 biosets and had an average fold change across those biosets which showed significant expression of at least 1.5 in either direction. Genes in the biomarker that overlapped with those in the Nrf2 biomarker (SLC7A11, GCLM, ADM, SRXN1, TXNRD1, SLC3A2, LARP6)²⁴ were excluded from the MTF-1 biomarker. The final biomarker consisted of 81 genes and average fold-change levels. The biomarker was uploaded to BSCE without any further filtering. Comparison of the biomarker to microarray compendium biosets. The MTF-1 biomarker was uploaded to BSCE and compared with all biosets in the database using the Running Fisher test. This method provides an assessment of the statistical significance of the correlation of the overlapping genes between the biomarker and each bioset providing a summary p-value. A complete description of the Running Fisher test is provided in 22. The results were exported and each p-value was converted to a $-\log(p\text{-value})$, with negative values used to indicate negative correlation between the biomarker and the bioset. Biosets with $-\log(p\text{-value}) \geq 4$ or ≤ -4 were considered significant based on prior studies using this threshold 20, 25. A column in the human gene expression spreadsheet was populated with the $-\log(p\text{-value})$ for each bioset. Biosets that were positively correlated with the biomarker were predicted to activate MTF-1, either directly or indirectly, such as by inducing oxidative stress or increasing labile cellular Zn; biosets that were negatively correlated were predicted to suppress MTF-1. The minor number of biosets identified here that exhibited negative correlation were not considered further.

Selection of positive and negative controls and calculation of biomarker accuracy. In the database of human gene expression comparisons, 57 biosets were selected as true positives as they were expected to activate MTF-1, based on the fact that they came from human cells treated with known activators of MTF-1 such as zinc, arsenic, iron, gold, or silver. The ten biosets used to create the biomarker (the training set) were not included in this list. Negative controls were identified using data from the EPA ToxCast program. In the EPA ToxCast online dashboard (<https://comptox.epa.gov/dashboard>), the ATG_MRE_CIS_up assay (carried out by contract with the EPA by Attagene, Inc) was selected, and all chemical data was downloaded. Tested chemicals that were examined in both the HTS assay and the human microarray compendium were identified. This list was filtered to include only 1) chemicals with no activity in the ATG_MRE_CIS_up assay and 2) microarray biosets from chemical treatments that used 75-100% of the maximum concentration tested in the ToxCast assay. Using these criteria, there were 23 chemicals classified

as true negatives. The values for predictive accuracy were calculated as follows: sensitivity (true positive rate) = True Positive (TP)/(TP+False Negative (FN)); specificity (true negative rate) = True Negative (TN)/(False Positive (FP)+TN); positive predictive value (PPV) = TP/(TP+FP); negative predictive value (NPV) = TN/(TN+FN); balanced accuracy = (sensitivity+specificity)/2.

Virtual screen for potential MTF-1 activators using the biomarker. Biosets were ranked by $-\log(p\text{-value})$, and a list was generated of the top organic chemicals predicted to activate MTF-1 (Supplemental File 1). Chemicals were excluded if they were not activated in the majority (more than half) of available biosets in which the chemical was tested in human cells. Of the top predicted MTF-1-activating chemicals remaining, nine commercially-available chemicals were selected for in vitro validation in our study.

Ingenuity pathway analysis. The MTF-1 biomarker genes were analyzed using the canonical pathway and upstream analysis functions of Ingenuity Pathway Analysis (IPA, Qiagen Bioinformatics, Redwood City, California). IPA calculates significance using a right-tailed Fisher's Exact test. The p-value is the probability of the overlap between the MTF-1 biomarker gene list and the IPA pathway gene list. Upstream analysis used the number of differentially expressed genes to predict upstream regulators of the biomarker genes.

Creation of a MTF1 knockout cell pool. The Knockout Cell Pool, MTF-1 Knockout MCF7, was created by Synthego, Inc. The MCF-7 cells were engineered by CRISPR-Cas9 to knock out the MTF-1 gene. The guide RNA was designed to bind to the antisense strand, in exon 2, of MTF1 gene to create a double stranded break at location 37,857,591 on chromosome 1. The knockout efficiency was determined by Synthego's Inference of CRISPR Edits (ICE) analysis; 96% of cell pool exhibited a disruption of the MTF-1 open reading frame.

Culture and treatment of wild-type and MTF-1-null MCF-7 cells. Chemicals were purchased from Sigma-Aldrich. All chemicals were $\geq 95\%$ purity except for thiostrepton and sulforaphane (90%). All chemical stock solutions were made in DMSO. MCF-7 cells were cultured in DMEM media (GIBCO) supplemented with 10% FBS (Omega Scientific, Australia) and 1x penicillin/streptomycin/glutamine. An initial range finding experiment (data not shown) was conducted to determine the optimal concentration of the tested chemicals. Based on these preliminary results, the highest concentration that was clearly non-cytotoxic was chosen as the test concentration for each chemical. Cells were plated at 15×10^5 cells per well in 12-well plates. After 48 hours, media was replaced with dosing solutions containing DMSO (0.05%) or 1-100 μM of each chemical (final concentration of DMSO = 0.05%). After 6 hours of exposure, media was removed; cells were lysed in 0.5 mL Trizol, followed by RNA extraction.

Evaluation of gene expression by RT-qPCR. Expression of selected genes were quantified using reverse transcription quantitative PCR (RT-qPCR) analysis. Briefly, $1\mu\text{g}$ of total RNA

was reverse transcribed with the SensiFAST™ cDNA Synthesis Kit (Bioline, London, UK). The primers were designed with Primer3 and listed in Supplemental Table 1. Real-time PCR was conducted with Sso Advanced Universal SYBR Green Supermix (Bio-Rad) in 384-well format. Expression of genes was calculated by the $2^{-\Delta\Delta Ct}$ method and normalized with β -actin of the same sample, and the relative transcript levels were expressed as percentage of control.

Evaluation of gene expression using Tempo-Seq. Gene expression in wild-type and MTF1-null cells were evaluated for gene expression changes after chemical exposure in a focused study using the human 1500+ Tempo-Seq platform (BioSpyder, Inc, Carlsbad, CA). After RNA extraction, samples were sent to BioSpyder for analysis. Raw read counts were normalized and statistically-filtered gene lists (p -value < 0.05 with no multiple test correction) were generated using the DESeq2 module in Partek Flow. The data is publicly available at Gene Expression Omnibus, accession number GSE152703.

Measurement of oxidative stress. MCF-7 cells were treated with 10 μ M of the reactive oxygen species probe 2', 7'-dichlorofluorescein diacetate (DCFDA, Sigma-Aldrich) or no probe in Hank's balanced salt solution (HBSS) for 30 min. This solution was removed and cells were washed twice before treatment with each compound or DMSO control for two hours. Fluorescence was measured (excitation 490 nm, emission 530 nm) with a SpectraMax i3X, Molecular Devices, San Jose, CA plate reader. The fluorescence for the no-probe controls was subtracted from the fluorescence of DCFDA conditions and normalized to the DMSO condition. The data was subjected to ANOVA analysis, followed by Duncan's multiple range tests with significance criteria of $p < 0.05$.

Measurement of free zinc. MCF-7 cells were treated with each compound or DMSO control in DMEM and incubated for one hour. Media was removed and cells were washed twice with HBSS. New media containing 5 μ M of the fluorescent zinc probe Zinpyr-1 (Santa Cruz Biotechnology) or no probe was added and cells were incubated for 30 minutes. Media was removed and cells were washed twice again before reading fluorescence (excitation 490 nm, emission 530 nm) with a SpectraMax i3X, Molecular Devices, San Jose, CA plate reader. The fluorescence for the no-probe controls was subtracted from the fluorescence of Zinpyr-1 conditions and normalized to the DMSO condition. The analysis was conducted on 4-5 independent experiments. The data was subjected to ANOVA analysis, followed by multiple range tests with significance criteria of $p < 0.05$.

RT-qPCR analysis of MTF-1 biomarker genes. Wild-type MCF-7 cells and MTF-1 null cells were treated with 10 selected MTF-1 activators at concentrations indicated in Figure 6 for 6 hr in 3-4 independent experiments. Total RNA was isolated in Trizol, reverse transcribed (Bioline, UK), and subjected to qPCR analysis with SYBR Green master mix (BioRad, CA) on Bio-Rad CFX real-time PCR system. The primers were designed with Primer3 (v 4.0). The Ct values were used to calculate the

relative expression by the $2^{-\Delta\Delta Ct}$ method and normalized with β -actin, setting WT control as 100%. The Student t-test was used to compare the difference between WT and MTF-1 null cells, with significance criteria of $p < 0.05$. Preparation of chemicals for clustering and association rule mining. Chemicals from the CMAP 2.0 collection of chemicals with defined structural information and observed MTF1 activity in at least 2 of the 3 cell lines were used to perform the clustering and association rule mining algorithms. Structures in the form of simplified molecular-input line entry system (SMILES) notation were taken from the EPA CompTox Chemicals dashboard (<https://comptox.epa.gov/dashboard/>, accessed November 19th 2018) 26. Substances which did not meet the condition of activating MTF1 in 2 or 3 of the 3 cell lines were removed from the cheminformatic analysis. A chemical fingerprint file (a matrix of binary values [1,0] for each chemical-feature pair) was derived using the publicly available ToxPrint feature set (<https://toxprint.org/>) generated within the associated Chemotyper application (<https://chemotyper.org/> accessed November 19th 2018), developed by Altamira [Altamira, Columbus, OH] and Molecular Networks [Molecular Networks, Erlangen, GmbH] under contract from FDA. The ToxPrint set (V2.0_r711) consists of 729 uniquely defined features designed to provide broad feature coverage of inventories consisting of tens of thousands of environmental and industrial chemicals, including pesticides, cosmetics ingredients, food additives, and drugs 27.

Chemical clustering. To generate chemical clusters, the ChemmineR package (v.3.32.1) was used to calculate a distance matrix of the Tanimoto similarity index between each pair of ToxPrint fingerprints and, subsequently, chemicals were binned together using the single-linkage (nearest neighbor) clustering function 28. A range of different similarity thresholds from 0.6 (60%) to 0.9 (90%) in 5% increments were used to cluster the chemicals. The final threshold was chosen based upon a visual inspection of the clusters generated by each similarity threshold.

Identification of association rules. Association rule mining (ARM), as implemented within the arules package in R 29, 30, was employed to investigate whether certain (sets of) ToxPrints were associated with a chemical having a positive designation. It should be noted that if a chemical was tested at two different concentrations and given a different designation at each concentration, only the concentration with the positive designation was retained. Briefly, ARM is a data mining method that can identify highly correlated (groups of) variables and has been utilized in a variety of application domains including bioinformatics, and cheminformatics 31-37. The ARM workflow comprises two main steps: frequent itemset mining (FIM) and generation of rules. More information on FIM and ARM can be found in studies by Agrawal, Bagelt and colleagues 38, 29, 39.

The first step of FIM, aims to identify items in a dataset that frequently co-occur. Within FIM, there are 3 components

to consider; a set of items I , where $I = \{i_1, i_2, i_3, \dots, i_n\}$, a set of transactions T , where $T = \{t_1, t_2, t_3, \dots, t_m\}$ and a transaction t which contains a subset of the items in I (also called an itemset). In this analysis, the individual ToxPrints and the MTF-1 activity call (positive or negative) represent the items i , whereas a given chemical represents a transaction t .

Once identified, frequent itemsets can then be examined and expressed as a collection of probabilistic if/then statements called association rules. They take the form of $X \rightarrow Y$ where X and Y are itemsets. Each association rule comprises two parts: 1) the “if” or antecedent (i.e. the left-hand side, LHS) and 2) the “then” or consequent (i.e. the right-hand side, RHS). The basic premise behind the generation of association rules is splitting a frequent itemset X into two disjoint subsets X and Y such that items present in the LHS are not present in the RHS (i.e. $X \cap Y = \emptyset$). In this study, the ToxPrints served as the antecedent, whereas the MTF1 activity call served as the consequent. It should be noted that for the association rules generated the antecedent could be a set of up to three separate ToxPrints. Additionally, we filtered for only those rules containing a positive designation as the consequent. Several interest measures were calculated to help identify a pragmatic and practical number of frequent itemsets and their resulting association rules for use. These measures themselves are defined as follows:

Support: The support of an association rule ($X \rightarrow Y$) describes the proportion of transactions that contain all items in the rule (i.e. the union of X and Y).

$$\text{support}(X \rightarrow Y) = t(X \cup Y) / T$$

Confidence: The confidence of a rule meanwhile describes the proportion of all transactions containing X that also contain Y . Alternatively, confidence is the conditional probability of the transaction containing Y given the presence of X .

$$\text{confidence}(X \rightarrow Y) = \text{support}(X \cup Y) / (\text{support}(X))$$

Lift: The lift of a rule describes the factor by which the observed co-occurrence of X and Y exceeds what would be expected if X and Y were independent.

$$\text{lift}(X \rightarrow Y) = \text{support}(X \cup Y) / (\text{support}(X) \cdot \text{support}(Y))$$

Odds ratio: The odds ratio of a rule describes the odds of X occurring given the presence of Y compared against X occurring in the absence of Y .

$$\text{odds ratio}(X \rightarrow Y) = (\text{support}(X \cup Y) / (\text{support}(X) - \text{support}(X \cup Y))) / ((\text{support}(Y) - \text{support}(X \cup Y)) / (1 - \text{support}(X \cup Y)))$$

A combination of the support and confidence interest measures were used to limit the number of association rules generated, specifically, a rule had to have a minimum support of 4 (i.e., ToxPrints and positive activity outcome needed to appear in at least 4 chemicals within the dataset) and a minimum confidence of 75% (i.e., the ToxPrints needed to be associated with a positive activity call at least 75% of the time). The derived rules were subsequently applied to the set of 284 unique chemicals from the list of CMAP chemicals where 1) no overall designation had been determined in the MTF1 assays,

and 2) structural information (in terms of SMILES notation) were present in the EPA’s CompTox Chemicals Dashboard (<https://www.epa.gov/chemical-research/comptox-chemicals-dashboard>; accessed March 17, 2020). This was carried out to identify other chemicals that contain the structural features present in at least one of the rules derived from the ARM algorithm.

Inventory comparison. As the ToxPrint chemotypes are a predefined set of structural features that cover a broad cross section of the available chemical space, they were used to compare the structural diversity of the chemicals studied within this analysis to those tested in the ATG_MRE_CIS_up assay as part of the ToxCast program. The comparison would provide an indication of whether the rules derived were generalizable. To do this the ToxPrint fingerprints were generated for those chemicals tested in the ATG_MRE_CIS_up assay. Subsequently, the fingerprints for both sets of chemicals were folded to reduce their length from 729 features to 70 features so the comparison could be more easily visualized. This folding took advantage of the hierarchical nature of the ToxPrints such that more specific ToxPrints were collapsed into their more generalised form. For example, the “bond:C#N_cyano_acylcyanide”, “bond:C#N_cyano_cyanamide”, and “bond:C#N_nitrile_ab-unsaturated” chemotypes would all fall under the more general “bond:C#N” chemotype. The relative quantities of each of these more general chemotypes were then calculated and compared between the two data sets.

Results and Discussion

Construction of an MTF-1 biomarker. A set of biomarker genes was identified to predict modulation of MTF-1 activity. Genes were identified from ten microarray comparisons (biosets) representing diverse metal treatment conditions and human cell types (Table 1). Raw gene expression data from the studies was used to generate the gene lists using Partek Genomics Suite. These statistically-filtered gene lists were compared to the corresponding gene lists found in BSCE (Figure S1); strong correlations were found between the two computationally-derived bioset gene lists. The Partek-generated gene lists were used to identify biomarker genes. Genes that exhibited relatively robust and consistent expression changes across the ten biosets were selected as candidate biomarker genes (absolute average fold-change ≥ 1.5 -fold in either direction). Given that metal exposure can lead to oxidative stress that activates Nrf2, we filtered out any genes that overlapped with a previously characterized biomarker for Nrf2. The behavior of the resulting 81 genes across the 10 biosets is shown in Figure 1A. The gene list and fold-change values were used as the MTF-1 biomarker. The biomarker consisted of 48 upregulated genes and 33 downregulated genes. The genes with the highest fold-change values included seven MT family members (MT1E, MT1F, MT1G, MT1H, MT1M, MT1X, MT2A) and the zinc transporter

1
2
3 SLC30A1, all well known to be regulated by MTF-1 11. Several
4 of the other top ranked genes in the biomarker are not known
5 MTF-1 targets but have indirect connections to the oxidative
6 stress response or metal detoxification and homeostasis. For
7 example, a top ranked gene in the biomarker was AKR1C2
8 (aldo-keto reductase family 1 member C2), which has known
9 antioxidant effects in cancer cells 40. Other highly ranked
10 genes in the biomarker were ATF3 (activating transcription
11 factor 3) and CLU (clusterin), known to be activated under
12 conditions of oxidative stress 41-43.

13 To identify pathways enriched in biomarker genes, the
14 canonical pathway and upstream analysis functions in
15 Ingenuity Pathway Analysis were used (Figure 1B). The analysis
16 identified the Nrf2-mediated oxidative stress response
17 pathway and the iron homeostasis signaling pathways as the
18 top overlapping pathways, with 3.5% ($p = 9 \times 10^{-6}$) and 2.9% ($p = 2 \times 10^{-3}$)
19 of the genes overlapping, respectively. Several of the
20 top upstream regulators predicted by the analysis were
21 transcription factors and included MTF-1 ($p = 6.0 \times 10^{-6}$).

22 Comparison of the biomarker to individual biosets used to
23 make the biomarker. BSCE was used to determine the
24 correlation between the biomarker and each of the ten Partek-
25 generated biosets used in the construction of the biomarker.
26 BSCE uses the rank-based Running Fisher algorithm to
27 calculate the statistical significance of each pair-wise
28 correlation. Statistically significant correlations were defined
29 as those with a $-\log(p\text{-value}) \geq 4$ or ≤ -4 , where negative values
30 indicate a negative correlation. Chemical treatment conditions
31 that resulted in positively correlated biosets were predicted to
32 be activators of MTF-1, while those conditions that resulted in
33 negatively correlated biosets were predicted to be suppressors
34 of MTF-1. (Suppressors of MTF-1 were not examined in this
35 study.) Of the ten biosets used to create the biomarker, nine
36 were found to be significantly correlated with the biomarker
37 (Figure 1C). The one bioset not significantly correlated ($-\log(p\text{-value}) = 2.57$)
38 had only two significantly altered genes
39 overlapping with the biomarker. We also compared the
40 biomarker to the same 10 BSCE-derived biosets in the human
41 compendium. Figure S2 shows that the biomarker correlations
42 were similar independent of the method used to derive the
43 bioset, although most of the biosets generated by the Partek
44 approach had higher $-\log(p\text{-value})$ s.

45 MTF-1 dependence of biomarker genes. To determine if
46 any of the biomarker genes exhibited MTF-1 dependence in
47 their expression profiles, we examined expression of the genes
48 after MTF-1 siRNA knockdown from a recent microarray study
49 using colorectal cancer cells 23. For this analysis, the data from
50 this bioset was processed in Partek to generate a gene list
51 using a fold-change cutoff of 1.2 in either direction and a
52 significance cutoff of $p = 0.05$. There were 20 significantly
53 altered genes that overlapped between the biomarker and a
54 comparison in which MTF-1 knockdown was compared to cells
55 treated with control siRNA in which both groups were treated
56 with 150 μM zinc. Out of 20 genes, 9 of the biomarker
57
58
59
60

upregulated genes exhibited decreased expression by MTF-1
siRNA (Figure 1D), and 5 of the biomarker downregulated
genes exhibited increased expression by MTF-1 siRNA. This
bioset was significantly negatively correlated with the
biomarker ($-\log(p\text{-value}) = -4.68$). These results indicated that
at least a subset of the biomarker genes were dependent on
MTF-1 for metal-induced expression. We also show MTF-1-
dependence of the biomarker gene responses in experiments
described below.

Correlation between MTF-1 and Nrf2 activation. Because
metal toxicity is known to induce oxidative stress, we
examined the relationship between activation of MTF-1 and
Nrf2 by comparing biomarker responses across the biosets in
the human compendium. Figure S3 compares the $-\log(p\text{-value})$ s for each of the 11,725 biosets for both biomarkers. The
correlation coefficient was calculated to be $r = 0.433$ with $p < 10^{-300}$, indicating that predicted activation of the two
transcription factors is correlated in a positive manner. This
correlation does not change meaningfully when the biosets
were filtered for time (e.g. $r = 0.432$, > 6 h treatment) or cell
line (e.g. $r = 0.408$, > 6 h treatment, MCF7 cells). This analysis
indicates a tight relationship between activation of MTF1 and
Nrf2. The inclusion of a number of stress response genes in the
MTF-1 biomarker including those that may not be directly
regulated by MTF-1 indicates that the biomarker represents
the gene expression profile of a general cellular response to
metal-induced stress, not just the direct induction of metal
regulatory genes by MTF-1. Using this biomarker as an
indicator of the metal stress response may therefore be more
useful as a measure of systems changes due to metal-induced
toxicity than simply measuring changes in MT gene expression.

Predictive accuracy of the biomarker. In the compendium
of human gene expression biosets, there were many examples
of cells treated with metals under conditions expected to
activate MTF-1 (excluding those used to make the biomarker).
We randomly selected 57 to use as those that activate MTF-1.
Using the Running Fisher test, correlation between each of
these biosets and the biomarker was determined and 52 of the
57 were predicted to activate MTF-1 resulting in a sensitivity of
91%. Results from the EPA's ToxCast chemical screening
program (MTF-1 activation assay run by Attagene, Inc.) were
used to identify 9 chemicals represented by 23 biosets in the
database which were classified as true negatives for MTF-1
activation. Of these, only one of the 23 was predicted by the
biomarker to activate MTF-1, demonstrating a specificity of
96%. Combining these results, the biomarker gave a balanced
accuracy of 93%. Figure 2 shows the distribution of the $-\log(p\text{-value})$ s for the true positives and true negatives,
demonstrating that the biomarker reliably identifies MTF-1
activators from non-activators.

Virtual screen for chemical activators of MTF-1. Given the
high predictive accuracy of the biomarker, we performed an
in-silico screen to identify novel chemicals that activate MTF-1.
In the human compendium, there were 11,725 biosets

representing human cell lines exposed to 2582 chemicals. Excluding the biosets used to create the biomarker, the analysis identified 1547 biosets representing 700 chemicals that were positively correlated and 122 biosets representing 90 chemicals that were negatively correlated with the MTF-1 biomarker. The remaining 10,066 biosets were not significantly correlated with the biomarker in either direction. The statistically filtered fold-change values of the biomarker genes for all 11,725 biosets ranked by $-\log(p\text{-value})$ are shown in Figure 3, top, and the $-\log(p\text{-value})$ s for individual biosets are shown in Figure 3, bottom. The top most correlated chemical treatments are shown. For those biosets that were significantly positively correlated with the biomarker, the expression of the biomarker genes was remarkably similar to the fold-change values of the biomarker genes. For those biosets that were significantly negatively correlated, the expression of the biomarker genes exhibited opposite expression compared to the biomarker genes.

Characterization of putative MTF-1 activators. In addition to the chemicals used in the prediction analysis, there were additional metal treatments in the database. We compared those to the biomarker to examine which metal treatment conditions were correlated with the MTF-1 biomarker. There were a total of 89 metal treatment biosets identified (including the 58 true positives) discussed above, representing ten different metals. For this analysis, all metal-containing compounds of the same metal were grouped together. The metals with the greatest number of database biosets were zinc, silver, and arsenic. The MTF-1 biomarker predicted that the majority of zinc treatments (87.1%), silver treatments (95.7%), and arsenic treatments (100%) activated MTF-1 (Table 2). All other metals in the database either have very few representative biosets in the database, or they are most often not predicted to activate MTF-1. Most surprisingly, cadmium, which is represented by seven biosets and known to activate MTF-1, was only predicted to activate MTF-1 in two of seven biosets. However, the majority of the cadmium biosets not predicted to activate MTF-1 were derived from one study using one cell line. The lack of MTF-1 activation in these experiments may be particular to this cell line or due to other experimental conditions used in this study. Comparison of the biomarker to a larger and more diverse sample of metal treatment experiments would be necessary to confidently assess the differences in MTF-1 activation by different metals as predicted by the biomarker.

Identification of structural features associated with MTF-1 activation. We next asked whether there were any structural features in the organic compounds associated with MTF-1 activation. (Metals were not considered in this analysis.) To facilitate the analysis, we focused on a set of 616 unique CMAP chemicals that had been examined in three cell lines (HL60, MCF-7, PC3), had an overall activity call for MTF-1 consistent across the cell lines, and for which chemical structures could be readily identified. There was activation of MTF-1 by 105

unique chemicals in two or more cell lines, and of these there were 81 in which chemical structures could be identified. Two approaches were undertaken to address whether there were structural features associated with MTF-1 activation: 1) clustering based upon structural similarity, enabling assessment of how similar the chemicals were to one another and 2) association rule mining to specifically examine which sets of structural features discriminated for MTF-1 activation.

ChemmineR's agglomerative clustering using a Tanimoto threshold of 80% was found to result in reasonable clusters of all 616 active and inactive chemicals in terms of total number of clusters, their membership size, and chemistry makeup. At this similarity threshold, 21 clusters were extracted that had at least 3 or more members. Six clusters contained at least 1 active chemical whereas five of these six clusters contained multiple active chemicals. The 5 clusters with multiple active chemicals covered 32 unique chemicals, of which 18 chemicals were active. The remaining active chemicals were captured in "clusters" of either 1 or 2 chemicals. Figure 4A represents a dendrogram of the clustering highlighting two specific clusters (cluster ID 25 and 125) for illustrative purposes. The full set of clusters and their membership are provided in Supplemental File 1. Figure 4B shows the consistency in the structures identified in clusters 25 and 125.

Utilizing the ToxPrint fingerprints for the 616 unique chemicals and their corresponding activity calls (i.e., positive or negative), the apriori algorithm identified 235 rules associated with a chemical being given an overall activity call of "positive". The majority of these rules identified the same or similar sets of chemicals. These rules can be grouped into 4 larger bins. The reason behind this grouping is the hierarchical nature of the ToxPrints features themselves. For example, if a chemical contains one of the more specific ToxPrint features, then its molecular fingerprint will contain an "on" bit for both this more specific ToxPrint feature and the corresponding more general ToxPrint feature. Thus, both the more specific and more general ToxPrint features may be used in the generation of association rules, provided they meet the inclusion criteria (i.e., a minimum support of 4 and minimum confidence of 75%). All of the association rules generated by the apriori algorithm are available in Supplemental File 1. Figure 5 illustrates one of the association rules generated in this study. A description of the support and confidence of this and other rules are found in Supplemental File 1. The association rules predicted 70 unique chemicals out of the 284 chemicals from the CMAP set with no activity designation (either positive or negative) as MTF-1 actives.

The structural diversity of the chemicals tested within this study were compared to those chemicals tested in the ATG_MRE_CIS_up assay. Figure S4 shows the chemicals in this study contain a larger proportion of chemicals with aromatic and hetero ring systems. Additionally, the chemicals in this study contain a larger proportion of chemicals with amine, alcohol, ether, and amide bonds. However, the

ATG_MRE_CIS_up data set contained more metal containing chemicals; although, it is likely that the majority of these metal-containing chemicals are salts rather than organometallics.

The initial clustering approach used in this study could identify the closest chemical neighbors; thereby, facilitating predictive inferences between chemicals within a given cluster. This approach could also be expanded to categorizing novel chemicals and making inferences as to the likely (in)activity towards MTF-1 using a data gap filling approach. Additionally, we employed ARM to identify substructural features associated with chemicals being assigned a positive designation: a total of 253 rules were identified. These rules were applied prospectively to a set of 284 chemicals from the CMAP data set for which incomplete activity data was available. Confirmatory testing would help evaluate the performance of the rules inferred. To facilitate the generalization of the rules derived, the ToxPrint feature set was utilized which are a fixed set of defined features that are not specific to the dataset being analyzed. To explore the general ability of the rules, the set of chemicals of interest were compared to a larger inventory. The enrichments across the feature set was very similar between the CMAP chemicals and the chemicals tested in the ATG_MRE_CIS_up assay suggesting that the rules should be broadly applicable.

Confirmation of the MTF-1 activating chemicals. Nine organic chemicals were selected for testing of MTF-1 activation in vitro (Table 3). Most of these came from the CMAP 2.0 dataset (GSE5258) in which ~1300 mostly pharmaceutical chemicals were examined in three cell lines 44; they had some of the highest activation scores of the putative positive chemicals. Figure S5A shows the average $-\log(p\text{-value})$ s for significantly correlated biosets for each of these positive chemicals. The chemical structures of the chemicals are shown in Figure S5B. Of these nine chemicals, two are known to interact with cellular metals. Clioquinol is an antibiotic that is known to act as a zinc and copper ionophore and has been shown to have metal-dependent activity against cancer and Alzheimer's disease 45–47, and disulfiram is an alcohol aversion therapeutic that has also been shown to act as a copper and zinc ionophore with anticancer activity 48, 4948, 4948–50. The other seven chemicals, however, have no known direct relationships to metal homeostasis. Some of these chemicals, such as mefloquine 51, thioestrepton 52, and deoxycholic acid 53, are known to induce oxidative stress, which may explain why they activate MTF-1.

The ability of the nine chemicals to regulate MTF-1-dependent genes in MCF-7 cells was tested by treating MCF7 wild-type and MTF-1-null cells with each chemical and measuring gene expression by RT-qPCR. We chose four genes from the MTF-1 biomarker to measure (MT1M, MT2A, HMOX1, ATF3). Zinc activated three of the genes in a MTF-1-dependent manner. Exposure to all of the chemicals led to increases in expression of one or more of the four genes

(Figure 6). For all but two of the chemicals (chloroquine, phenoxybenzamine), the increases were abolished or decreased in MTF-1-null cells compared to wild-type cells demonstrating dependence of the increases on MTF-1. While ATF3 appeared to be regulated by the majority of metals in the microarray studies, there may be cell line differences in responses between those used to create the biomarker and MCF-7 cells used for validation.

Assessment of changes in intracellular zinc and oxidative stress by the nine chemicals. Having confirmed that most of the chemicals do activate MTF-1, we sought to determine the mechanism by which MTF-1 is being activated. Possible mechanisms of activation include induction of zinc influx, generation of reactive oxygen species that induce release of zinc from MTs, or direct binding to MTF-1 or another protein necessary for MTF-1 transcriptional activation. We used fluorescent probes Zinpyr-1 and 2',7'-dichlorofluorescein diacetate (DCFDA) to detect intracellular amounts of labile zinc and reactive oxygen species (ROS), respectively. Our results indicate that in addition to Zn itself six of the nine chemicals cause an increase in cellular labile zinc (Figure 7A). The positive control tert-butyl hydroperoxide (TBHP) caused increases in ROS production (Figure 7B). Two the nine chemicals caused increases in ROS while one chemical decreased ROS. Overall, the results suggest that increased labile zinc may play an important role in the mechanism of activation of MTF-1 by many of the predicted MTF-1-activating chemicals.

Transcript profiling of chemicals in wild-type and MTF-1-null cells. In future HTr studies, we propose that the use of cell lines nullizygous for important targets of environmental chemicals will greatly facilitate the interpretation of regulated genes. To provide proof of this concept, we generated transcript profiles of Zn and 6 putative activators of MTF-1 in wild-type and MTF-1-null cells generated by TempO-Seq targeted sequencing of ~3000 human genes. In the absence of chemical exposure, knockout of MTF-1 resulted in effects on 44 genes (Figure S6) indicating that MTF-1 has some effect on gene expression in the absence of overt activation. Figure 8A shows the heat maps of the filtered changes altered by each chemical in the two cell lines. Figure 8B shows the number of genes altered by each chemical in wild-type cells and the number of those genes that are altered in the MTF-1-null cells. Almost all of the genes regulated by Zn and phenoxybenzamine were no longer regulated in the MTF-1-null cells. For the rest of the chemicals, more than half of the genes regulated in the wild-type cells were no longer regulated in the MTF-1-null cells.

The profiles derived from the wild-type cells were each compared to those found in the CMAP 2.0 collection to determine if the same chemical would exhibit high correlation despite differences in transcriptomic platform and number of genes evaluated. Each profile was compared to 3442 biosets in the CMAP collection. The chemicals ranked highest to lowest

to the same chemical examined in MCF-7 cells were astemizole and disulfiram (both 1st), suloctidil (29th), alexidine (41st) and phenoxybenzamine (816th). Zn and deoxycholic acid were not examined in the CMAP collection. The top hit for Zn was the comparison "Lung epithelial BEAS-2B cells + 3uM zinc sulfate heptahydrate for 4hr_vs_24hr untreated" from GSE80733 (p-value = 3.3E-12). The top hit for deoxycholic acid was the comparison "Hepatic progenitor HepaRG cells differentiated 96hr - 200037.13nM chenodeoxycholic acid_vs_DMSO" (p-value = 8.9E-42) examined on the 1500+ platform 54. Except for phenoxybenzamine, these rankings indicate that the 1500+ gene set allows for a high correlation to the same chemical within an independent dataset.

These studies demonstrate that many of the changes in gene expression after exposure to the putative MTF-1 activators were MTF-1-dependent. We also found that the chemical-induced profiles generated in wild-type cells exhibited correlation to the same chemicals profiled in various cell lines in our compendium. These studies demonstrate that wild-type vs. null cell line comparisons can unequivocally identify targets of environmental chemicals. To our knowledge, this is the first study to compare profiles of chemicals between wild-type and nullizygous cell lines. Although the study was limited in scope, it provides support for the use of nullizygous cell lines in HTTr screening, especially in targeted screening to confirm predicted targets from primary HTTr screens.

Summary

Gene expression biomarkers have been shown to be useful for identifying chemical modulators of transcription factors in previous work by our group 19, 20. In the present study, we used a similar approach to identify activators of MTF-1 in a compendium of gene expression profiles derived from chemically-treated human cell lines. The biomarker was built by identifying genes with consistent expression behavior across comparisons from metal-treated cells under conditions known to activate MTF-1. The resulting biomarker consisted of 48 up-regulated and 33 down-regulated genes. The genes in the biomarker with the highest fold-change values included seven MT family members and SLC30A1 encoding the zinc transporter ZnT1 that are well-established targets of MTF-1-mediated gene regulation 11. A subset of the biomarker genes was shown to be MTF-1-dependent based on a previously published MTF-1 siRNA knockdown experiment and transcript profile comparisons between wild-type and MTF-1-null cell lines. We tested the biomarker for predictive accuracy using a set of biosets with known MTF-1 responses including a number of metal activators of MTF-1 and found that the biomarker was very accurate (balanced accuracy = 93%). The biomarker was used to identify chemical activators of MTF-1 in our human

microarray compendium. We identified 700 chemicals in 1547 biosets that were putative activators, and they included a large number of metals, metal complexes, and chemicals known to modulate metal homeostasis. Nine of the positive chemicals were examined further and we found that in RT-qPCR and transcript profile experiments most of the nine of the chemicals were confirmed to be MTF-1 activators. We performed a chemoinformatic analysis to identify structural features associated with MTF-1 activity and used these features to reprofile chemicals with an equivocal result to demonstrate their utility in chemical prioritization. The results indicate that our methods can readily identify MTF-1 activators. Our approach will not only be useful in analyzing HTTr data of environmental chemicals in ToxCast screens to identify those that disrupt metal homeostasis but will allow identification of environmental samples with contamination by metals.

We envision that batteries of validated biomarkers (including the the MTF-1 biomarker described here) with known accuracies and context of use⁵⁵ could be systematically applied to HTTr data to identify the modulation of tens if not hundreds of factors important in chemical toxicity. There are a number of examples in which characterized biomarkers have been used to screen large sets of chemicals for those that modulate the factor similar to the screening strategy used in this study. A set of 1152 chemicals were screened for androgen receptor (AR) agonism/antagonism in AR+ human prostate cancer cell lines 19. A set of ~1950 chemicals were screened for those that potentially cause DNA damage using the TGx-DDI biomarker identifying ~10 novel compounds not previously known to cause DNA damage 56. The same set of chemicals were screened using a biomarker for Nrf2 and identified both known and novel activators 24. Finally, an estrogen receptor (ER) biomarker was characterized that could be used to screen chemicals in ER+ human breast cancer cell lines 20. Prioritization of identified chemicals for further analysis could then be carried out using the network of adverse outcome pathways (AOP)s as a starting point 57, 58, especially if strategies are in place that stratify the AOPs in terms of human relevance, number of KEs perturbed, and overlap with AOPs of special concern including developmental or reproductive toxicity. This approach would then allow targeted follow-up studies for confirmation of phenotypic perturbation in lower throughput organotypic-based assays, zebrafish, or short-term rodent studies. Benchmark dose analysis will be used to identify transcriptional points of departure for the genes in the biomarkers to estimate minimal concentrations of chemical that lead to activation or suppression of the factor similar to methods carried out to determine transcriptional points of departure for pathways 59. These concentration values could then be used for in vitro to in vivo extrapolation (IVIVE) to predict tissue concentrations of the chemical in humans followed by use of exposure models to

determine the margin between estimated human exposures and those that induce responses in cell-based assays 60. These efforts would allow a preliminary assessment of “safe” levels of exposure 60. Our efforts to characterize a biomarker that predicts MTF-1 modulation contributes to a broader effort within EPA to comprehensively and systematically identify targets of chemicals 17 and to move towards reduction of the use of animals in toxicity testing 61.

The main text of the article should appear here with headings as appropriate.

Conflicts of interest

The authors have no conflicts of interest.

Acknowledgements

We wish to thank Drs James Samet and Keith Houck for critical review of this manuscript and Molly Windsor for assistance with creating the figures. Funding for this study came from the U.S. Environmental Protection Agency Office of Research and Development. A.C.J. acknowledges support from the NSF Graduate Research Fellowship Program and Graduate Research Internship Program (DGE 1644868). This work was carried out as part of the US-EPA Chemical Safety for Sustainability (CSS) research program 12.01: Adverse Outcome Pathways, Project 1.1g (Cancer AOPs).

Notes and references

References

1. A. Rolfs and M. A. Hediger, Metal ion transporters in mammals: structure, function and pathological implications, *The Journal of Physiology*, 1999, 518, 1-12.
2. T. O'Halloran, Transition metals in control of gene expression, *Science (New York, N.Y.)*, 1993, 261, 715-725.
3. S. Y. Choi and A. J. Bird, Zinc'ing sensibly: controlling zinc homeostasis at the transcriptional level, *Metallomics*, 2014, 6, 1198-1215.
4. M. Jaishankar, T. Tseten, N. Anbalagan, B. B. Mathew and K. N. Beeregowda, Toxicity, mechanism and health effects of some heavy metals, *Interdisciplinary Toxicology*, 2014, 7, 60-72.
5. Z. L. He, X. E. Yang and P. J. Stoffella, Trace elements in agroecosystems and impacts on the environment, *Journal of trace elements in medicine and biology : organ of the Society for Minerals and Trace Elements (GMS)*, 2005, 19, 125-140.
6. J. Ferguson, *The Heavy Elements*, Chemistry, Environmental Impact and Health Effects, 1990, Oxford, 211-212.
7. A. T. Miles, G. M. Hawksworth, J. H. Beattie and V. Rodilla, Induction, Regulation, Degradation, and Biological Significance of Mammalian Metallothioneins, *Critical Reviews in Biochemistry and Molecular Biology*, 2000, 35, 35-70.
8. V. Gunther, U. Lindert and W. Schaffner, The taste of heavy metals: gene regulation by MTF-1, *Biochimica et biophysica acta*, 2012, 1823, 1416-1425.
9. G. Dong, H. Chen, M. Qi, Y. Dou and Q. Wang, Balance between metallothionein and metal response element binding

transcription factor 1 is mediated by zinc ions (review), *Mol Med Rep*, 2015, 11, 1582-1586.

10. U. Lindert, M. Cramer, M. Meuli, O. Georgiev and W. Schaffner, Metal-Responsive Transcription Factor 1 (MTF-1) Activity Is Regulated by a Nonconventional Nuclear Localization Signal and a Metal-Responsive Transactivation Domain, *Molecular and cellular biology*, 2009, 29, 6283-6293.

11. J. E. Hardyman, J. Tyson, K. A. Jackson, C. Aldridge, S. J. Cockell, L. A. Wakeling, R. A. Valentine and D. Ford, Zinc sensing by metal-responsive transcription factor 1 (MTF1) controls metallothionein and ZnT1 expression to buffer the sensitivity of the transcriptome response to zinc, *Metallomics*, 2016, 8, 337-343.

12. B. Ruttkay-Nedecky, L. Nejdil, J. Gumulec, O. Zitka, M. Masarik, T. Eckschlager, M. Stiborova, V. Adam and R. Kizek, The Role of Metallothionein in Oxidative Stress, *International Journal of Molecular Sciences*, 2013, 14, 6044-6066.

13. W. Maret, Metallothionein/disulfide interactions, oxidative stress, and the mobilization of cellular zinc, *Neurochemistry International*, 1995, 27, 111-117.

14. W. Maret, Oxidative metal release from metallothionein via zinc-thiol/disulfide interchange, *Proceedings of the National Academy of Sciences*, 1994, 91, 237.

15. W. Wu, P. A. Bromberg and J. M. Samet, Zinc ions as effectors of environmental oxidative lung injury, *Free radical biology & medicine*, 2013, 65, 57-69.

16. R. Judson, K. Houck, M. Martin, T. Knudsen, R. S. Thomas, N. Sipes, I. Shah, J. Wambaugh and K. Crofton, In vitro and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast programme, *Basic & clinical pharmacology & toxicology*, 2014, 115, 69-76.

17. R. S. Thomas, T. Bahadori, T. J. Buckley, J. Cowden, C. Deisenroth, K. L. Dionisio, J. B. Frithsen, C. M. Grulke, M. R. Gwinn, J. A. Harrill, M. Higuchi, K. A. Houck, M. F. Hughes, E. S. Hunter, K. K. Isaacs, R. S. Judson, T. B. Knudsen, J. C. Lambert, M. Linnenbrink, T. M. Martin, S. R. Newton, S. Padilla, G. Patlewicz, K. Paul-Friedman, K. A. Phillips, A. M. Richard, R. Sams, T. J. Shafer, R. W. Setzer, I. Shah, J. E. Simmons, S. O. Simmons, A. Singh, J. R. Sobus, M. Strynar, A. Swank, R. Tornero-Valez, E. M. Ulrich, D. L. Villeneuve, J. F. Wambaugh, B. A. Wetmore and A. J. Williams, The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency, *Toxicological sciences : an official journal of the Society of Toxicology*, 2019, 169, 317-332.

18. J. M. Yeakley, P. J. Shepard, D. E. Goyena, H. C. VanSteenhouse, J. D. McComb and B. E. Seligmann, A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling, *PLoS one*, 2017, 12, e0178302.

19. J. P. Rooney, B. Chorley, N. Kleinstreuer and J. C. Corton, Identification of Androgen Receptor Modulators in a Prostate Cancer Cell Line Microarray Compendium, *Toxicological sciences : an official journal of the Society of Toxicology*, 2018, 166, 146-162.

20. N. Ryan, B. Chorley, R. R. Tice, R. Judson and J. C. Corton, Moving Toward Integrating Gene Expression Profiling Into High-Throughput Testing: A Gene Expression Biomarker Accurately Predicts Estrogen Receptor alpha Modulation in a Microarray Compendium, *Toxicological sciences : an official journal of the Society of Toxicology*, 2016, 151, 88-103.

21. J. Rooney, K. Oshida, N. Vasani, B. Vallanat, N. Ryan, B. N. Chorley, X. Wang, D. A. Bell, K. C. Wu, L. M. Aleksunes, C. D. Klaassen, T. W. Kensler and J. C. Corton, Activation of Nrf2 in the liver is associated with stress resistance mediated by suppression of the growth hormone-regulated STAT5b transcription factor, *PLOS ONE*, 2018, 13, e0200004.

22. I. Kupersmidt, Q. J. Su, A. Grewal, S. Sundaresh, I. Halperin, J. Flynn, M. Shekar, H. Wang, J. Park, W. Cui, G. D. Wall, R. Wisotzkey, S. Alag, S. Akhtari and M. Ronaghi, Ontology-based meta-analysis of global collections of high-throughput public data, *PLoS one*, 2010, 5.

23. J. E. J. Hardyman, J. Tyson, K. A. Jackson, C. Aldridge, S. J. Cockell, L. A. Wakeling, R. A. Valentine and D. Ford, Zinc sensing by metal-responsive transcription factor 1 (MTF1) controls metallothionein and ZnT1 expression to buffer the sensitivity of the transcriptome response to zinc, *Metallomics*, 2016, 8, 337-343.
24. J. Rooney, Chorley, B, Corton JC, Mining a human transcriptome database for chemical modulators of Nrf2, Submitted, 2020.
25. K. Oshida, N. Vasani, R. S. Thomas, D. Applegate, M. Rosen, B. Abbott, C. Lau, G. Guo, L. M. Aleksunes, C. Klaassen and J. C. Corton, Identification of modulators of the nuclear receptor peroxisome proliferator-activated receptor alpha (PPARalpha) in a mouse liver gene expression compendium, *PloS one*, 2015, 10, e0112655.
26. A. J. Williams, C. M. Grulke, J. Edwards, A. D. McEachran, K. Mansouri, N. C. Baker, G. Patlewicz, I. Shah, J. F. Wambaugh, R. S. Judson and A. M. Richard, The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, *J Cheminform*, 2017, 9.
27. C. Yang, A. Tarkhov, J. Maruszczyk, B. Bienfait, J. Gasteiger, T. Kleinoeder, T. Magdziarz, O. Sacher, C. H. Schwab, J. Schwobel, L. Terfloth, K. Arvidson, A. Richard, A. Worth and J. Rathman, New publically available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling, *J Chem Inf Model*, 2015, 55, 510-528.
28. Y. Cao, T. Backman, Y. Wang and T. Girke, ChemmineR - V2: Analysis of small molecule and screening data, *ChemMineR Manual*, 2011.
29. R. Agrawal and R. Srikant, Santiago, Chile, 1994.
30. M. Hahsler, B. Grun and K. Hornik, arules - A Computational Environment for Mining Association Rules and Frequent Item Sets, *J. Stat. Softw.*, 2005, 14.
31. N. O. Oki and S. W. Edwards, An integrative data mining approach to identifying adverse outcome pathway signatures, *Toxicol*, 2016, 350-352, 49-61.
32. S. M. Bell and S. W. Edwards, Philadelphia, PA, 2014.
33. S. M. Bell and S. W. Edwards, Identification and Prioritization of Relationships between Environmental Stressors and Adverse Human Health Impacts, *Environ Health Perspect*, 2015, 123, 1193-1199.
34. E. J. Gardiner and V. J. Gillet, Perspectives on Knowledge Discovery Algorithms Recently Introduced in Chemoinformatics: Rough Set Theory, Association Rule Mining, Emerging Patterns, and Formal Concept Analysis, *J Chem Inf Model*, 2015, 55, 1781-1803.
35. S. Naulaerts, P. Meysman, W. Bittremieux, T. N. Vu, W. Vanden Berghe, B. Goethals and K. Laukens, A primer to frequent itemset mining for bioinformatics, *Brief Bioinform*, 2015, 16, 216-231.
36. P. Yu and D. J. Wild, Fast rule-based bioactivity prediction using associative classification mining, *J Cheminformatics*, 2012, 4.
37. S. H. Park, S. Y. Jang, H. Kim and S. W. Lee, An association rule mining-based framework for understanding lifestyle risk behaviors, *PloS one*, 2014, 9, e88859.
38. R. Agrawal, T. Imielinski and A. Swami, Washington, DC, 1993.
39. C. Borgelt, Frequent item set mining, *WIREs Data Mining Knowl Discov*, 2012, 2, 437-456.
40. A. Shirato, T. Kikugawa, N. Miura, N. Tanji, N. Takemori, S. Higashiyama and M. Yokoyama, Cisplatin resistance by induction of aldo-keto reductase family 1 member C2 in human bladder cancer cells, *Oncology Letters*, 2014, 7, 674-678.
41. T. Hai and M. G. Hartman, The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis, *Gene*, 2001, 273, 1-11.
42. T. Hai, C. D. Wolfgang, D. K. Marsee, A. E. Allen and U. Sivaprasad, ATF3 and Stress Responses, *Gene expression*, 1999, 7, 321-335.
43. G. B. Schwochau, K. A. Nath and M. E. Rosenberg, Clusterin protects against oxidative stress in vitro through aggregative and nonaggregative properties, *Kidney International*, 1998, 53, 1647-1653.
44. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander and T. R. Golub, The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease, *Science (New York, N.Y.)*, 2006, 313, 1929-1935.
45. W.-Q. Ding and S. E. Lind, Metal ionophores – An emerging class of anticancer drugs, *IUBMB Life*, 2009, 61, 1013-1018.
46. X. Mao, X. Li, R. Sprangers, X. Wang, A. Venugopal, T. Wood, Y. Zhang, D. A. Kuntz, E. Coe, S. Trudel, D. Rose, R. A. Batey, L. E. Kay and A. D. Schimmer, Clioquinol inhibits the proteasome and displays preclinical activity in leukemia and myeloma, *Leukemia*, 2008, 23, 585.
47. B. Regland, W. Lehmann, I. Abedini, K. Blennow, M. Jonsson, I. Karlsson, M. Sjögren, A. Wallin, M. Xilinas and C. G. Gottfries, Treatment of Alzheimer's Disease with Clioquinol, *Dementia and Geriatric Cognitive Disorders*, 2001, 12, 408-414.
48. J. L. Allensworth, M. K. Evans, F. Bertucci, A. J. Aldrich, R. A. Festa, P. Finetti, N. T. Ueno, R. Safi, D. P. McDonnell, D. J. Thiele, S. Van Laere and G. R. Devi, Disulfiram (DSF) acts as a copper ionophore to induce copper-dependent oxidative stress and mediate anti-tumor efficacy in inflammatory breast cancer, *Molecular Oncology*, 2015, 9, 1155-1168.
49. H. L. Wiggins, J. M. Wymant, F. Solfa, S. E. Hiscox, K. M. Taylor, A. D. Westwell and A. T. Jones, Disulfiram-induced cytotoxicity and endo-lysosomal sequestration of zinc in breast cancer cells, *Biochem Pharmacol*, 2015, 93, 332-342.
50. J. L. Allensworth, M. K. Evans, F. Bertucci, A. J. Aldrich, R. A. Festa, P. Finetti, N. T. Ueno, R. Safi, D. P. McDonnell, D. J. Thiele, S. Van Laere and G. R. Devi, Disulfiram (DSF) acts as a copper ionophore to induce copper-dependent oxidative stress and mediate anti-tumor efficacy in inflammatory breast cancer, *Molecular oncology*, 2015, 9, 1155-1168.
51. J. E. Hood, J. W. Jenkins, D. Milatovic, L. Rongzhu and M. Aschner, Mefloquine induces oxidative stress and neurodegeneration in primary rat cortical neurons, *NeuroToxicology*, 2010, 31, 518-523.
52. S. Qiao, S. D. Lamore, C. M. Cabello, J. L. Lesson, J. L. Muñoz-Rodríguez and G. T. Wondrak, Thioestrepton is an inducer of oxidative and proteotoxic stress that impairs viability of human melanoma cells but not primary melanocytes, *Biochemical Pharmacology*, 2012, 83, 1229-1240.
53. S. Lechner, K. Muller-Ladner U Fau - Schlottmann, B. Schlottmann K Fau - Jung, M. Jung B Fau - McClelland, J. McClelland M Fau - Ruschoff, J. Ruschoff J Fau - Welsh, J. Welsh J Fau - Scholmerich, F. Scholmerich J Fau - Kullmann and F. Kullmann, Bile acids mimic oxidative stress induced upregulation of thioredoxin reductase in colon cancer cell lines, *Carcinogenesis*, 2002, 23, 1281-1288.
54. S. C. Ramaiahgari, S. S. Auerbach, T. O. Saddler, J. R. Rice, P. E. Dunlap, N. S. Sipes, M. J. DeVito, R. R. Shah, P. R. Bushel, B. A. Merrick, R. S. Paules and S. S. Ferguson, The Power of Resolution: Contextualized Understanding of Biological Responses to Liver Injury Chemicals Using High-throughput Transcriptomics and Benchmark Concentration Modeling, *Toxicological sciences : an official journal of the Society of Toxicology*, 2019, 169, 553-566.
55. J. C. Corton, N. C. Kleinstreuer and R. S. Judson, Identification of potential endocrine disrupting chemicals using gene expression biomarkers, *Toxicology and applied pharmacology*, 2019, DOI: 10.1016/j.taap.2019.114683, 114683.

ARTICLE

Journal Name

56. J. C. Corton, A. Williams and C. L. Yauk, Using a gene expression biomarker to identify DNA damage-inducing agents in microarray profiles, *Environmental and molecular mutagenesis*, 2018, 59, 772-784.

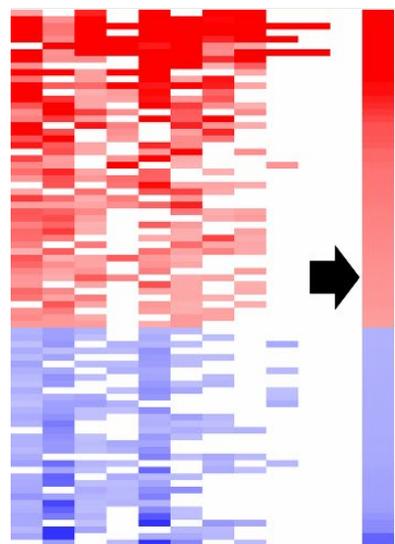
57. S. W. Edwards, Y. M. Tan, D. L. Villeneuve, M. E. Meek and C. A. McQueen, Adverse Outcome Pathways-Organizing Toxicological Information to Improve Decision Making, *The Journal of pharmacology and experimental therapeutics*, 2016, 356, 170-181.

58. E. J. Perkins, R. Ashauer, L. Burgoon, R. Conolly, B. Landesmann, C. Mackay, C. A. Murphy, N. Pollesch, J. R. Wheeler, A. Zupanec and S. Scholz, Building and Applying Quantitative Adverse Outcome Pathway Models for Chemical Hazard and Risk Assessment, *Environmental toxicology and chemistry / SETAC*, 2019, 38, 1850-1865.

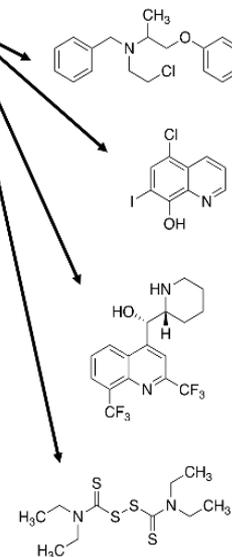
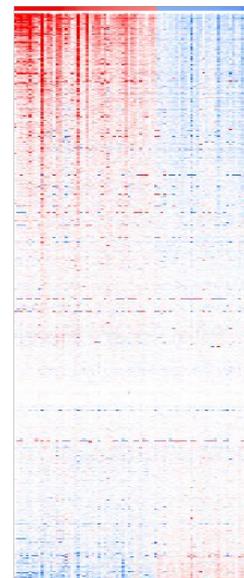
59. J. R. Phillips, D. L. Svoboda, A. Tandon, S. Patel, A. Sedykh, D. Mav, B. Kuo, C. L. Yauk, L. Yang, R. S. Thomas, J. S. Gift, J. A. Davis, L. Olszyk, B. A. Merrick, R. S. Paules, F. Parham, T. Saddler, R. R. Shah and S. S. Auerbach, *BMDExpress 2: enhanced transcriptomic dose-response analysis workflow*, *Bioinformatics (Oxford, England)*, 2019, 35, 1780-1782.

60. S. M. Bell, X. Chang, J. F. Wambaugh, D. G. Allen, M. Bartels, K. L. R. Brouwer, W. M. Casey, N. Choksi, S. S. Ferguson, G. Fraczkiwicz, A. M. Jarabek, A. Ke, A. Lumen, S. G. Lynn, A. Paini, P. S. Price, C. Ring, T. W. Simon, N. S. Sipes, C. S. Sprankle, J. Strickland, J. Troutman, B. A. Wetmore and N. C. Kleinstreuer, In vitro to in vivo extrapolation for high throughput prioritization and decision making, *Toxicology in vitro : an international journal published in association with BIBRA*, 2018, 47, 213-227.

61. W. AR, Directive to prioritize efforts to reduce animal testing, <https://www.epa.gov/sites/production/files/2019-09/documents/image2019-09-09-231249.pdf>, 2019.

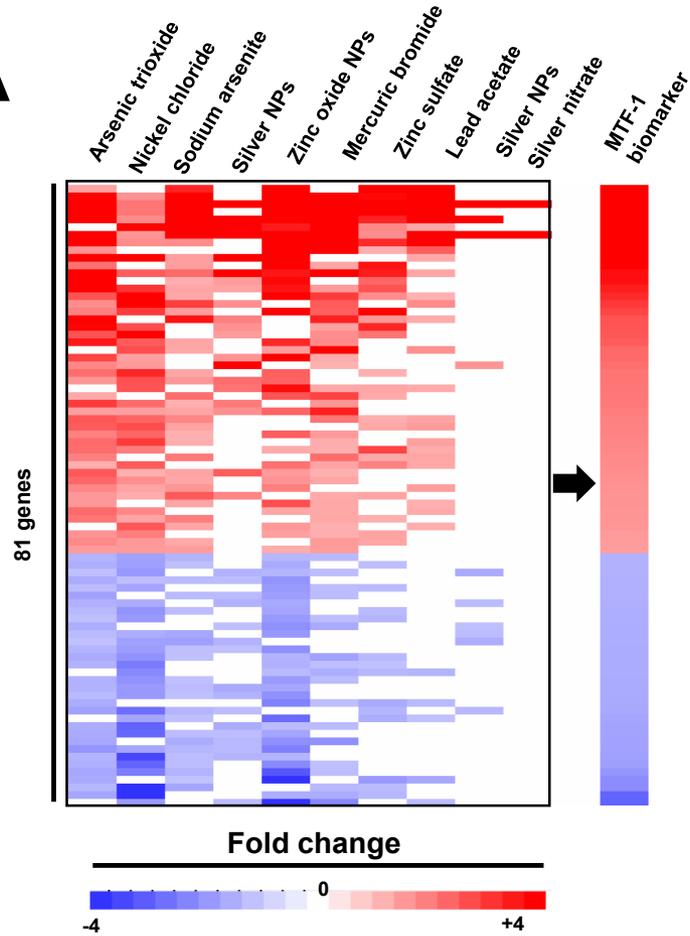


Human
microarray
compendium

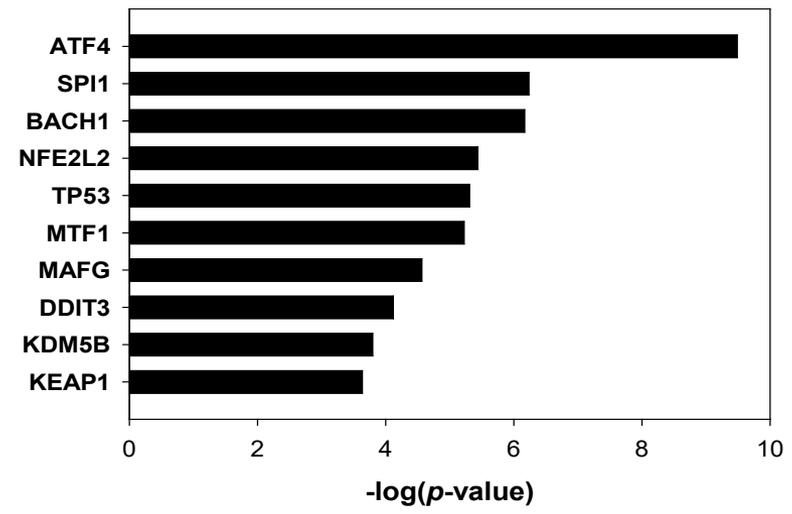
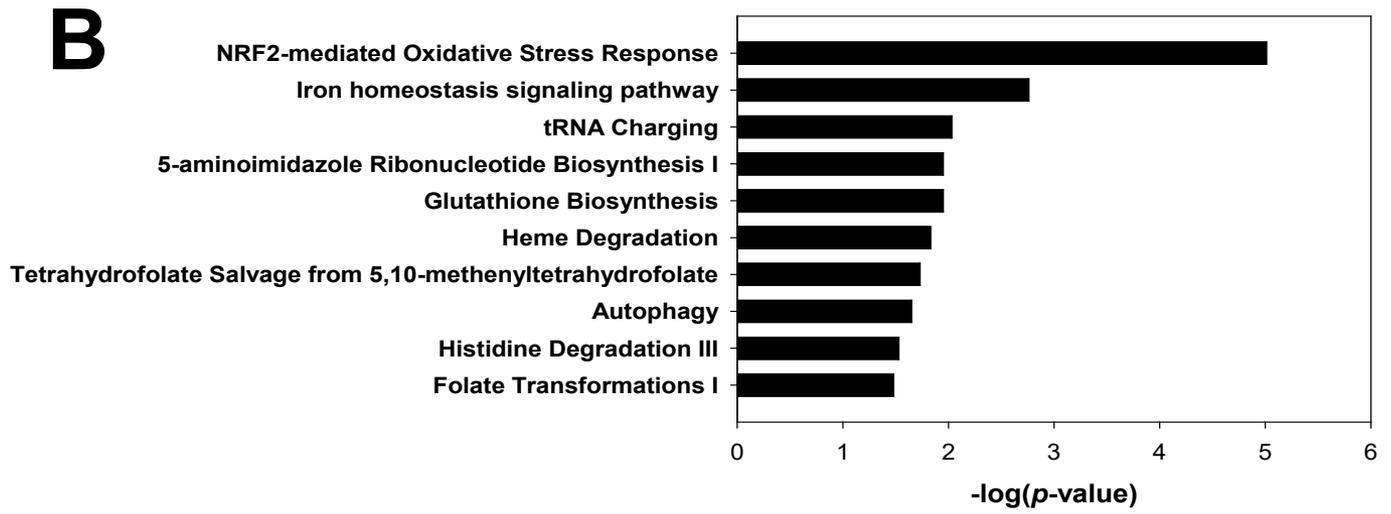


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

A

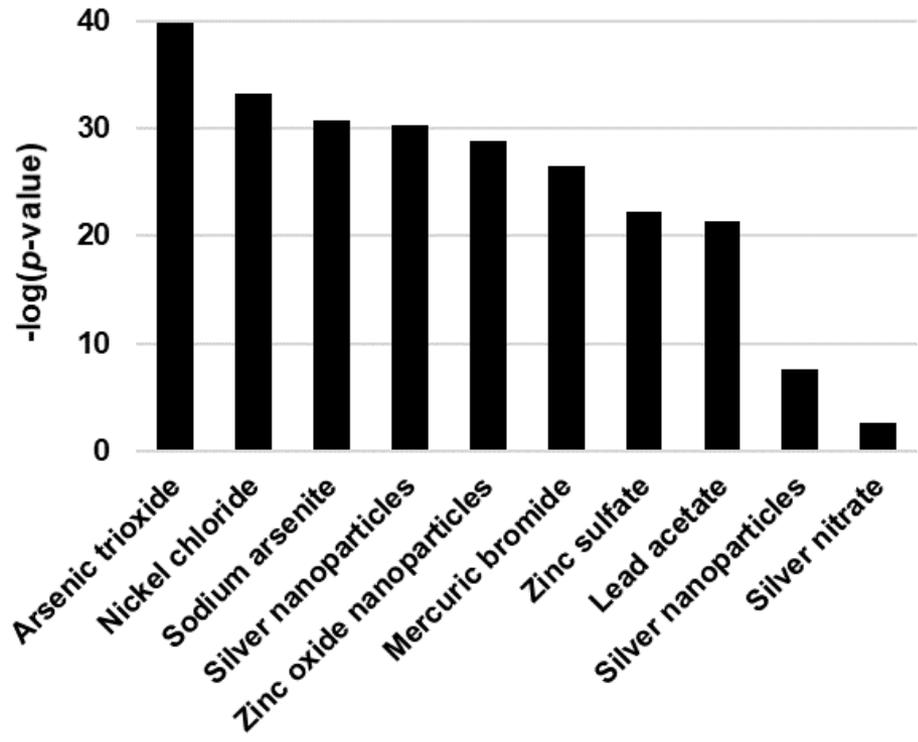


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41



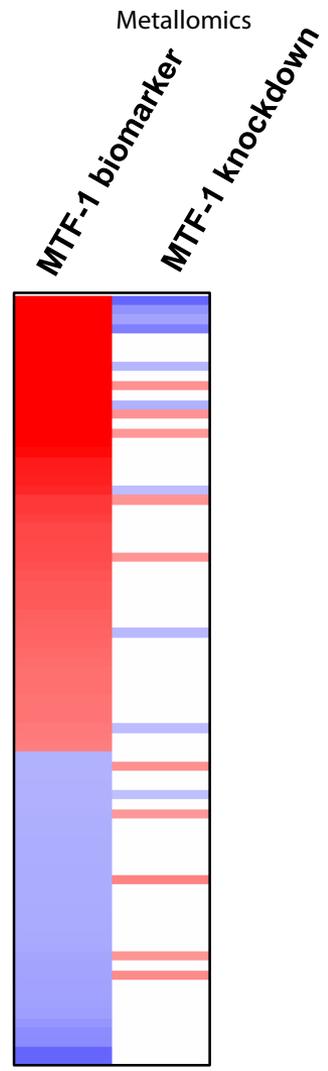
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

C

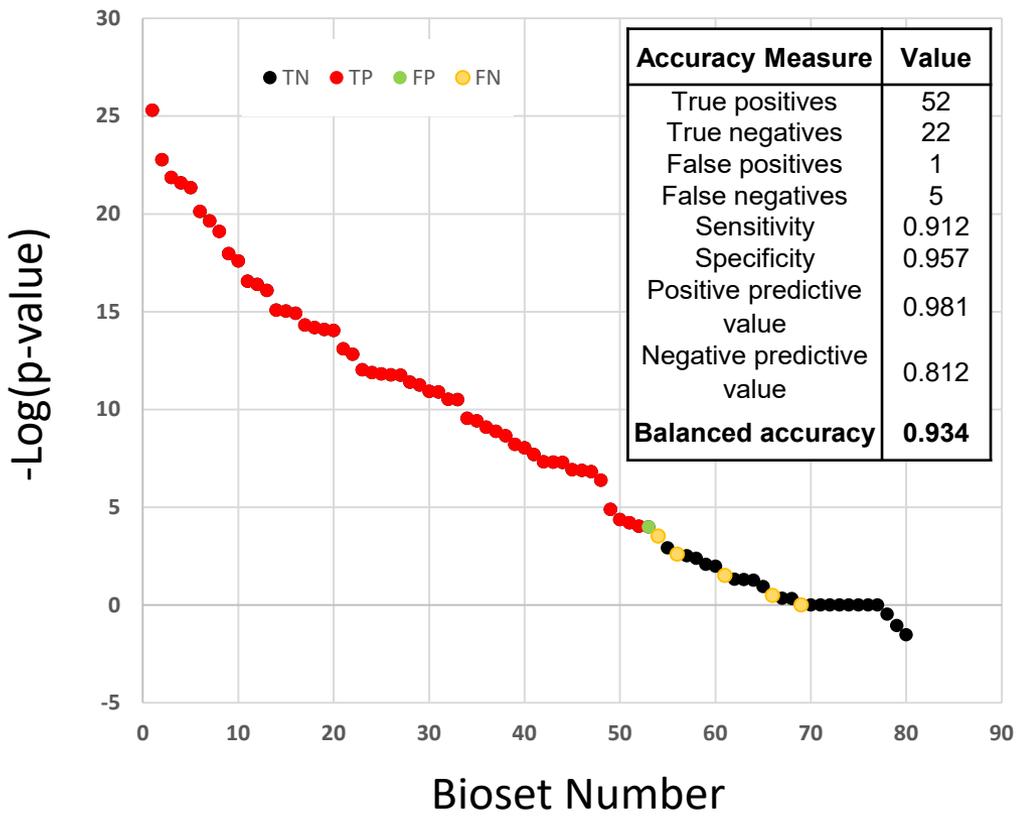


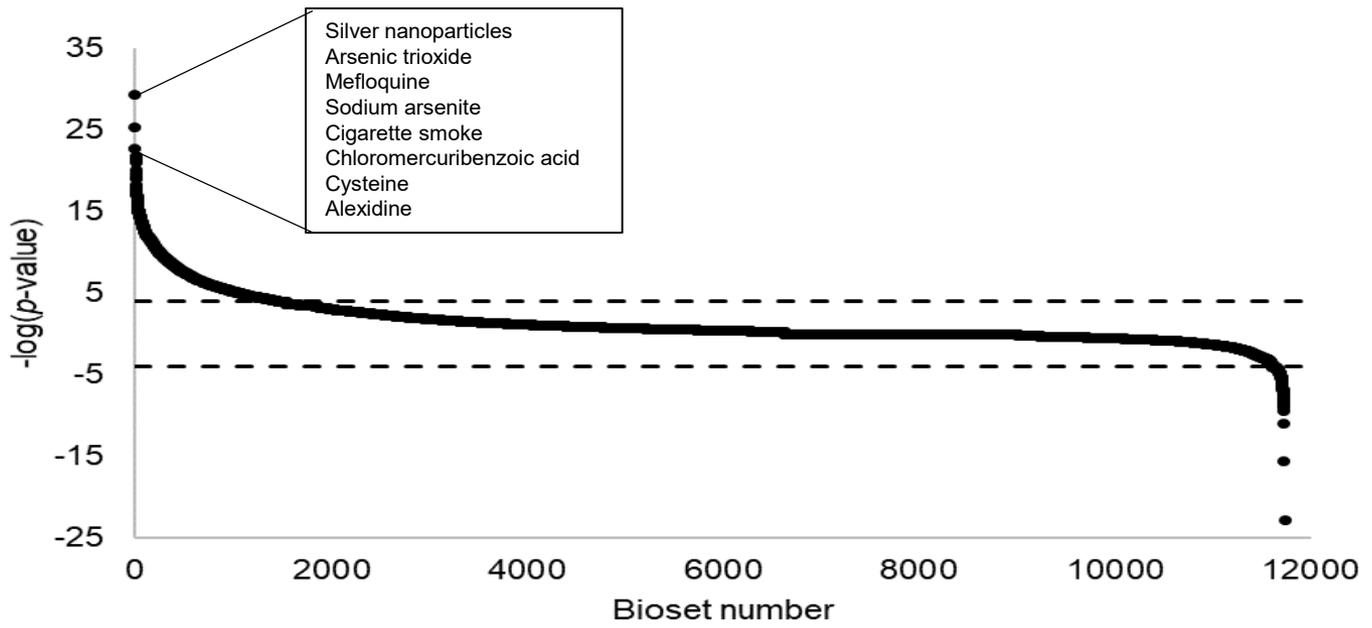
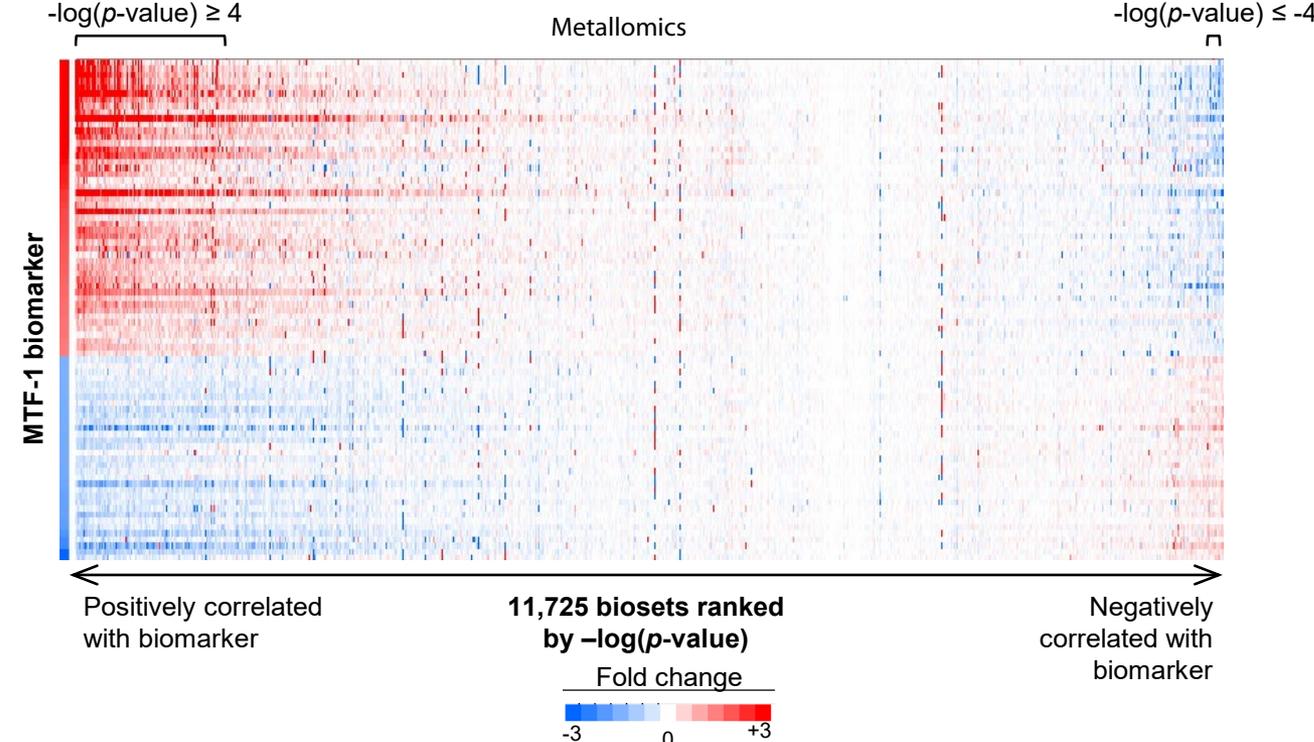
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

D



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

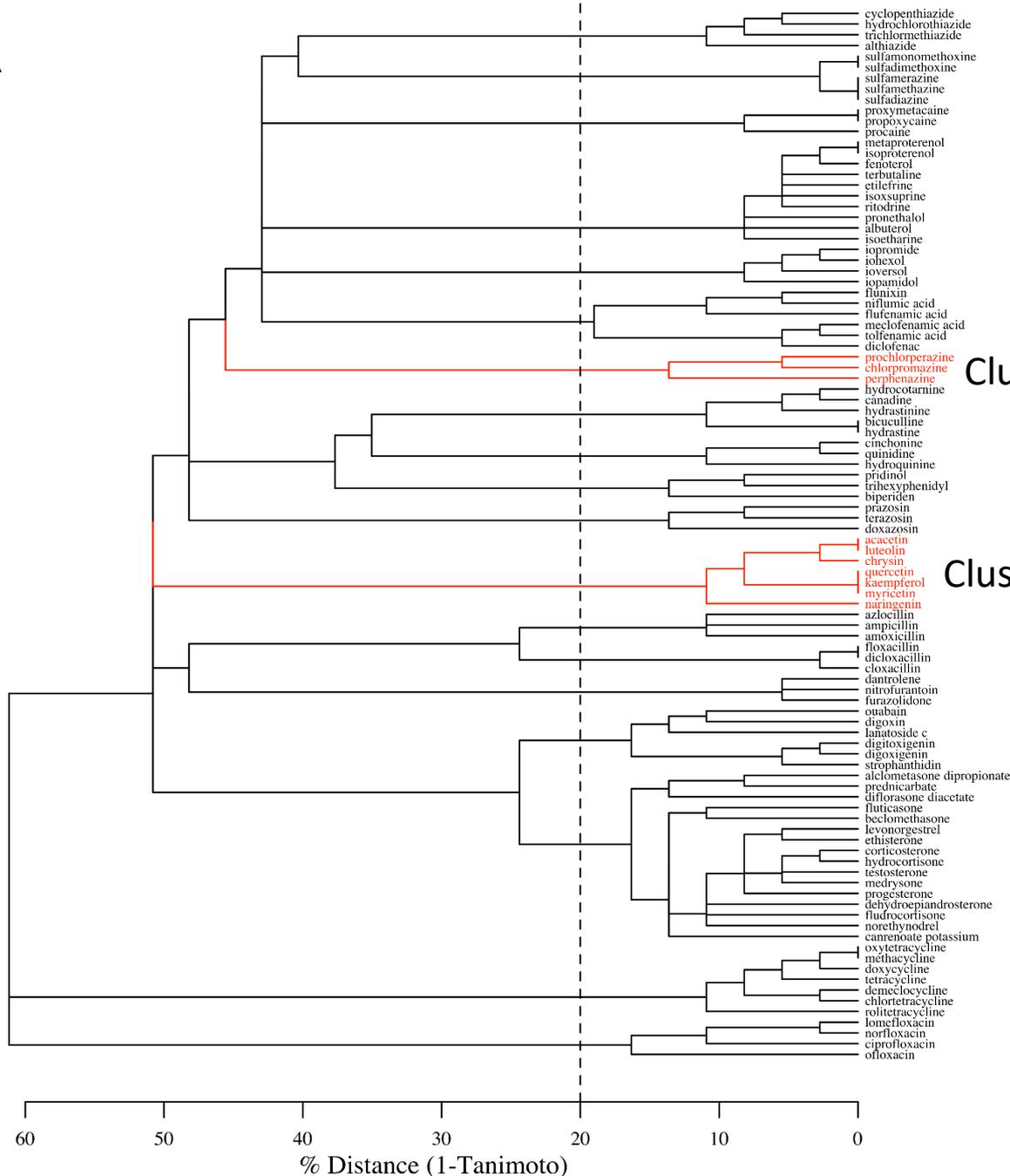




A

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

Metallomics



Cluster 125

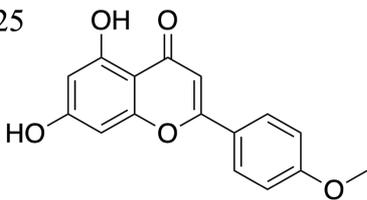
Cluster 25

60 50 40 30 20 10 0
% Distance (1-Tanimoto)

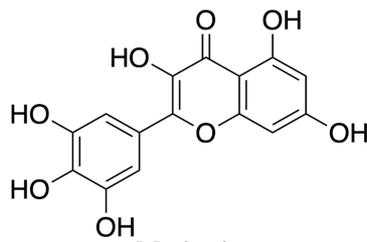
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

B

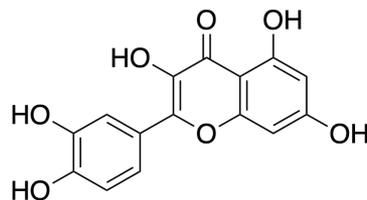
Cluster 25



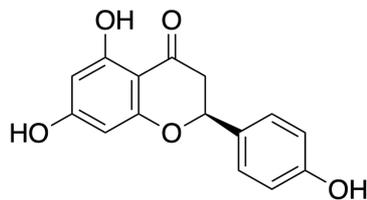
Acacetin
Positive



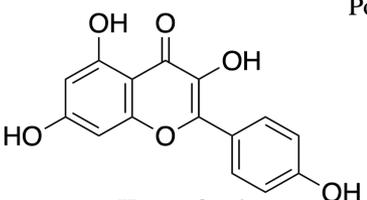
Myricetin
Positive



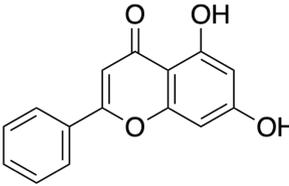
Quercetin
Negative



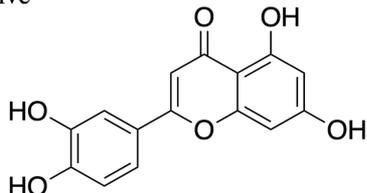
Naringenin
Negative



Kaempferol
Positive

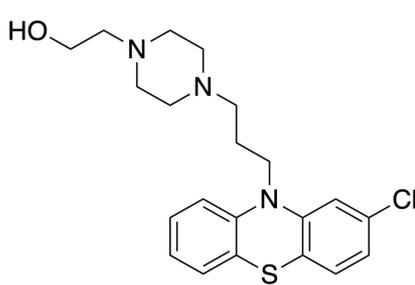


Chrysin
Negative

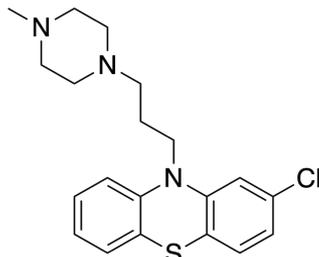


Luteolin
Negative

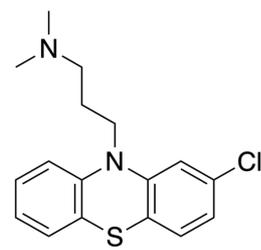
Cluster 125



Perphenazine
Positive



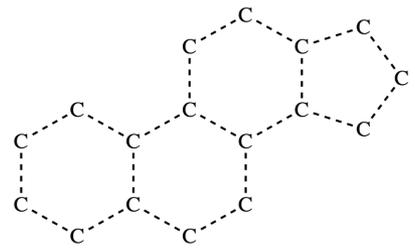
Prochlorperazine
Positive @ 10 μ M
Negative @ 6.6 μ M



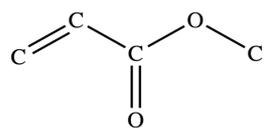
Chlorpromazine
Positive @ 11.2 μ M
Negative @ 1 μ M

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

A

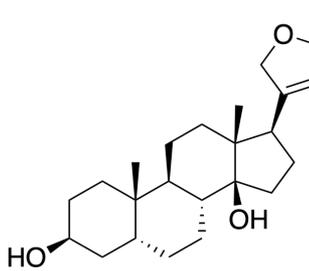


ring:fused_steroid_generic_[5_6_6_6]

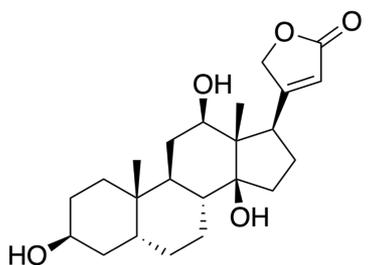


bond:C(=O)O_carboxylicEster_alkenyl

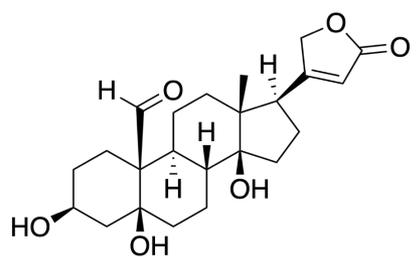
B



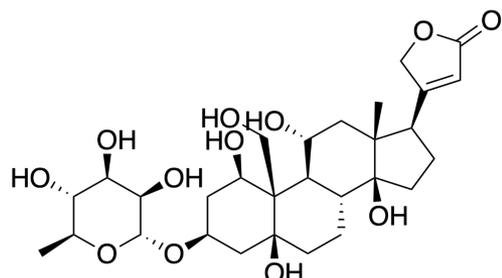
Digitoxigenin
Positive



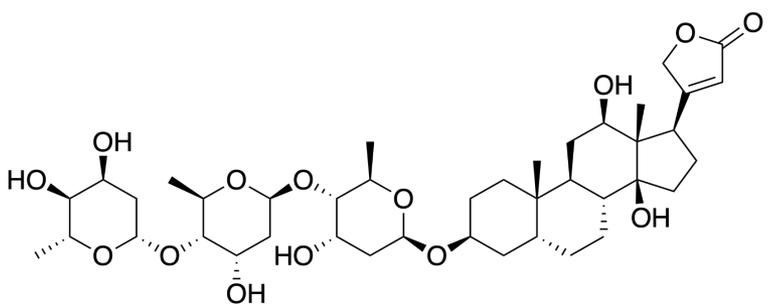
Digoxigenin
Positive



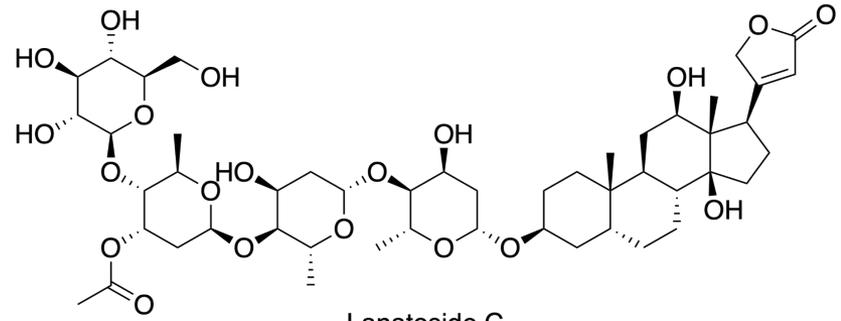
Strophanthidin
Positive



Ouabain
Positive

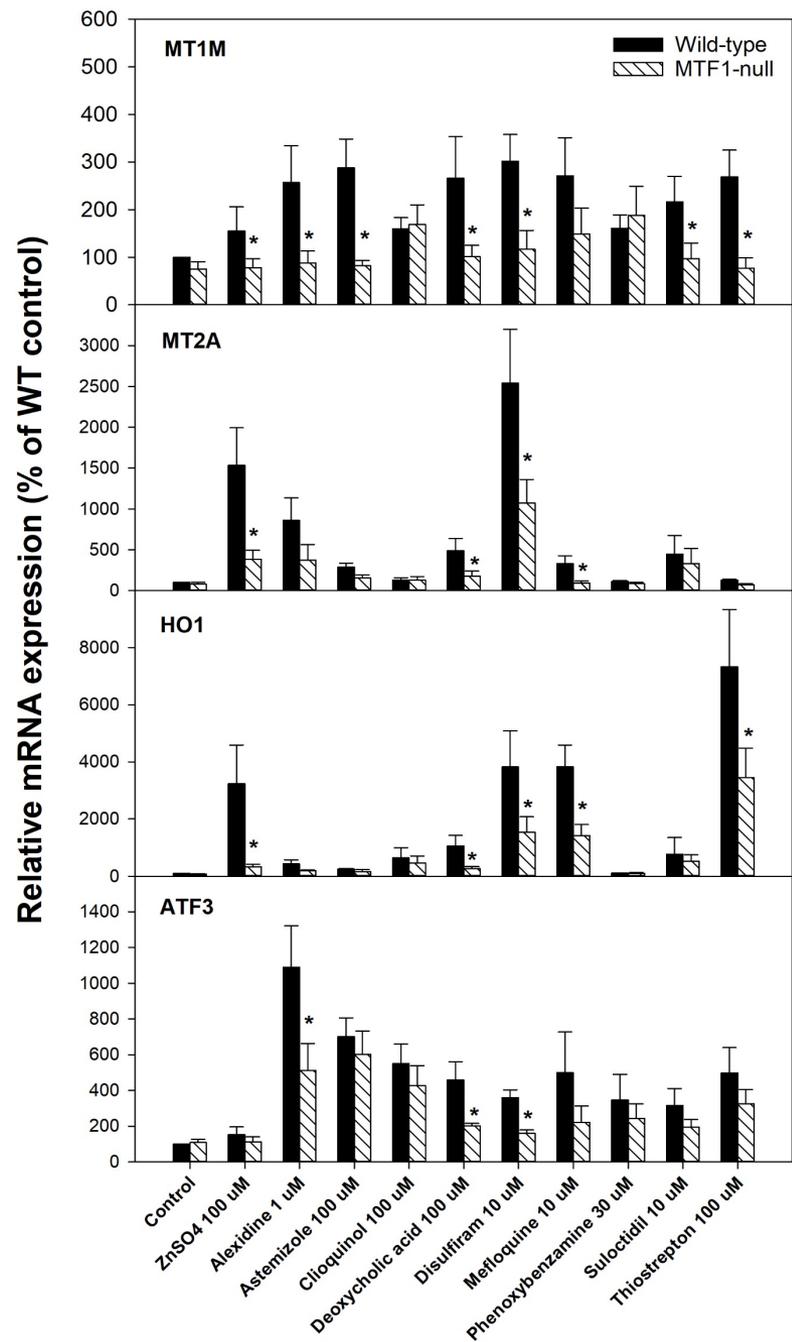


Digoxin
Positive



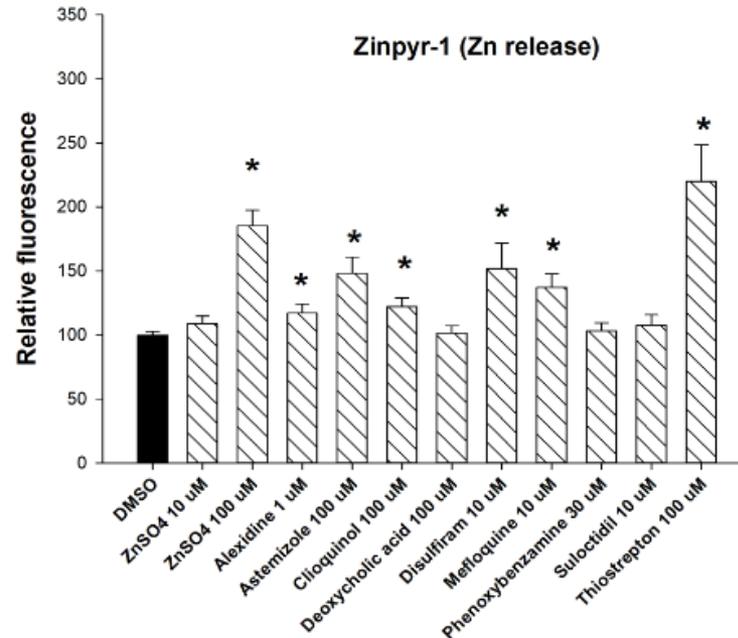
Lanatoside C
Positive

Metallomics

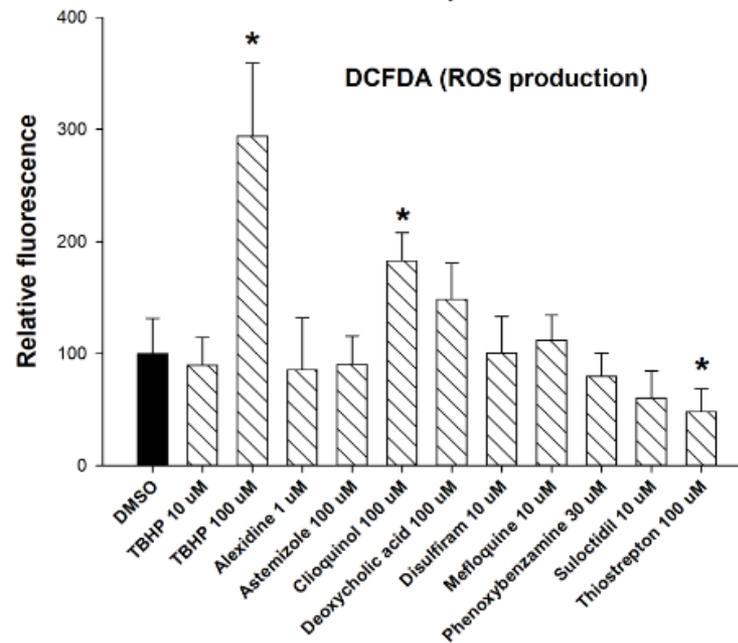


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

A



B

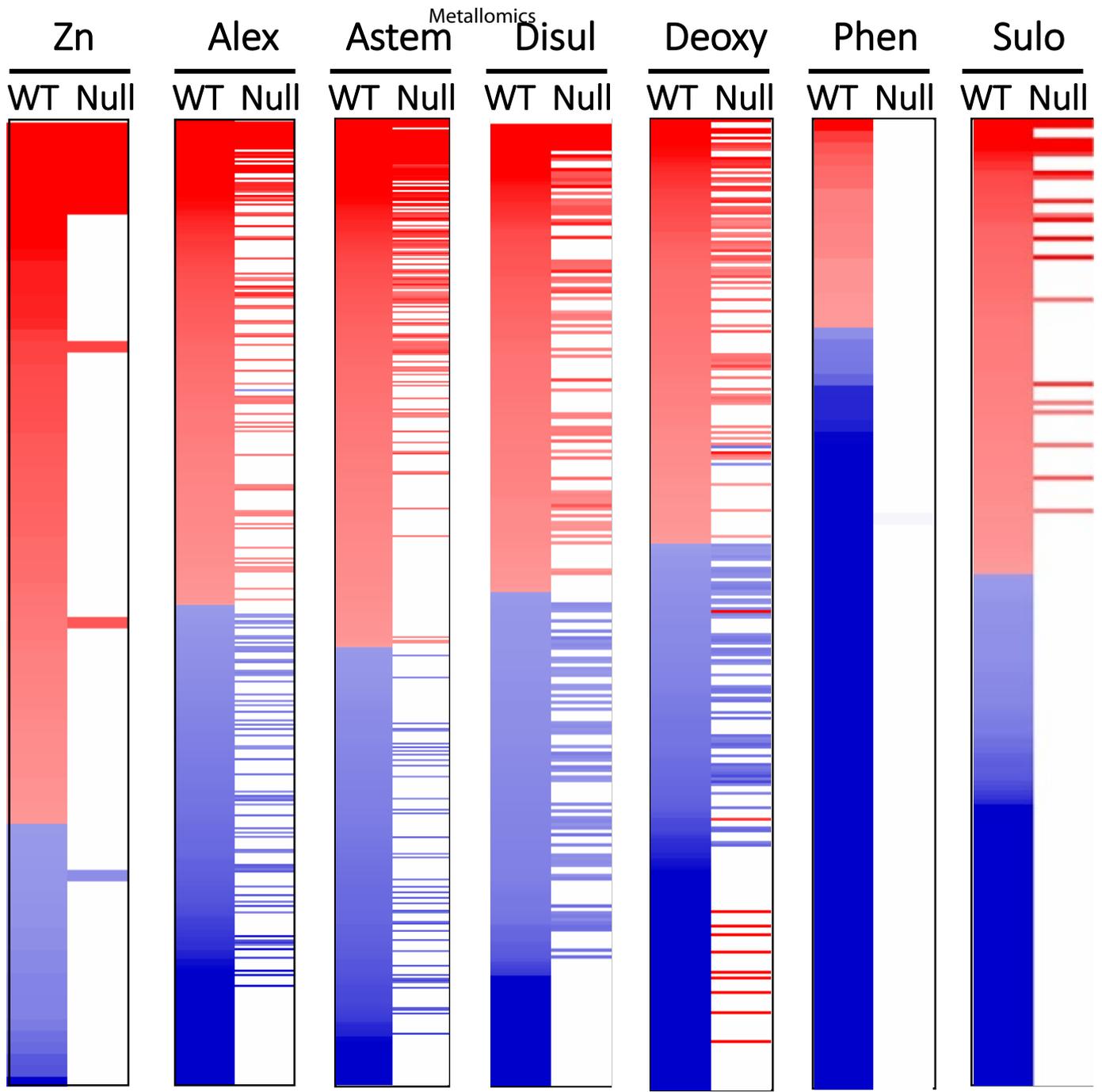


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

Figure 8

A

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41



Fold-change
-3 0 +3

B