EPA Public Access

Author manuscript

Comput Toxicol. Author manuscript; available in PMC 2018 October 17.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Comput Toxicol. 2017 May ; 2: 39–44. doi:10.1016/j.comtox.2017.03.001.

Chemical-agnostic hazard prediction: statistical inference of *in vitro* toxicity pathways from proteomics responses to chemical mixtures

Jeffrey A Ross, Barbara Jane George, Maribel Bruno, and Yue Ge

National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711

Abstract

Toxicity pathways have been defined as normal cellular pathways that, when sufficiently perturbed as a consequence of chemical exposure, lead to an adverse outcome. If an exposure alters one or more normal biological pathways to an extent that leads to an adverse toxicity outcome, a significant correlation must exist between the exposure, the extent of pathway alteration, and the degree of adverse outcome. Biological pathways are regulated at multiple levels, including transcriptional, post-transcriptional, post-translational, and targeted degradation, each of which can affect the levels and extents of modification of proteins involved in the pathways. Significant alterations of toxicity pathways resulting from changes in regulation at any of these levels therefore are likely to be detectable as alterations in the proteome. We hypothesize that significant correlations between exposures, adverse outcomes, and changes in the proteome have the potential to identify putative toxicity pathways, facilitating selection of candidate targets for high throughput screening, even in the absence of *a priori* knowledge of either the specific pathways involved or the specific agents inducing the pathway alterations. We explored this hypothesis *in vitro* in BEAS-2B human airway epithelial cells exposed to different concentrations of Ni²⁺, Cd²⁺, and Cr⁶⁺, alone and in defined mixtures. Levels and phosphorylation status of a variety of signaling pathway proteins and cytokines were measured after 48 hours exposure, together with cytotoxicity. Least Absolute Shrinkage and Selection Operator (LASSO) multiple regression was used to identify a subset of these proteins that constitute a putative toxicity pathway capable of predicting cytotoxicity. The putative toxicity pathway for cytotoxicity of these metals and metal mixtures identified by LASSO is composed of phospho-RPS6KB1, phospho-p53, cleaved CASP3, phospho-MAPK8, IL-10, and Hif-1 α . As this approach does not depend on knowledge of the chemical composition of the mixtures, it may be generally useful for identifying sets of proteins predictive of adverse effects for a variety of mixtures, including complex environmental mixtures of unknown composition.

Keywords

Proteomics; toxicity pathways; cytotoxicity; *in vitro* assays; chromium; cadmium; nickel

INTRODUCTION

The National Academy of Sciences has proposed approaching the assessment of toxicological risk of chemical exposures from the standpoint of toxicity pathways, which are defined as normal biological pathways that, if sufficiently perturbed, lead to possible adverse health outcomes (Committee on Toxicity Testing and Assessment of Environmental Agents, 2007). The premise is that if such pathways can be identified, then the risk posed by toxicant exposure may be inferred from the extent of pathway disruption. The appeal of this approach is that it focuses primarily on the extent of pathway perturbation, and thus is largely independent of the specific chemical or chemical mixture exposure inducing the pathway alteration. Accordingly, the approach should be applicable to assessing the toxicity of chemical mixture exposures as well as single agent exposures. Once a toxicity pathway is identified, measurement of alteration of a relatively small number of key events within that pathway in response to toxicant exposures can provide a rapid means of predicting the extent of toxicity that the exposure will induce. A key conceptual difficulty presented by this approach for predictive screening is the identification of appropriate pathways and key events to measure.

An inherent requirement for the application of toxicity pathway key event monitoring to predict an adverse outcome is that there must exist robust relationships between the pathway disrupting exposures, the extent of alteration of the key events being monitored, and the adverse outcome. If a node in a toxicity pathway is relatively unaffected by the exposure, or if alterations of that node are not strongly linked to the adverse outcome, then assessment of the status of that node in response to an exposure is not likely to yield a useful prediction of the adverse outcome induced by that exposure. Therefore, for an altered pathway node to be a useful indicator of toxicity pathway alteration, the extent of its alteration will exhibit significant correlations with both exposure and adverse outcome. In addition, the magnitude of node alteration in response to the exposure ideally should be large enough at the time of measurement that the measured change in the node can be accurately assessed experimentally.

Alteration of biological pathways in response to chemical exposures can occur at many different levels, including large scale chromatin reorganization, induction of mutations in pathway genes, modulation of transcription, post-transcriptional regulation of translation, post-translational protein modifications, and targeted degradation of pathway proteins. Each of these levels of regulation has the potential of affecting the levels and extents of modification of pathway proteins, and thus pathway function, in response to pathway perturbing exposures. Alterations at the level of the proteome reflect the cumulative result of regulation at all of these levels, and are therefore more likely to be reflective of pathway dysregulation and predictive of adverse health effects than alterations at upstream levels.

These considerations suggest that, at a practical level, it should be possible to identify key nodes in putative toxicity pathways by a relatively straightforward assessment of the correlations between a series of exposures, a series of measurements of cellular molecular events, and the extent of induction of an adverse effect elicited in cells by the exposures, even in the absence of detailed *a priori* knowledge of the toxicity pathways involved. We

have examined this approach capitalizing on the availability of a previously published data set (Ge *et al.*, 2015) generated by *in vitro* exposure of BEAS-2B human airway epithelial cells to different concentrations of Cd^{2+} , Cr^{6+} , and Ni^{2+} , both individually and in combinations at different concentrations, and then assessing the levels of induced cytotoxicity as well as levels of selected proteins and extent of post-translational protein modifications in the cells. The proteins assessed were selected based upon a literature survey of the effects of the metals on cells, including induction of reactive oxygen stress, inflammation, and altered cell signaling. Multiple regression analysis was performed to identify a parsimonious subset of altered proteins for which the extent of alteration is predictive of the extent of cytotoxicity among the mixtures examined.

MATERIALS and METHODS

Cell Treatments

Details of the cell treatments, determinations of protein level and extent of modification, and cytotoxicity have been explained in a previous publication (Ge *et al.*, 2015). Briefly, BEAS-2B human respiratory epithelial cells (American Type Culture Collection, Rockville, MD) were cultured in LHC-9 medium. The cells were maintained at $37 \pm 1^\circ\text{C}$, $85 \pm 5\%$ relative humidity, and $5 \pm 1\%$ CO_2 . Medium was replaced every second day and cells were sub-cultured at approximately 80% confluence by washing with HEPES-buffered saline solution, followed by a wash with Versene and dislodged with TrypLe. For cytotoxicity measurements, cells were seeded at an initial density of 30,000 cells/cm² in 0.125 ml of media per well of a 96-well plate. For proteomic measurements, cells were seeded at 40,000 cells/cm² in 150 cm² flask.

Treatment with metals and cytotoxicity assay

Forty eight hours after seeding, 0.125ml of media containing the metal treatment was applied to each well of the cytotoxicity assay plate. The metal treatment conditions used are shown in Table 1. The flasks for the proteomic endpoints were treated at the same time as the cytotoxicity plates. The culture medium was removed and replaced with 15 ml of LHC-9 and 15 ml of the appropriate metal dilution. All exposures to the metal treatments were forty eight hours for both cytotoxicity and proteomic endpoints.

Cytotoxicity of the metals in BEAS-2B cells was determined by the neutral red assay. Briefly, medium was removed and sterile replaced with LHC-9 medium containing 0.003% neutral red dye. The cells were incubated for 3 hours at 37°C . Neutral red medium was removed, wells were rinsed with Dulbecco's phosphate buffered saline, and 100 μl of extraction buffer (50% ethanol, 1% acetic acid in water) was added to each well. Plates were gently shaken for 20–40 minutes and absorbance measured at 540 nm to measure dye uptake. Each plate included cell-free blank wells and untreated control cell wells.

Protein symbols

Proteins are referenced by their HGNC approved symbols. In our previous publication (Ge *et al.*, 2015), MAPK8, mitogen-activated protein kinase 8, was referred to as JNK, and RPS6KB1, ribosomal protein S6 kinase B1, was referred to as p70S6K.

Protein assays

Assessment of levels of expression or post-translational modifications of 21 proteins across 29 different treatments varying in the amounts of Ni²⁺, Cd²⁺, or Cr⁶⁺ present was carried out using MSD® Multi-Spot immunoassays with electrochemiluminescent detection (Meso Scale Discovery, Inc., Rockville, MD) as previously described (Ge *et al.*, 2015). Endpoints measured were total protein for Hif-1 α , p53, GM-CSF, IFN- γ , IL-1 β , IL-10, IL-12p70, IL-2, IL-6, IL-8, and TNF- α , as well as phospho-p53 (Ser15), phospho-GSK-3 β (Ser9), phospho-RPS6KB1 (Thr421/Ser424), phospho-Akt (Ser474), phospho-EGFR (Tyr1173), phospho-MAPK8 (Thr183/Tyr185), phospho-ERK1/2 (Thr202/Tyr204; Thr185/Tyr187), phospho-ErbB2 (Tyr1248), cleaved PARP (Asp214), and cleaved CASP3 (Asp 175). Assays were conducted according to manufacturer's protocols. Results from each assay were expressed as mean fold change relative to levels in untreated control cell cultures.

Statistical Analyses

Statistical analyses were performed using SigmaPlot for Windows version 12.5 (Systat Software, Inc., San Jose, CA), and SAS version 9.4 (SAS Institute, Cary, NC). Feature selection by multiple regression for modeling cytotoxicity as a function of protein fold-changes was performed with R version 3.2.2 (<http://www.r-project.org/>) using the cv.glmnet routine in the glmnet package version 2.0–2 for cross-validated LASSO feature selection (Friedman *et al.*, 2010). We used 10-fold cross validation to select values of λ yielding minimum mean squared error (MSE). Since the folds for cross validation are selected at random, repeated runs of the routine may return models with different feature sets yielding minimum MSEs. Therefore, the cross-validated LASSO was repeated 100 times to identify different possible model feature sets yielding minimal MSE values. We applied the “one-standard error” rule (Hastie *et al.*, 2009), selecting the model with the fewest non-zero parameters whose error is no more than one standard error greater than the MSE of the best model.

We repeated the parameter selection for the model using the elastic net method (Zou and Hastie, 2005), also using the glmnet package. Cross validation to minimize MSE for α selection was performed using a precomputed fold vector for values of α iteratively sampled between 0.05 and 1.00 at a step size of 0.05. This value of α was used to identify the cross validated λ yielding minimum MSE plus one standard-error values to identify the parameter set for the predictive model.

Values for the model intercept and coefficients for the selected parameters, together with the variance inflation factor (VIF), were calculated using the multiple linear regression routine of SigmaPlot for Windows version 13.0 (Systat Software, Inc., San Jose, CA). The VIF provides a measure of the extent of multi-collinearity present in the selected parameter set. Excessive multi-collinearity is undesirable, as it increases the variance and numerical instability of parameter estimates. Network diagrams were constructed using Cytoscape version 3.3.0 (<http://www.cytoscape.org/download.php>) (Shannon *et al.*, 2003).

Proteins identified by the LASSO procedure as being predictive of survival were mapped to canonical pathways using the Database for Annotation, Visualization and Integrated

Discovery (DAVID), version 6.7 (<https://david.ncifcrf.gov/home.jsp>) as described by Huang *et al.* (2009a; b) The list of genes corresponding to the selected proteins were mapped against KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways and assessed for significance of representation in the pathway using a modified Fisher exact p-value. Gene networks based on pathway and genetic interactions were identified using the GeneMANIA plug-in for Cytoscape (Montejo *et al.*, 2010; Warde-Farley *et al.*, 2010). The list of genes was also examined for enriched transcription factor targets using the iRegulon plug-in for Cytoscape (Janky *et al.*, 2014), with a maximum false discovery rate set to 0.00025.

RESULTS

The cross-validated LASSO, using mean relative cell survival as the dependent variable and mean relative levels of expression or phosphorylation of 21 proteins as independent variables across 29 different treatments varying in the amounts of Ni^{2+} , Cd^{2+} , or Cr^{6+} present, was effective in selecting a subset of proteins whose expression was predictive of cytotoxicity. The cytotoxicity model with the lowest prediction error yielded a minimum MSE at $\lambda = 0.0234$. The model with the fewest parameters one standard-error greater, at $\lambda = 0.040$, included six proteins: phospho-RPS6KB1, phospho-p53, cleaved CASP3, phospho-MAPK8, IL10, and Hif-1 α . The coefficient estimates for this model and their p-values are given in Table 2. The MSE for this model was 0.018, and the overall regression p-value was <0.001 . The agreement between the survival values predicted by this model and measured survival is shown in Figure 2. Variance Inflation Factor (VIF) values for each of the parameters in the model are less than 4.0, indicating that there is little remaining multicollinearity among the model parameters selected (O'Brien, 2007). Using the elastic net procedure with the one standard-error rule to independently select parameters resulted in identification of the same subset of proteins as the LASSO

In order to identify putative toxicity pathway members, we developed a second round of cross-validated LASSO regression models using each of the proteins identified in the cytotoxicity LASSO as dependent variables and the remaining protein expression levels as the independent variables. The sets of identified predictors, their coefficient estimates selected at λ values yielding minimum cross validation MSEs, and p-values are shown in Table 3. For each of these second round models, the p-values for the overall F tests of the regression models were each <0.001

A network map was constructed with nodes for each of the proteins that the initial LASSO identified as highly correlated with cytotoxicity and for the predictors of these proteins (Figure 1). Arrows connect the nodes pointing from the predictor protein toward the protein highly correlated with cytotoxicity. The signs of the coefficients are indicated by the colors of the edges connecting the nodes, green for positive and red for negative. It should be noted that it would not be appropriate to attempt to predict either cytotoxicity or levels of proteins identified by the first round LASSO from these second round LASSO models. Rather, these should be viewed only as providing a higher level view of possible interactions among putative pathway members to serve as a starting point for comparisons to identified signaling pathways.

We mapped the proteins included in the identified predictive network against canonical KEGG pathways using DAVID bioinformatics tools. The top pathways to which our protein set map are the Toll-like receptor signaling pathway, pathways in cancer, colorectal cancer pathways, and the ErbB signaling pathway, each of which are mapped to our gene list with $p < 2 \times 10^{-8}$.

For the pathway level analyses, we included caspase 8 and 9 rather than caspase 3, as our assay measures cleaved caspase 3, and caspase 8 and 9 are major effectors of caspase 3 cleavage. Using the GeneMANIA plug-in for Cytoscape, all of the nodes identified within the predictivity network can be connected by known pathway or genetic interactions. We also used the DAVID bioinformatics tools to analyze canonical KEGG pathways for significant overlap with the genes within the identified predictivity network. We found four pathways within which our identified proteins were highly significantly represented ($p < 2 \times 10^{-8}$), including two signaling pathways, ErbB and Toll-like receptor pathways, as well as two cancer-related pathways. Analysis of the prediction network genes for transcription factor target overrepresentation using iRegulon identifies a cluster of 12 genes regulated by transcription factor ETV2 and a cluster of 9 genes regulated by transcription factor AP4.

DISCUSSION

The toxicity pathway approach for *in vitro* toxicological screening depends upon identifying a relatively small set of measurable key molecular pathway alterations that are robustly associated to the adverse health outcome being assessed. While it may be ultimately desirable to understand the complete molecular details of all affected pathway nodes, it is not necessary to assess the status of every node in the pathway being altered in order for the approach to be useful in predicting adverse effects. Indeed, since the biology of the cell is a complex network of interacting pathways, any particular protein's inclusion in, or exclusion from, a specific canonical pathway is arguably somewhat arbitrary. Any given toxicant exposure may also induce alterations in multiple biological pathways, each of which may contribute to one or more apical adverse health outcomes, particularly when the exposure is to a mixture of chemicals. Additionally, compensatory signaling mechanisms not normally active in the cell may be invoked by chemical exposure, as exemplified by the development of resistance to pathway-specific chemotherapeutic agents by some cancer cells (Logue and Morrison, 2012; Ramos and Bentires-Alj, 2015). From a pragmatic viewpoint, if the goal of the toxicity pathway approach is to identify a tractable number of measurable cellular markers predictive of an adverse outcome in response to toxicant exposure, then a direct assessment of the predictive relationships among a set of potential markers and the adverse outcome over a series of different exposures may prove more efficient than identification of all the individual canonical pathways and pathway members perturbed by the exposures.

Genes within biological pathways tend to be highly co-regulated (Huang *et al.*, 2006). In general, if a biological pathway is altered by exposure to a series of stressors, we may expect that multiple nodes within the affected pathway would be altered, and that the degree of alteration of pathway nodes reflects in some manner the extent of overall pathway alteration. It is therefore reasonable to expect that some nodes within a toxicant response pathway would exhibit a high degree of multicollinearity. Inclusion of highly multicollinear factors in

the model can lead to overall poor model predictivity (Dormann *et al.*, 2013). The cross-validated LASSO technique was used for parameter selection because it is relatively insensitive to multicollinearity, as it tends to retain only one from among a set of highly correlated coefficients and eliminate the others from the model (Friedman *et al.*, 2010). The success of this approach in eliminating collinearity from the set of selected parameters is reflected in the relatively low VIF values observed (Table 2).

The set of cytotoxicity predictor proteins identified in this study meet the functional requirements for indicators of toxicity pathway perturbations, i.e., the proteins are altered in a dose-responsive manner by different toxicant exposures and the extent of protein alteration is predictive of the toxicity resulting from the exposure (Figure 2). However, the NRC guidelines also state that “models of toxicity-pathway perturbations need to be developed to interpret results from toxicity tests in a mechanistic rather than simply empirical manner...” (Committee on Toxicity Testing and Assessment of Environmental Agents, 2007). It is therefore of interest to determine the extent to which the predictivity network identified by this approach corresponds to known cellular response networks. In general, there are published studies reporting almost all of the relationships identified in our predictivity network. For example, EGFR has been shown to induce Hif-1 α (Secades *et al.*, 2015), as well as RPS6KB1 via activation of a Raf/MEK/ERK cascade (Steelman *et al.*, 2011). Similarly, expression of proinflammatory cytokines IL-6 and IL-8 is highly correlated with increased cytotoxicity (Veranth *et al.*, 2004), and both are positive predictors of cleaved-CASP3 in our predictivity network, supporting a pro-apoptotic role. However, the connections between proteins in the predictivity network should not be considered to be directional, e.g., the arrow leading from cleaved-PARP to cleaved-CASP3 does not indicate that cleaved-PARP induces or is a precursor to cleaved-CASP3, but rather that the levels of both are positively correlated. In this example, cleavage of CASP3 proceeds the cleavage of PARP in the apoptotic pathway, but because the responses of both are highly collinear, only one, cleaved-CASP3, was retained by the first LASSO procedure as a predictor of cell survival. The second-level LASSO to identify the protein changes predictive of the selected survival predictors correctly identified that level of cleaved-PARP are positively correlated with cleaved CASP3 levels. To be clear, the network shown in Figure 1 is not proposed to be a toxicity pathway *per se*, but is a predictivity network illustrating the relationships between the proteins identified in the second level LASSO and the proteins in the top level LASSO, as well as between the top level LASSO proteins and cell survival. The utility of the second level predictors is intended to aid in the generation of hypotheses related to possible signaling pathway connections impacted over a range of cellular perturbations, not for making quantitative predictions.

The identification of a cluster of 9 genes in our predictivity network that are targets of transcription factor AP4 suggests a possible role for AP4 in mediating the cytotoxic effects of the metal mixtures examined in this study. In support of this hypothesis, transcription factor AP4 has been demonstrated to be induced by c-myc expression, and directly suppresses p21 and p16 (Liu *et al.*, 2012). Furthermore, c-myc expression was demonstrated to be induced by exposure to Ni²⁺ ions in BEAS-2B cells, and increased c-myc expression was responsible for induction of apoptosis mediated via the Ras/Erk signaling pathway (Li *et al.*, 2009). Elevated AP4 expression has been shown to mediate epithelial-mesenchymal

transition and correlates with increased metastasis in colorectal cancer and poor clinical prognosis in hepatocellular carcinoma (Hu *et al.*, 2013; Jackstadt and Hermeking, 2014; Xinghua *et al.*, 2012). Similarly, the identification of a cluster of 12 of the predictivity network genes as targets of transcription factor ETV2 suggests a role its involvement in the regulation of cellular responses to cytotoxic metals. ETV2 has been found to be upregulated by both nickel and chromium in the 1HAEO- human respiratory epithelial cell line (Kaczmarek *et al.*, 2007). ETV2 is an important regulator of vasculogenesis and hematopoiesis (Lammerts van Bueren and Black, 2012) and plays a role in the differentiation of endothelium (Veldman *et al.*, 2013; Morita *et al.*, 2015).

The adverse outcome modeled in this study was cytotoxicity, but the approach used is not specific to any particular endpoint. The modelling approach would work equally well for any quantifiable cellular outcome, either physiological or molecular. The general approach should also be applicable to identifying relationships between chemical exposures and adverse effects in tissues *in vivo*. Indeed, verification of the applicability of *in vitro* toxicity pathway perturbations to predicting *in vivo* toxicity pathway perturbations and concomitant tissue or organism level effects is central to the utilization of toxicity pathway approaches in risk assessment.

Alternative multiple regression methods for identifying predictive subsets of proteins could also be employed. However, the cross-validated LASSO is relatively efficient computationally, and handles the inherent multi-collinearity in pathway responses reasonably well. The cross-validated elastic net method resulted in identification of the same set of proteins predictors of cytotoxicity as did the LASSO, but requires additional computation to optimize values of both α and λ to identify the combination yielding the minimum MSE. In situations where the number of candidate predictors is much greater than the number of observations, the elastic net may perform better than the LASSO (Zou and Hastie, 2005). Alternative numerical approaches for subset selection may be considered, but must be capable of adequately dealing with multi-collinearity.

Although we chose to conduct our analysis based on changes at the protein level, an analogous approach could be contemplated at the transcriptomic level. It is worth noting that four of the six proteome alterations predictive of survival that we identified represent post-translational alterations, by either phosphorylation or specific cleavages. Such alterations would not be directly detectable by changes in the transcriptome. Although the prevalence of post-translational alterations identified is to some extent a reflection of the set of endpoints selected for measurement, the ability to detect such changes directly represents a significant advantage to assessing toxicity pathway alterations at the proteome level. Functional assays of protein activity may also be useful adjuncts to the approach.

It is important to note that, in our approach, no information relating to the composition of the mixtures inducing the perturbations of biological pathways in the cells was required, or indeed used, in the construction of the predictive model. Only the protein levels and survival data following exposure to the mixtures were used in the analyses. This chemical agnosticism suggests the potential utility of this approach to estimating the relative toxicities of mixtures, including complex environmental mixtures of unknown composition. Such

complex mixtures could be used to construct a series of related mixtures suitable for identification of predictive protein networks by a variety of approaches, e.g., partial chromatographic fractionation and recombination of sub-fractions in different proportions, or spiking the mixture with varying amounts of known constituents if partial chemical characterization is available. The use of such reconstructed mixtures using the approach described in this manuscript for identification of a set of proteins whose expression is predictive of an adverse effect may be useful for estimating relative hazards and for prioritization of chemicals.

ACKNOWLEDGEMENTS

The authors wish to thank Drs. R. Woodrow Setzer Jr. and E. Sidney Hunter III for their helpful comments on this manuscript. The information in this document has been funded wholly by the U. S. Environmental Protection Agency. It has been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

References

- Committee on Toxicity Testing and Assessment of Environmental Agents, N. R. C. (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy, The National Academies Press.
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Marquez JRG, Gruber B, Lafourcade B, Leita PJ, Munkemuller T, McClean C, Osborne PE, Reineking B, Schroder B, Skidmore AK, Zurell D & Lautenbach S (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46.,
- Friedman J, Hastie T & Tibshirani R (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*, 33, 1–22. [PubMed: 20808728]
- Ge Y, Bruno M, Wallace K, Leavitt S, Andrews D, Spassova MA, Xi M, Roy A, Haykal-Coates N, Lefew W, Swank A, Winnik WM, Chen C, Woodard J, Farraj A, Teichman KY & Ross JA (2015). Systematic proteomic approach to characterize the impacts of chemical interactions on protein and cytotoxicity responses to metal mixture exposures. *J Proteome Res*, 14, 183–92. [PubMed: 25285964]
- Hastie T, Tibshirani R & Friedman J (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer New York.
- Hu BS, Zhao G, Yu HF, Chen K, Dong JH & Tan JW (2013). High expression of AP-4 predicts poor prognosis for hepatocellular carcinoma after curative hepatectomy. *Tumour Biol*, 34, 271–6. [PubMed: 23055200]
- Huang DW, Sherman BT & Lempicki RA (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37, 1–13. [PubMed: 19033363]
- Huang DW, Sherman BT & Lempicki RA (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4, 44–57. [PubMed: 19131956]
- Huang R, Wallqvist A & Covell DG (2006). Comprehensive analysis of pathway or functionally related gene expression in the National Cancer Institute's anticancer screen. *Genomics*, 87, 315–28. [PubMed: 16386875]
- Jackstadt R & Hermeking H (2014). AP4 is required for mitogen- and c-MYC-induced cell cycle progression. *Oncotarget*, 5, 7316–27. [PubMed: 25261373]
- Janky R, Verfaillie A, Imrichova H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers M, Marine JC & Aerts S (2014). iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol*, 10, e1003731. [PubMed: 25058159]

- Kaczmarek M, Timofeeva OA, Karaczyn A, Malyguine A, Kasprzak KS & Salnikow K (2007). The role of ascorbate in the modulation of HIF-1 α protein and HIF-dependent transcription by chromium(VI) and nickel(II). *Free Radic Biol Med*, 42, 1246–57. [PubMed: 17382205]
- Lammerts van Bueren K & Black BL (2012). Regulation of endothelial and hematopoietic development by the ETS transcription factor ETV2. *Curr Opin Hematol*, 19, 199–205. [PubMed: 22406820]
- Li Q, Suen TC, Sun H, Arita A & Costa M (2009). Nickel compounds induce apoptosis in human bronchial epithelial Beas-2B cells by activation of c-Myc through ERK pathway. *Toxicol Appl Pharmacol*, 235, 191–8. [PubMed: 19135467]
- Liu X, Zhang B, Guo Y, Liang Q, Wu C, Wu L, Tao K, Wang G & Chen J (2012). Down-regulation of AP-4 inhibits proliferation, induces cell cycle arrest and promotes apoptosis in human gastric cancer cells. *PLoS One*, 7, e37096. [PubMed: 22615908]
- Logue JS & Morrison DK (2012). Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy. *Genes Dev*, 26, 641–50. [PubMed: 22474259]
- Montejo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q & Bader GD (2010). GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics*, 26, 2927–2928. [PubMed: 20926419]
- Morita R, Suzuki M, Kasahara H, Shimizu N, Shichita T, Sekiya T, Kimura A, Sasaki K, Yasukawa H & Yoshimura A (2015). ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. *Proc Natl Acad Sci U S A*, 112, 160–5. [PubMed: 25540418]
- O'Brien RM (2007). A caution regarding rules of thumb for Variance Inflation Factors. *Quality & Quantity*, 41, 673–690.
- Ramos P & Bentires-Alj M (2015). Mechanism-based cancer therapy: resistance to therapy, therapy for resistance. *Oncogene*, 34, 3617–26. [PubMed: 25263438]
- Secades P, de Santa-Maria IS, Merlo A, Suarez C & Chiara MD (2015). In vitro study of normoxic epidermal growth factor receptor-induced hypoxia-inducible factor-1- α , vascular endothelial growth factor, and BNIP3 expression in head and neck squamous cell carcinoma cell lines: Implications for anti-epidermal growth factor receptor therapy. *Head Neck*, 37, 1150–62. [PubMed: 24798801]
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13, 2498–504. [PubMed: 14597658]
- Steelman LS, Chappell WH, Abrams SL, Kempf RC, Long J, Laidler P, Mijatovic S, Maksimovic-Ivanic D, Stivala F, Mazzarino MC, Donia M, Fagone P, Malaponte G, Nicoletti F, Libra M, Milella M, Tafuri A, Bonati A, Basecke J, Cocco L, Evangelisti C, Martelli AM, Montalto G, Cervello M & McCubrey JA (2011). Roles of the Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR pathways in controlling growth and sensitivity to therapy-implications for cancer and aging. *Aging (Albany NY)*, 3, 192–222. [PubMed: 21422497]
- Veldman MB, Zhao C, Gomez GA, Lindgren AG, Huang H, Yang H, Yao S, Martin BL, Kimelman D & Lin S (2013). Transdifferentiation of fast skeletal muscle into functional endothelium in vivo by transcription factor ETV2. *PLoS Biol*, 11, e1001590. [PubMed: 23853546]
- Veranth JM, Reilly CA, Veranth MM, Moss TA, Langelier CR, Lanza DL & Yost GS (2004). Inflammatory cytokines and cell death in BEAS-2B lung cells treated with soil dust, lipopolysaccharide, and surface-modified particles. *Toxicol Sci*, 82, 88–96. [PubMed: 15310859]
- Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montejo J, Shao Q, Wright G, Bader GD & Morris Q (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, 38, W214–W220. [PubMed: 20576703]
- Xinghua L, Bo Z, Yan G, Lei W, Changyao W, Qi L, Lin Y, Kaixiong T, Guobin W & Jianying C (2012). The overexpression of AP-4 as a prognostic indicator for gastric carcinoma. *Med Oncol*, 29, 871–7. [PubMed: 21336989]
- Zou H & Hastie T (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320.

Figure 1.

Network of proteomic changes predictive of cytotoxicity. Nodes shown in orange ovals are the set of proteins identified by the first level LASSO procedure as predictive of cell survival. Nodes in blue rectangles are the proteins identified in second level LASSO procedures as predictive of the proteins identified in the first level LASSO. Green arrows indicate a positive coefficient for the predictive node pointing to the predicted node, and red inhibition indicators show a negative coefficient for the predictive node pointing to the predicted node.

Figure 2.
Comparison of survival predicted by the 6-parameter model described in Table 2 with measured survival.

Table 1.

Metal treatments. Each column in the table represents a tested treatment condition containing the indicated μM concentrations of each metal.

Metal	Concentration (μM)																	
	0	30	60	75	100	0	30	60	0	30	60	0	30	60	0	30	60	0
Ni^{2+}	0	30	60	75	100	0	30	60	0	30	60	0	30	60	0	30	60	0
Cd^{2+}	0	0	0	0	0	5.6	5.6	5.6	0	0	0	5.6	5.6	5.6	0	0	0	5.6
Cr^{6+}	0	0	0	0	0	0	0	0	0.3	0.3	0.3	0.3	0.3	0.3	1.8	1.8	1.8	1.8

Table 2.

Model parameters for predictors of cell survival

Model component	Coefficient	VIF	p-value
Overall regression			<0.001
intercept	1.523		0.002
phospho-RPS6KB1	0.477	3.432	0.039
phospho-p53	−0.228	1.704	0.058
cleaved-CASP3	−0.336	3.733	0.186
phospho-MAPK8	0.031	1.860	0.857
IL-10	−0.375	2.207	0.032
HIF-1 α	−0.044	2.682	0.183

Table 3.

Predictors of cell survival predictors

Cell Survival Predictor	Predictors of Cell Survival Predictor	Coefficient	p-value
phospho-RPS6KB1	intercept	1.168	<0.001
	phospho-Akt	0.490	<0.001
	cleaved-PARP	-0.496	<0.001
	EGFR	0.078	0.330
	phospho-ERK1/2	-0.063	.0228
	GM-CSF	-0.037	0.768
	IFN- γ	-0.144	0.246
	IL-6	0.059	0.102
phospho-p53	intercept	1.693	<0.001
	TNF- α	0.486	0.132
	cleaved-PARP	-0.448	.020
	phospho-ErbB2	-1.028	<0.001
	GM-CSF	1.187	0.012
	IL-1 β	-0.790	0.076
	IL-8	-0.050	0.192
cleaved-CASP3	intercept	0.498	0.001
	phospho-GSK-3 β	-0.145	0.020
	cleaved-PARP	0.502	<0.001
	phospho-ErbB2	0.132	0.238
	phospho-ERK1/2	0.008	0.911
	IL-6	0.084	0.211
	IL-8	0.026	0.426
phospho-MAPK8	intercept	2.688	<0.001
	cleaved-PARP	-0.907	<0.001
	p53	-0.273	0.033
	GM-CSF	-0.226	0.474
	IL-2	-0.324	0.238
	IL-6	0.099	0.069
IL-10	intercept	0.138	<0.001
	phospho-GSK-3 β	-0.074	0.007
	TNF- α	0.488	<0.001
	p53	-0.076	0.052

Cell Survival Predictor	Predictors of Cell Survival Predictor	Coefficient	p-value
	GM-CSF	0.097	0.406
	IL-12p70	0.398	0.037
	IL-2	0.028	0.835
HIF-1 α			
	intercept	-2.095	<0.001
	phospho-GSK-3 β	-2.202	<0.001
	phospho-Akt	2.092	0.008
	cleaved-PARP	1.353	0.013
	EGFR	6.712	<0.001
	phospho-ErbB2	-4.915	<0.001
	phospho-ERK1/2	-0.303	0.444
	GM-CSF	3.497	<0.001
	IL-1 β	-2.709	0.002
	IL-6	0.191	0.348