Unbiased Approach for the Identification of Molecular Mechanisms Sensitive to Chemical Exposures

Alexander Suvorov[1]*, Victoria Salemme[1], Joseph McGaunn[1], Anthony Poluyanoff [1], Menna Teffera[1], Saira Amir[1,2]

Affiliations:

[1]Department of Environmental Health Sciences, School of Public Health & Health Sciences, University of Massachusetts, 686 North Pleasant Street, Amherst, MA 01003, USA.

[2]Current affiliation: Department of Biosciences, COMSATS University Islamabad, Pakistan

*To whom correspondence should be addressed: Alexander Suvorov, 173B-Goessmann, 686 North Pleasant Street, Amherst, MA 01003, USA, asuvorov@schoolph.umass.edu

All authors declare to competing financial interests

Abstract

Background: Targeted methods that dominated toxicological research until recently did not allow for screening of all molecular changes involved in toxic response. Therefore, it is difficult to infer if all major mechanisms of toxicity have already been discovered, or if some of them are still overlooked.

Objectives: To identify molecular mechanisms sensitive to chemical exposures in an unbiased manner.

Methods: We used data on 641,516 unique chemical-gene interactions from the Comparative Toxicogenomic Database. Only data from high-throughput gene expression experiments with human, rat or mouse cells/tissues were extracted. The total number of chemical-gene interactions was calculated for every gene, and used as a measure of gene sensitivity to chemical exposures. These values were further used in enrichment analyses to identify molecular mechanisms sensitive to chemical exposures.

Results: Remarkably, use of different input subsets with non-overlapping lists of chemical compounds identified largely the same genes and molecular pathways as most sensitive to chemical exposures, indicative of an unbiased nature of our analysis. One of the most important findings of this study is that almost every known molecular mechanism may be affected by chemical exposures. Predictably, xenobiotic metabolism pathways and mechanisms of cellular response to stress and damage were among the most sensitive. Additionally, our analysis identified a range of highly sensitive molecular pathways, which are not widely recognized by modern toxicology as major targets of toxicants, including lipid metabolism pathways, longevity regulation cascade and cytokine mediated signaling.

Discussion: Molecular mechanisms identified as the most sensitive to chemical exposures are relevant for significant public health problems, such as aging, cancer, metabolic and autoimmune disease. Thus, public health system will likely benefit from future research focus on these sensitive molecular mechanisms. Additionally, approach used in this study may guide identification of priority adverse outcome pathways (AOP) for in-vitro and in-silico toxicity testing methods.

Introduction

The total burden of disease costs associated with exposures to environmental chemicals likely exceeds 10% of the global domestic product (Grandjean and Bellanger 2017). The number of new chemicals is rising rapidly, with the Chemical Abstract Service Registry growing from 20 million to 156 million chemicals between 2002 and 2019 (Escher et al. 2020). This situation poses a significant challenge for regulatory toxicology and requires the development of new, rapid, cost-efficient, and reliable methods of toxicity testing. Indeed, in recent decades, regulatory toxicology has begun a revolutionary transformation that was outlined in a strategic vision in 2007 (NRC 2007). One important component of this transformation consists in a transition from outdated animal-based tests to *in-vitro* tests, targeting sensitive molecular mechanisms – toxicity pathways or adverse outcome pathways (AOP) (Ankley et al. 2010; Haynes 2010; NRC 2007; OECD 2017; Vinken 2013). To ensure comprehensive coverage of all possible adverse effects using this approach, the battery of *in-vitro* tests should include all pathways that mechanistically connect exposures with adverse health outcomes. Given that the transformation of toxicology is going rapidly (EPA 2013; NASEM 2017), it is critically important to identify all of these pathways, or at least all major toxicity pathways.

Today, the toxicological community recognizes many molecular mechanisms responsible for a significant portion of all toxicity events, such as the formation of DNA adducts (Himmelstein et al. 2009; Jarabek et al. 2009), oxidative stress (Sies et al. 2017), activation of the AhR receptor (Strapacova et al. 2018), disruption of estrogenic signaling (Singleton and Khan 2003) and others. All these currently well-recognized mechanisms of toxicity were discovered during the period of toxicology development when high-throughput methods were unavailable. Therefore, it is impossible to infer if all major mechanisms of toxicity were already discovered, or if some of them were overlooked. Here, we hypothesized that data from toxicological -omics experiments rapidly accumulating in publicly accessible databases may help to answer this question.

Each chemical compound induces specific changes in gene expression in an exposed biological system. Some of these changes may be unique for a compound, but other may overlap with changes induced by other compound(s). Thus, by overlaying changes in genomic profiles resulting from many different exposures, we may identify genes and associated pathways that most commonly respond to exposure (Fig. 1A). In this study, we use publicly available data from the Comparative Toxicogenomic Database (CTD) (Davis et al. 2019) on changes in gene expression in response to a broad range of chemical compounds to identify, in an unbiased manner, the molecular mechanisms most sensitive to chemical exposure.

Materials and Methods

Database Creation

A database was created by extracting data on chemical-gene interactions from the CTD (Davis et al. 2019) on 08.24.2018 using the following filtering criteria. First, data was extracted only from experiments that used high-throughput approaches for gene expression analysis (microarray, RNA-seq). In addition, we selected data only from experiments that used human, rat or mouse cells, tissue or the whole organism for gene expression analysis in *in vitro* and *in vivo* studies. Further, we removed 1,846 olfactory receptor genes from the database, as the nomenclatures assigned to these genes in mouse, rat and human are different and independent. The resulting database included 641,516 entries, each representing one chemical-gene interaction (significant change in expression of a gene in response to exposure to a chemical compound) reported in a published study.
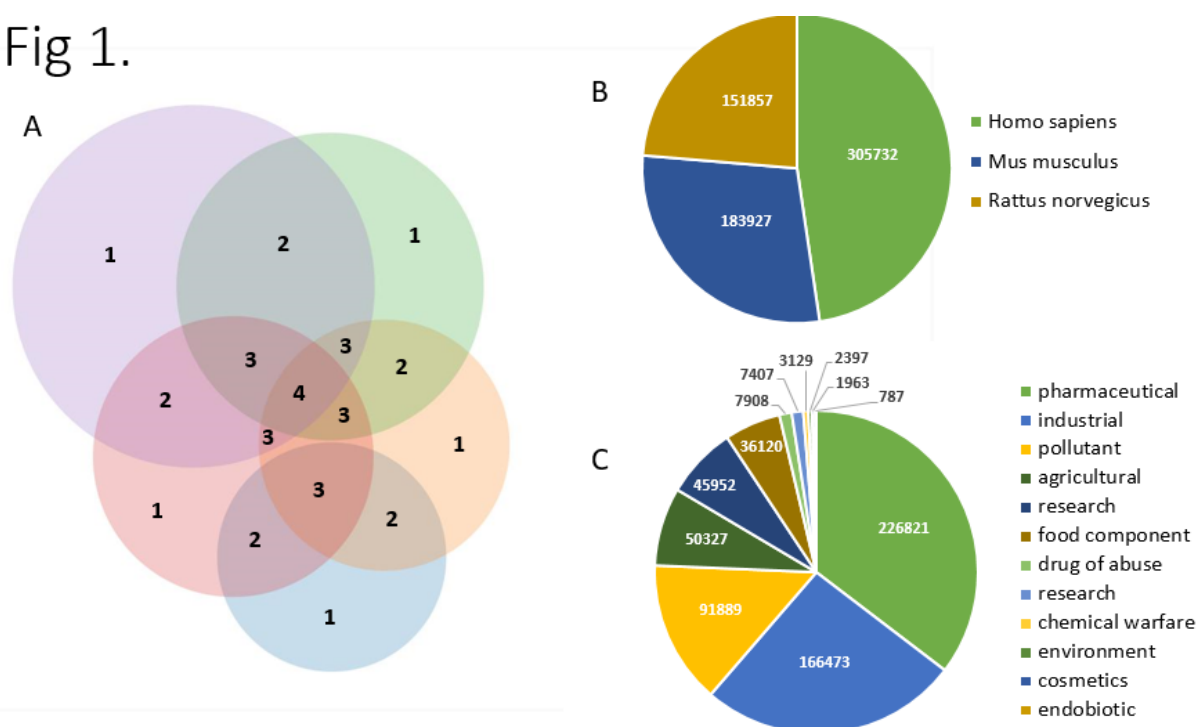
Fig. 1. Approach of this study for the unbiased identification of genes sensitive to chemical exposures. A - Each circle represents a group of genes differentially expressed in response to one chemical in one experiment. Genes sensitive to exposure to more than one chemical populate the zones where circles overlap. The numbers of chemical-gene interactions shown in every zone reflect sensitivity of corresponding genes to chemical exposures. B and C - The distribution of chemical-gene interactions by model organism and by annotation terms for chemical usage respectively in the analyzed database. Individual chemicals were annotated with up to 3 annotation terms in accordance with their most common uses.

At the next step, all chemical compounds in the database were manually annotated to identify their major uses. To obtain this information a search was conducted using the name of the chemical as the key word in Wikipedia, PubChem and PubMed. Based on the obtained descriptive annotation, every compound was categorized in accordance with the following terms: pharmaceutical, recreational drug, research, warfare, endobiotic, agricultural, cosmetics, environment, food components, industrial, and pollutant. The term "pharmaceutical" was used for all drugs, including prescription, over the counter, and traditional medicines. Chemicals that are currently going through preclinical and clinical trials were also categorized as "pharmaceutical". The term "research" was used for chemicals, which are mainly used for research purposes, such as components of molecular biology or analytical chemistry protocols. Some of these chemicals, such as protein kinase inhibitors, for example, have been or are currently tested as candidate pharmaceutical drugs. Given that information about preclinical testing is not always readily available (Piller 2020), we assigned both "pharmaceutical" and "research" annotation terms to them. The term "recreational drug" was used for compounds currently used as recreational drugs, their major metabolites, and compounds that are not currently used as recreational drugs but which possess these properties making them potential recreational drugs. The term "warfare" was used for compounds used as chemical weapons. The "endobiotic" annotation term was assigned to molecules synthetized in mammalian organisms as normal components of healthy physiology, such as cholesterol, hormones, bile acids and other. The term "agricultural" was used for fertilizers and pesticides. The term "cosmetics" was used for a broad range of chemical compounds used in cosmetics and perfumes. The annotation term "environment" was used for oxygen and ozone. The

term "food component" was used for a broad range of compounds that can be found in food, such as dietary nutrients, food additives, byproducts of food processing and others. The term "industrial" was used to annotate a broad range of chemicals used as intermediates or final products in a variety of industrial processes, applications, and final products resulting from industrial process. Finally, the term "pollutant" was used for these chemicals, which do not have current uses but are produced by natural processes and human activity such as products of incomplete combustion of organic material, air and water pollutants, toxins produced by algal blooms and similar. For chemicals that fell into multiple categories, we used up to three annotation terms reflecting the most common uses. All authors of the manuscript annotated an equal numbers of chemicals and all annotations were than checked by the corresponding author to insure uniformity of annotation approaches.

Bioinformatic analysis

We hypothesized that due to the presence of a high number of pharmaceuticals (40%) and other biologically active compounds in the database, some findings may be clouded by the fact that many chemical compounds were intentionally designed to target important molecular pathways in mammalian cells. To differentiate between unintended molecular responses to chemical exposure and compare with responses to biologically active chemicals, we split the original database into 2 subsets. All chemicals annotated with one or more of the following terms: pharmaceutical, drug of abuse, research, warfare and endobiotic, - were combined into a database subset, named "Bioactive" (BA). The second subset "Non-bioactive" (NBA) included all chemicals, which did not have these listed terms in their annotations. Thus, this second subset included only chemicals annotated with the following terms: agricultural, cosmetics, environment, food components, industrial, and pollutant.

For every gene in a database, we calculated the total number of chemical-gene interactions. Each individual interaction used for this calculation was one line in the database and represented a unique combination of the original study, biological model and chemical compound. Similarly, for each gene we calculated the total number of activating (gene expression increases in response to exposure) and suppressive (gene expression decreases in response to exposure) interactions.

The resulting lists of genes were used in bioinformatic tools developed for enrichment analysis of gene expression data. The numbers of chemical-gene interactions were used in place of differential expression values. The short lists of top genes with highest numbers of interactions were uploaded to Metascape (Tripathi et al. 2015), and enriched functional terms were explored with default settings. The lists of all genes with their respective chemical-gene interactions numbers were further used for gene set enrichment analysis (GSEA). This approach is particularly effective for the identification of biologically significant changes in activity of the whole pathway or other biological category, as its enrichment function stems from all, even small, changes in multiple members of a gene set (pathway) (Boylan et al. 2015). The details of the method and statistical approaches used by GSEA are described elsewhere (Mootha et al. 2003; Subramanian et al. 2005), see also Supplementary materials 1. Originally, GSEA was developed to characterize the cumulative shift of all genes in a particular pathway towards an increase or decrease of expression. As such, it was designed for an input in which values of gene expression changes have semi-symmetrical distribution, where comparable numbers of genes are up- and downregulated. To prepare datasets suitable for GSEA, we subtracted the same number from the values of chemical-gene interactions for every gene, to achieve equal negative and positive area under the curve for these values. The resulting gene lists were uploaded to GSEA and were analyzed against three independent open access databases of pathways: Hallmark (Liberzon et al. 2015), Reactome (Fabregat et al. 2018; Jassal et al. 2020) and KEGG (Kanehisa and Goto 2000). The following

stringent criteria were used to identify molecular mechanisms most and list sensitive to chemical exposures: normalized enrichment score (NES) ≥ 1.9 or ≤ -1.9 and FDR q ≤ 0.05.
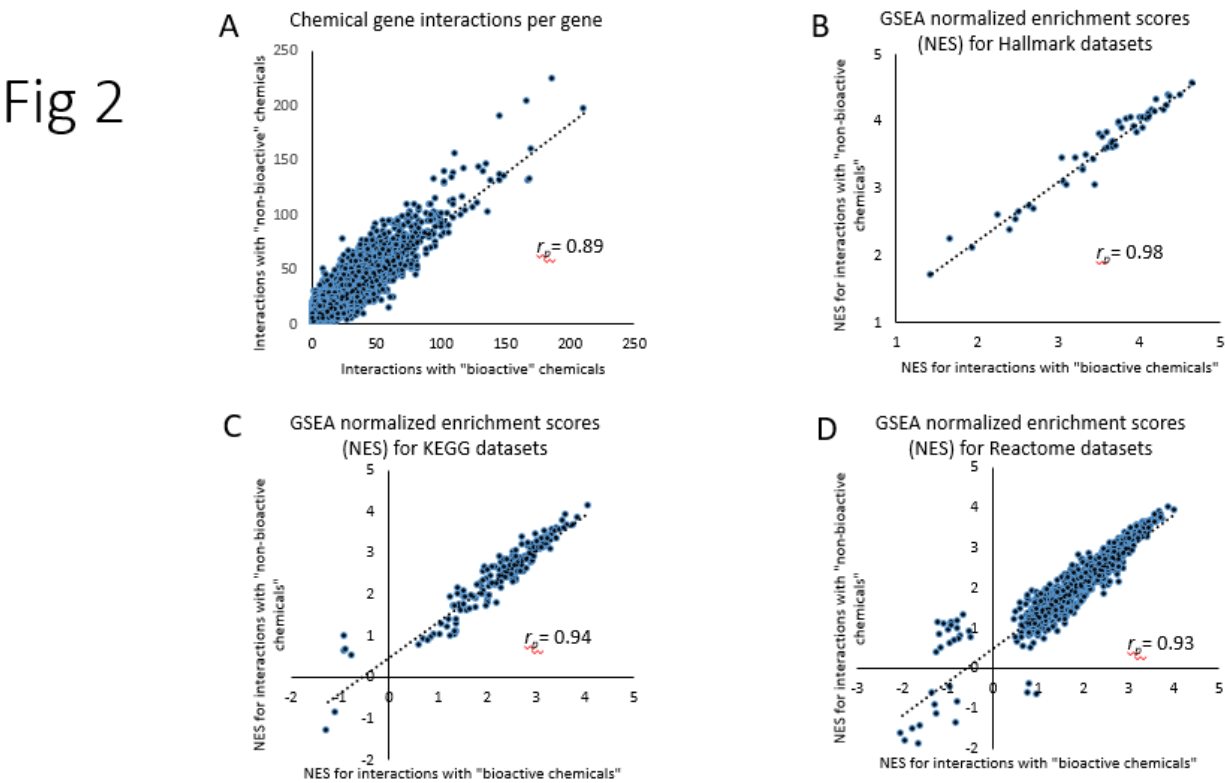

Fig. 2. Comparison of sensitivity of genes and pathways to "biologically active" (BA) and "non-biologically active" (NBA) chemicals. A – Number of chemical gene interactions per gene for BA and NBA subsets of the database. B-D – GSEA enrichment of pathways for BA and NBA subsets of the database. GSEA was done for Hallmark (B), KEGG (C) and Reactome (D) datasets. r_p – Pearson correlation.

Results

Database characteristics

The database used in this study (Supplementary Data File S1) consists of 641,516 individual chemical-gene interactions reported in 2,180 original studies using high-throughput gene expression analysis only. All these studies were done with 3 mammalian species (humans, mice, rats) or their cells and tissues (Fig. 1B). Chemical-gene interactions included 26,704 unique genes and 1,265 unique chemicals. Distribution of chemicals in accordance to their usage annotation terms is shown in Fig. 1C. This database was further split into 2 subsets defined by a purposeful design of chemicals to interact with essential molecular pathways. The bioactive (BA) subset included chemical compounds intentionally designed to interact with essential molecular mechanisms in mammalian cells. The non-bioactive (NBA) subset contained all other chemical compounds. Characteristics of these subsets are shown in Table 1.

Table 1. Characteristics of the whole database, NBA and BA subsets

Parameter	Whole database	NBA subset	BA subset
Annotation terms	Pharmaceutical, recreational drug, research, warfare, endobiotic, agricultural, cosmetics, environment, food components, industrial, pollutant	Excluding terms: pharmaceutical, recreational drug, research, warfare, endobiotic	Containing terms: pharmaceutical, recreational drug, research, warfare, endobiotic
Chemical-gene interactions - Homo sapiens - Mus musculus - Rattus norvegicus - Total:	 305,732 183,927 151,857 **641,516**	 119,410 121,447 87,203 **328,060**	 186,322 62,480 64,654 **313,456**
Genes	26,704	24,571	23,492
Chemicals	1,265	554	712
Source studies	2,180	1,295	1,092

Genes and pathways sensitive to "bioactive" and "non-bioactive" compounds

To test if our identification of molecular mechanisms sensitive to chemical exposures might be clouded by including chemicals intentionally designed to target specific molecular pathways in mammalian cells, we compared the number of chemical-gene interactions for every gene in the BA and NBA subsets of the database. Values of chemical-gene interactions were similar in both subsets, with Pearson correlation = 0.89 (Fig. 2A).

We further compared molecular pathways enriched with genes sensitive to "bioactive" and "non-bioactive" chemicals. To do so, we conducted gene-set enrichment analysis (GSEA) with gene lists from BA and NBA subsets and compared obtained normalized enrichment scores (NES) for Hallmark, KEGG and Reactome pathways. The results demonstrate a high level of similarity between pathways enriched with genes from both subsets (Fig. 2B-D). In these graphs, discordant distribution between BA and NBA datasets are mostly seen for pathways with non-significant enrichment and low positive and negative NES (≥ 1.5 or ≤ -1.5). The Pearson correlation for pathway enrichment in BA and NBA datasets was 0.97 for Hallmark, 0.94 for KEGG and 0.93 for Reactome collections.

Given that very similar lists of highly sensitive genes and molecular pathways were identified using 2 subsets of the database with non-overlapping lists of chemicals, we concluded that the number of chemical-gene interactions in both subsets is high enough to mask unique responses to individual compounds with specific bioactivity properties and subsequent analysis was done on the entire database.

Genes and pathways most sensitive to chemical exposures

The number of chemical-gene interactions in the database varied from 1 to 411 (Fig. 3A). The top genes with more than 200 chemical gene interactions are shown in Table 2. To identify molecular pathways most sensitive to chemical exposures, we first used the list of 302 genes with 120 or more interactions (Fig. 3A) for Metascape analysis. This threshold was arbitrary selected to produce reasonable input for Metascape. Top enriched categories were dominated by biological terms related to response to chemical exposure ("*response to toxic substance*", "*response to inorganic substance*", "*cellular response to drug*", "*response to xenobiotic stimulus*" and other) and response to damage ("*apoptotic signaling pathway*", "*positive regulation of cell death*",

"response to wounding", "vasculature development", "response to unfolded protein", "p53 signaling pathway", "HIF1 pathway" (=response to hypoxia (Simmons 2009)), and other) (Fig. 3B). Also among the top enriched pathways, were cytokine-mediated signaling (*"cytokine-mediated signaling pathway", "TNF signaling pathway"*) and *"regulation of lipid metabolic process".*
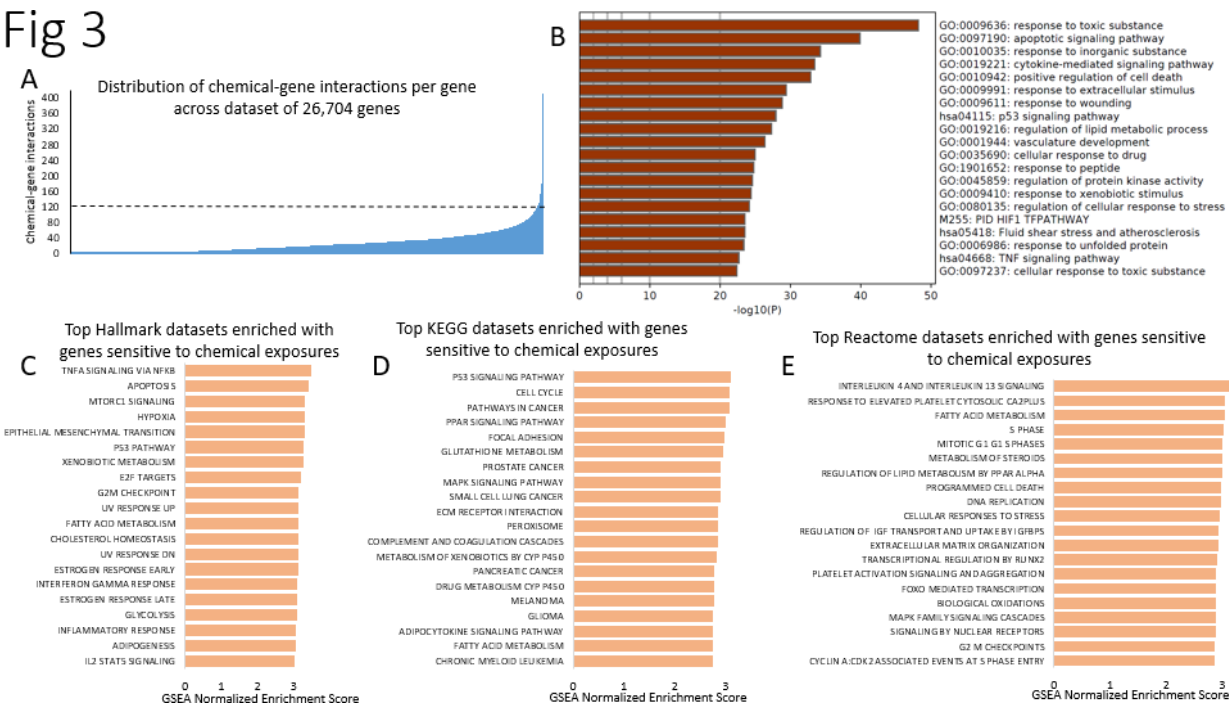

Fig. 3. Enrichment of molecular categories with genes sensitive to chemical exposures. A – Distribution of chemical-gene interaction numbers across the dataset of 26,704 genes. The dashed line illustrates the cutoff for the top genes with > 120 chemical gene interactions used for Metascape analysis. B - Metascape enrichment of biological categories by 302 genes with > 120 chemical-gene interactions. C-E - Top GSEA enriched categories for Hallmark (C), KEGG (D) and Reactome (E) collections of datasets, based on the list of 26,704 genes and their numbers of chemical-gene interactions.

In GSEA analysis with Hallmark collection all 50 datasets were positively enriched and 49 passed our criteria of significance. The top highly enriched datasets included *"TNF alpha signaling via NF-kappa B"* (NES = 3.47), *"apoptosis"* (NES = 3.42), *"mTORC1 signaling"* (NES = 3.32), *"hypoxia"* (NES = 3.30) and *"epithelial/mesenchymal transition"* (NES = 3.29) (Fig. 3C, 4A-D, Supplementary Table S1). Out of 186 KEGG datasets, none passed our threshold of significance for negative enrichment. One hundred forty-five KEGG datasets were significantly positively enriched. The most highly enriched datasets were *"p53 signaling pathway"* (NES = 3.09) (Fig. 5), *"cell cycle"* (NES = 3.08), *"pathways in cancer"* (NES = 3.08), *"PPAR signaling pathway"* (NES = 2.99) (Fig. 6) and *"focal adhesion"* (NES = 2.97) (Fig 3D, 4E-H, Supplementary Table S1). Out of 1499 Reactome datasets one, *"beta defensins"*, was negatively enriched with NES = -2.0 and FDR q ≤ 0.007. In total 689 Reactome datasets were significantly positively enriched. The most highly enriched datasets were *"interleukin 4 and interleukin 13 signaling"* (NES = 3.16), *"response to elevated platelet cytosolic Ca2+"* (NES = 3.04), *"fatty acid metabolism"* (NES = 3.04), *"S phase"* (NES = 3.03), and *"mitotic G1 phase and G1/S transition"* (NES = 3.01) and (Fig 3E, 4I-L, Supplementary Table S1).

Table 2. Top genes with more than 200 chemical gene interactions.

Gene ID	Interactions	Gene name
HMOX1	411	heme oxygenase 1
CDKN1A	408	cyclin dependent kinase inhibitor 1A
NQO1	371	NAD(P)H quinone dehydrogenase 1
CYP1A1	336	cytochrome P450 family 1 subfamily A member 1
EGR1	330	early growth response 1
GADD45A	302	growth arrest and DNA damage inducible alpha
CCL2	300	C-C motif chemokine ligand 2
GDF15	285	growth differentiation factor 15
FOS	282	Fos proto-oncogene, AP-1 transcription factor subunit
JUN	282	Jun proto-oncogene, AP-1 transcription factor subunit
MYC	277	MYC proto-oncogene, bHLH transcription factor
ATF3	273	activating transcription factor 3
DDIT3	273	DNA damage inducible transcript 3
CCND1	271	cyclin D1
CYP1B1	267	cytochrome P450 family 1 subfamily B member 1
SLC7A11	260	solute carrier family 7 member 11
GCLM	248	glutamate-cysteine ligase modifier subunit
GCLC	243	glutamate-cysteine ligase catalytic subunit
SERPINE1	243	serpin family E member 1
ALDH1A1	239	aldehyde dehydrogenase 1 family member A1
VEGFA	239	vascular endothelial growth factor A
IL1B	233	interleukin 1 beta
SRXN1	233	sulfiredoxin 1
ASNS	232	asparagine synthetase (glutamine-hydrolyzing)
DUSP1	232	dual specificity phosphatase 1
TXNRD1	227	thioredoxin reductase 1
CXCL8	223	C-X-C motif chemokine ligand 8
TRIB3	223	tribbles pseudokinase 3
CCN2	219	cellular communication network factor 2
HSPA1A	219	heat shock protein family A (Hsp70) member 1A
DDIT4	215	DNA damage inducible transcript 4
EPHX1	210	epoxide hydrolase 1
CDK1	207	cyclin dependent kinase 1
SLC3A2	206	solute carrier family 3 member 2
HMGCS1	202	3-hydroxy-3-methylglutaryl-CoA synthase 1
CXCL1	201	C-X-C motif chemokine ligand 1
IGF1	201	insulin like growth factor 1

In GSEA analysis with Hallmark collection all 50 datasets were positively enriched and 49 passed our criteria of significance. The top highly enriched datasets included "*TNF alpha signaling via NF-kappa B*" (NES = 3.47), "*apoptosis*" (NES = 3.42), "*mTORC1 signaling*" (NES = 3.32), "*hypoxia*" (NES = 3.30) and "*epithelial/mesenchymal transition*" (NES = 3.29) (Fig. 3C, 4A-D, Supplementary Table S1). Out of 186 KEGG datasets, none passed our threshold of significance for negative enrichment. One hundred forty-five KEGG datasets were significantly positively enriched. The most highly enriched datasets were "*p53 signaling pathway*" (NES = 3.09) (Fig. 5), "*cell cycle*" (NES = 3.08), "*pathways in cancer*" (NES = 3.08), "*PPAR signaling pathway*" (NES = 2.99) (Fig. 6) and "*focal adhesion*" (NES = 2.97) (Fig 3D, 4E-H, Supplementary Table S1). Out of 1499 Reactome datasets one, "*beta defensins*", was negatively enriched with NES = -2.0 and FDR q ≤ 0.007. In total 689 Reactome datasets were significantly positively enriched. The most highly enriched datasets were "*interleukin 4 and interleukin 13 signaling*" (NES = 3.16), "*response to elevated platelet cytosolic Ca2+*" (NES = 3.04), "*fatty acid metabolism*" (NES = 3.04), "*S phase*"

(NES = 3.03), and "*mitotic G1 phase and G1/S transition*" (NES = 3.01) and (Fig 3E, 4I-L, Supplementary Table S1).
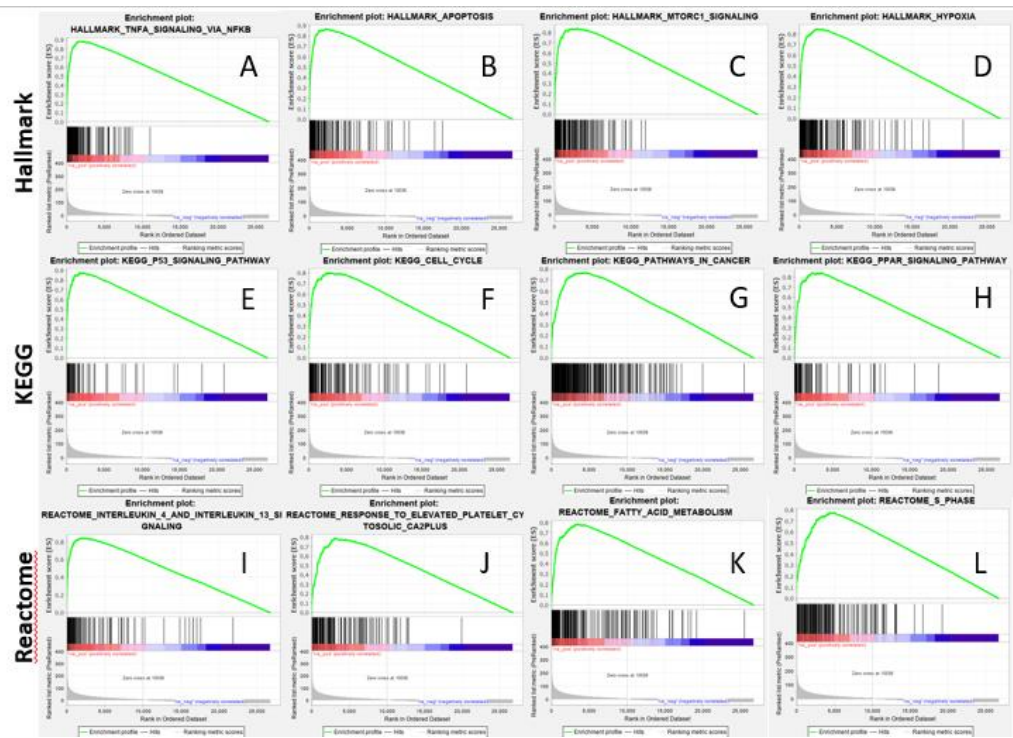

Fig. 4. GSEA enrichment plots for top enriched categories from Hallmark (A-D), KEGG (E-H) and Reactome (I-L) collections, based on the list of 26,704 genes and their numbers of chemical-gene interactions. A - TNF alpha signaling via NF-kappa B, B - apoptosis, C - mTORC1 signaling D – hypoxia, E - p53 signaling pathway, F - cell cycle, G - pathways in cancer, H - PPAR signaling pathway, I - interleukin 4 and interleukin 13 signaling, J - response to elevated platelet cytosolic Ca2+, K - fatty acid metabolism, L - S phase.
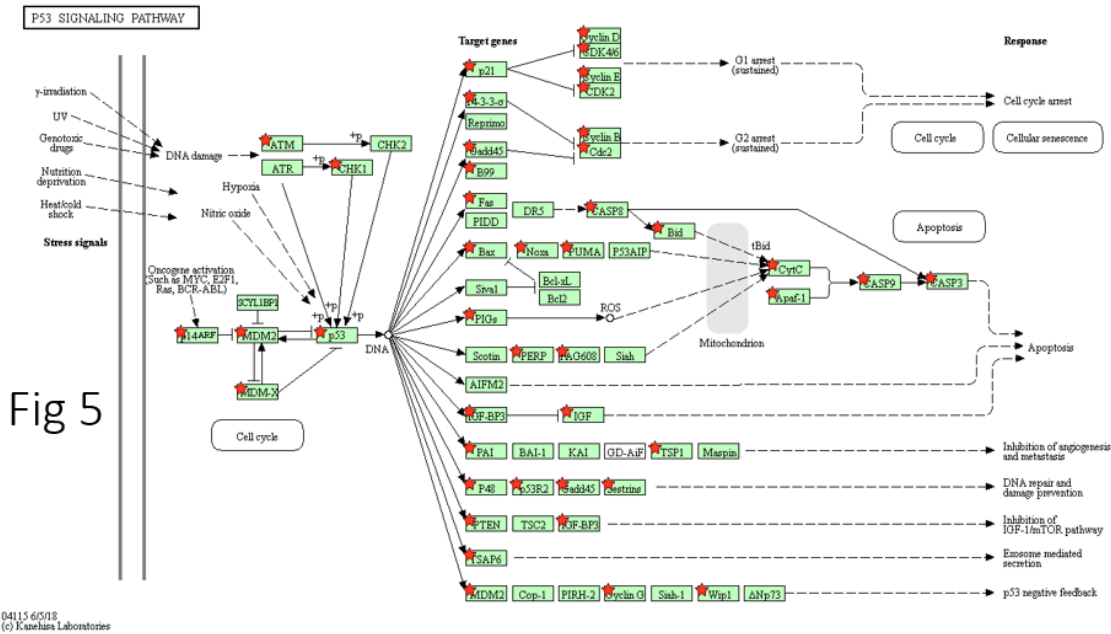

Fig. 5. P53 signaling pathway highly enriched (GSEA NES = 3.01, FDR q < 0.0005) with genes sensitive to chemical exposures. Red stars indicate genes that contributed to enrichment.

Top genes and pathways activated or suppressed by chemicals

At the next step, we identified the number of activating or suppressive chemical-gene interactions for every gene in the database. The top 20 genes with the highest numbers of activating and suppressive interactions are shown in Table 3. Among the highest scoring genes for suppressive interactions there were important members of the GH-IGF signaling cascade (GHR, IGFBP3, and IGF1), cytokines (CXCL8, CXCL12, CCL2), cyclins (CCNB1, CDK1, CCNA2, CCND1), lipid metabolism genes (THRSP, HMGCS1, FASN) and more. Among the highest scoring genes for activating interactions, there were components of the AP-1 transcription factor (FOS, JUN, ATF3), and other transcription factors - regulators of growth, proliferation, and apoptosis (MYC, EGR1), proteins that promote cell cycle arrest and/or apoptosis in response to stress stimuli (GADD45A, DDIT3, CDKN1A), microsomal (CYP1B1, CYP1A1, HMOX1) and cytosolic (NQO1, SRXN1, TXNRD1) oxidoreductases, and enzymes of glutathione synthesis (SLC7A11, GCLC, GCLM), including 2 rate-limiting enzymes (CCL2 and GDF15).

Table 3. Top 20 genes with highest number of suppressive or activating chemical-gene interactions.

Top genes with highest number of suppressive interactions				Top genes with highest number of activating interactions			
Gene ID	Interaction		Gene name	Gene ID	Interaction		Gene name
	-	+			-	+	
CCND1	117	126	cyclin D1	HMOX1	46	333	heme oxygenase 1
CCL2	116	163	C-C motif chemokine ligand 2	NQO1	32	312	NAD(P)H quinone dehydrogenase 1
NREP	108	17	neuronal regeneration related protein	CDKN1A	63	291	cyclin dependent kinase inhibitor 1A
IGF1	107	59	insulin like growth factor 1	GDF15	30	229	growth differentiation factor 15
EGR1	97	187	early growth response 1	CYP1A1	83	228	cytochrome P450 family 1 subfamily A member 1
CCNA2	95	72	cyclin A2	GADD45A	41	222	growth arrest and DNA damage inducible alpha
CDK1	94	86	cyclin dependent kinase 1	ATF3	35	209	activating transcription factor 3
VEGFA	92	119	vascular endothelial growth factor A	SRXN1	11	203	sulfiredoxin 1
FASN	91	47	fatty acid synthase	GCLM	33	197	glutamate-cysteine ligase modifier subunit
CXCL12	91	69	C-X-C motif chemokine ligand 12	DDIT3	45	196	DNA damage inducible transcript 3
DDC	90	23	dopa decarboxylase	SLC7A11	48	190	solute carrier family 7 member 11
RRM2	90	63	ribonucleotide reductase regulatory subunit M2	CYP1B1	53	189	cytochrome P450 family 1 subfamily B member 1
IGFBP3	90	73	insulin like growth factor binding protein 3	JUN	53	188	Jun proto-oncogene, AP-1 transcription factor subunit
HMGCS1	90	94	3-hydroxy-3-methylglutaryl-CoA synthase 1	**EGR1**	97	187	early growth response 1
GHR	87	19	growth hormone receptor	TXNRD1	19	185	thioredoxin reductase 1
THRSP	87	29	thyroid hormone responsive	SERPINE1	35	181	serpin family E member 1
CCNB1	87	86	cyclin B1	FOS	72	180	Fos proto-oncogene, AP-1 transcription factor subunit
TOP2A	85	72	DNA topoisomerase II alpha	GCLC	49	172	glutamate-cysteine ligase catalytic subunit
CXCL8	84	122	C-X-C motif chemokine ligand 8	MYC	77	165	MYC proto-oncogene, bHLH transcription factor
SQLE	84	67	squalene epoxidase	**CCL2**	116	163	C-C motif chemokine ligand 2

The distribution of all genes in the database in accordance to the numbers of suppressive and activating chemical-gene interactions is shown in Fig. 7A. This figure illustrates that most of the genes sensitive to chemical exposures have both suppressive and activating interactions. To identify genes, showing differential expression in one direction in response to chemical exposure, the ratio of activating to suppressive interactions for genes (RASIG) with more than 50 total chemical-gene interactions was calculated. Application of this threshold resulted in a list of 799 genes. Log2 RASIG for these genes has semi normal distribution, indicating that most genes sensitive to a variety of chemical exposures undergo expression changes in both directions (Fig. 7B). The top 20 genes on each margin of this distribution are shown in Table 4.
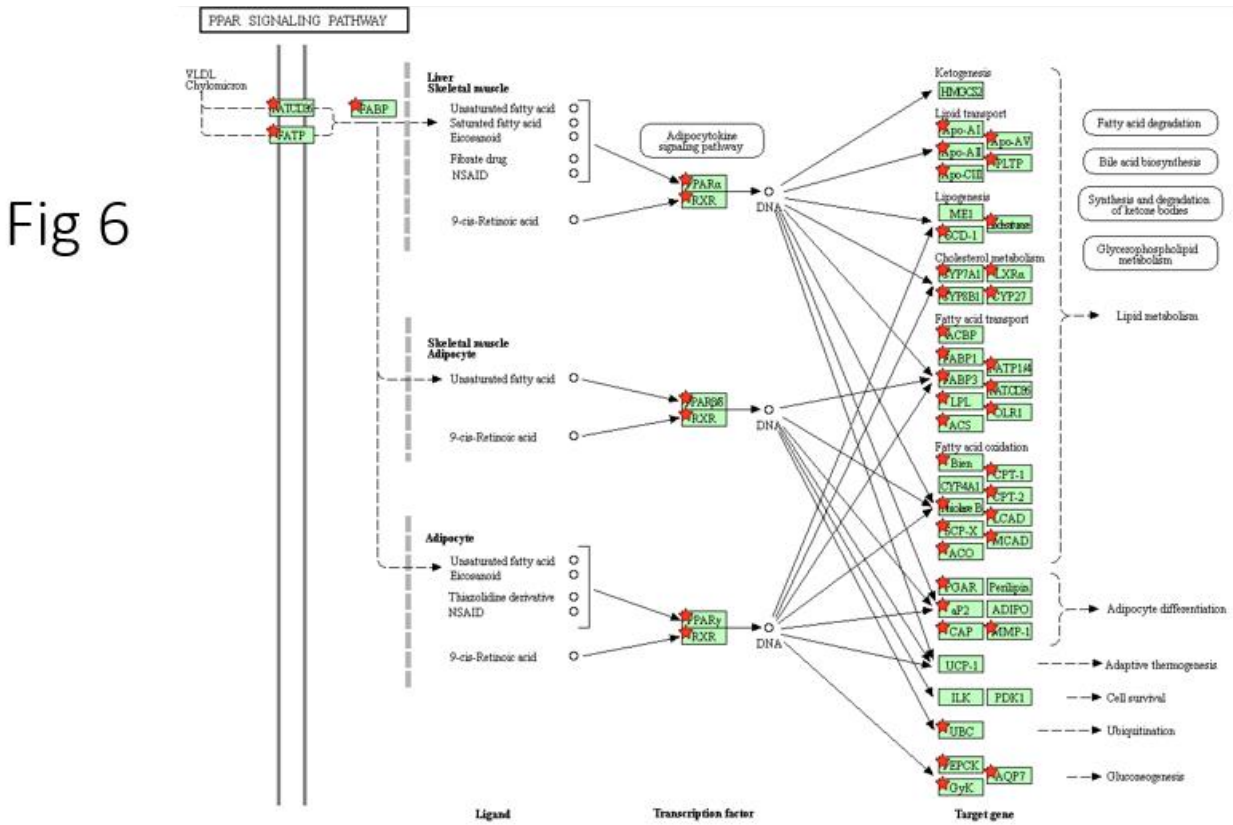

Fig. 6. PPAR signaling pathway highly enriched (GSEA NES = 2.99, FDR q < 0.0005) with genes sensitive to chemical exposures. Red stars indicate genes that contributed to enrichment.

Next, we used GSEA to identify which pathways are mostly regulated by chemical exposures in the direction of activation, suppression, or both. Pathway enrichment were identified for two lists of genes, one including only suppressive and another including only activating interactions. Pearson correlations (r_p) for enrichment (NES) of pathways by the 2 gene lists were as follows: 0.80 for Hallmark (Fig. 7C), 0.79 for KEGG (Fig. 7D) and 0.74 for Reactome datasets (Fig. 7E). Thus, most pathways sensitive to exposures are sensitive to both activation and suppression by chemical compounds. The top pathways highly enriched by both gene lists are shown in Table 5. These pathway include response to chemical exposures and associated stress (*"Xenobiotic metabolism", "UV response up", "Hypoxia", "Apoptosis", "Glutathione metabolism" and "P53 signaling pathway"*); immune response (*"Interferon gamma response", "TNF alpha signaling via NF-kappa B", IL4 and IL13 signaling", "Focal adhesion", and "Response to elevated platelet Ca2+"*); regulation of lipid metabolism (*"Fatty acid metabolism", "mTORC1 signaling", "PPAR signaling pathway", "Regulation of lipid metabolism by PPAR alpha"*); response to xenoestrogens

Table 4. Top 20 genes with highest positive and negative ratios of activating to suppressive chemical-gene interactions out of 799 genes most sensitive to chemical exposures (>50 chemical-gene interactions).

Gene Symbol	Description	Suppressive	Activating	Activating/ Suppressive Ratio
OTC	ornithine carbamoyltransferase	51	3	-17
IGFALS	IGF binding protein acid labile subunit	60	7	-8.6
SLC2A2	solute carrier family 2 member 2	74	10	-7.4
SLC22A7	solute carrier family 22 member 7	57	8	-7.1
RGN	regucalcin	71	11	-6.5
NREP	neuronal regeneration related protein	108	17	-6.3
ABCC6	ATP binding cassette subfamily C member 6	56	10	-5.6
HAL	histidine ammonia-lyase	77	15	-5.1
GNMT	glycine N-methyltransferase	66	13	-5.1
GAS1	growth arrest specific 1	55	12	-4.6
PAH	phenylalanine hydroxylase	55	12	-4.6
GHR	growth hormone receptor	87	19	-4.6
CPS1	carbamoyl-phosphate synthase 1	54	12	-4.5
CA3	carbonic anhydrase 3	79	18	-4.4
SLC10A1	solute carrier family 10 member 1	69	16	-4.3
DDC	dopa decarboxylase	90	23	-3.9
ST6GAL1	ST6 beta-galactoside alpha-2,6-sialyltransferase 1	58	15	-3.9
PKLR	pyruvate kinase L/R	77	20	-3.9
NR1H4	nuclear receptor subfamily 1 group H member 4	57	15	-3.8
FGFR2	fibroblast growth factor receptor 2	55	15	-3.7
GDF15	growth differentiation factor 15	30	229	7.6
EIF4EBP1	eukaryotic translation initiation factor 4E-BP1	7	54	7.7
PLA2G12A	phospholipase A2 group XIIA	9	71	7.9
ABHD4	abhydrolase domain containing 4	8	68	8.5
METRNL	meteorin like, glial cell differentiation regulator	7	60	8.6
SLC3A2	solute carrier family 3 member 2	19	163	8.6
ZMAT3	zinc finger matrin-type 3	8	69	8.6
AKR1B10	aldo-keto reductase family 1 member B10	6	56	9.3
PLK3	polo like kinase 3	8	77	9.6
TXNRD1	thioredoxin reductase 1	19	185	9.7
NQO1	NAD(P)H quinone dehydrogenase 1	32	312	9.8
ZFAND2A	zinc finger AN1-type containing 2A	9	88	9.8
GSR	glutathione-disulfide reductase	11	131	11.9
HSPA6	heat shock protein family A (Hsp70) member 6	6	72	12
SQSTM1	sequestosome 1	13	156	12
MAFK	MAF bZIP transcription factor K	5	64	12.8
MAFF	MAF bZIP transcription factor F	11	149	13.5
STX3	syntaxin 3	4	60	15
SRXN1	sulfiredoxin 1	11	203	18.5
FTL	ferritin light chain	1	74	74

("Estrogen response late", Estrogen response early"); and cell cycle, development and cancer-related pathways (*"Pancreatic cancer", Pathways in cancer", "Cell cycle", "Mitotic G1, G1/S phases", "Epithelial-mesenchymal transition", "Focal adhesion", "P53 signaling pathway", "mTORC1 signaling"*).
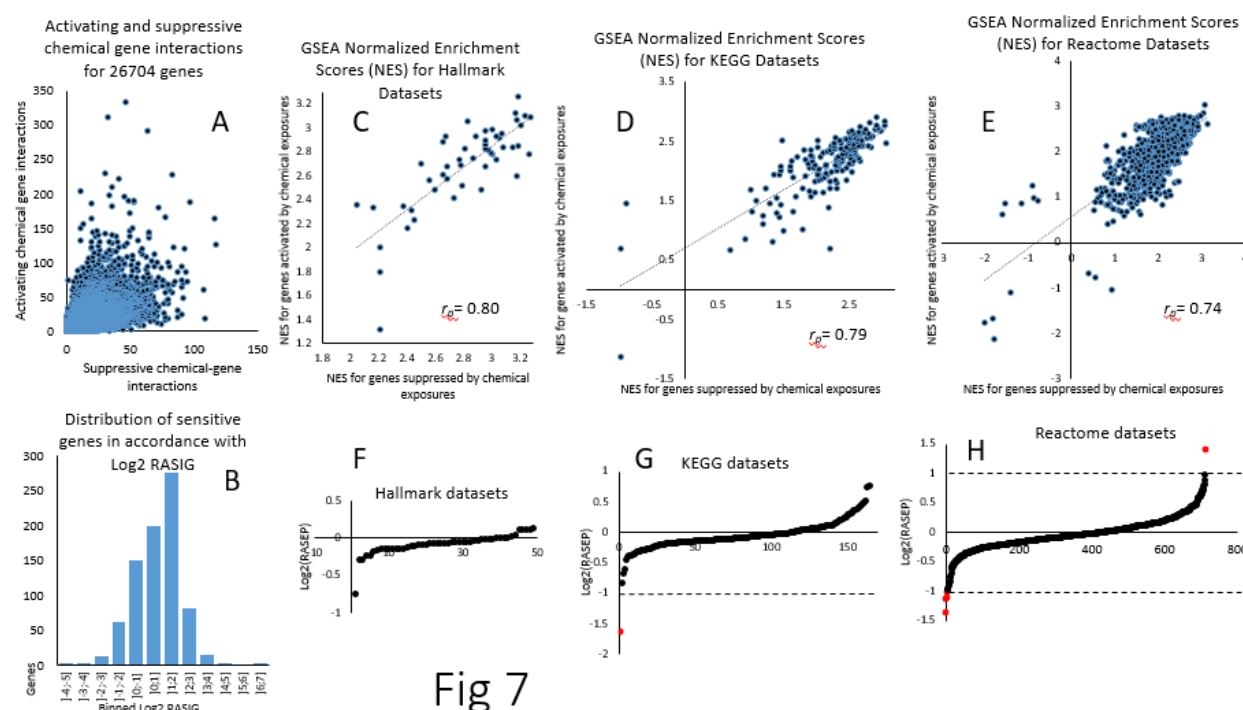

Fig 7

Fig. 7. Gene and pathway distribution in accordance with numbers of activating and suppressive chemical-gene interactions. A – Distribution of 26,704 individual genes. B – Binned distribution of 799 genes with > 50 suppressive or activating interactions in accordance with Log2 ratio of activating to suppressive interactions (RASIG). C-E - Sensitivity of molecular pathways to suppressive and activating effects of exposures. GSEA was conducted for Hallmark (A), KEGG (B) and Reactome (C) datasets against gene lists with only suppressive and only activating interactions. F-H - Distribution of molecular pathways from Hallmark (F), KEGG (G) and Reactome (H) datasets in accordance with their Log2 values of the ratio of activating to suppressive enrichment of a pathway (RASEP). Dashed lines illustrate our reporting cutoffs of Log2(RASEP) < 1 or > -1, corresponding to a 2 fold difference between pathway enrichments by activating and suppressive chemical-gene interactions. Each dot on the plot represents one pathway, and pathways shown in red passed the cutoff threshold. r_p – Pearson correlation.

To identify pathways predominantly activated or predominantly suppressed by chemical exposures we ranked pathways in accordance with the ratio of NESs identified using lists of genes with suppressive or activating interactions respectively. Only pathways for which at least one NES was ≥ 1.9 or ≤ -1.9 and was identified with FDR q ≤ 0.05 were used for this analysis. We call the ratio of NESs used here RASEP – ratio of activating to suppressive enrichment of a pathway. Pathways with RASEP > 2 are enriched 2-fold higher by genes that are activated by chemicals than by genes that are suppressed by chemicals, and vice versa, pathways with RASEP < -2 are enriched 2-fold higher by genes that are suppressed by chemicals than by genes that are activated. No dataset from the Hallmark collection passed this threshold (Fig. 7F). In fact, all pathways, except one (*"Pancreas beta cells"*, RASEP = -1.7), had RASEP values in the interval between -1.2 and 1.2. Among KEGG datasets, one passed our 2-fold criteria, - *"Maturity onset diabetes of the young"* (RASEP = -3.1) (Fig. 7G). Among Reactome datasets, three had a RASEP < -2 (*"Regulation of gene expression in beta cells"*, RASEP = -2.6; "Regulation of beta cell development", RASEP =

-2.2; "*Complex I biogenesis*", RASEP = -2.2 and "*PKMTS methylate histone lysines*", RASEP = -2) (Fig. 7H). One Reactome dataset had a RASEP > 2 ("*Acyl chain remodeling of PE*", RASEP = 2.7).

Discussion

Our study is an attempt to identify molecular mechanisms sensitive to chemical exposures in an unbiased manner. The approach used in this study is based on the assumption that sensitive molecular pathways can be identified as overlapping changes in gene expression from many individual experiments using different biological systems and chemical exposures.

We have enough data to identify molecular pathways sensitive to chemical exposures in an unbiased manner

To identify in an unbiased manner genes and pathways that represent the most generalized response to chemical exposure large numbers of overlaying datasets is needed. We asked, whether the database used in our study is truly big enough (saturated) to identify most sensitive genes and pathways, or if the results of the analysis change (unsaturated database) when more chemical-gene interactions from new studies are added. To answer to this question, we split our database into 2 non-overlapping subsets in terms of lists of chemicals (BA and NBA) and identified the number of chemical-gene interactions per gene and enrichment values of pathways for these 2 subsets independently. The results for the 2 subsets were highly concordant at the gene level (Fig. 2A). Concordance was even higher at the pathway level (Fig. 2B-D), with Pearson correlation of pathway enrichment values in the range of 0.93-0.98. Given that 2 non-overlapping subsets of the database produce highly similar results in the identification of the most sensitive genes and pathways, we concluded that our database is saturated enough to produce unbiased results.

The majority of known molecular pathways are sensitive to chemical exposures

One important finding of this study consists in the observation that majority of known molecular pathways is sensitive to chemical exposures. In fact, out of 1,735 combined Hallmark, KEGG and Reactome datasets 1,173 (67.6%) were positively enriched in GSEA analysis (Supplementary Table S1) and 883 (50.9%) were significantly positively enriched (NES $\geq$ 1.9 or $\leq$ -1.9 and FDR q $\leq$ 0.05). The numbers of negative and significant negative enrichment (corresponding pathways were non-sensitive to chemical exposures) were 10 (0.58%) and 1 (0.06%) respectively. In the previous decade, a significant effort was made to develop *in vitro* toxicity testing approaches, which test perturbation of specific molecular mechanisms known to be connected with adverse phenotypic outcomes (toxicity pathways, adverse outcome pathways) (Ankley et al. 2010; Haynes 2010; NRC 2007; OECD 2020). One important question that remains unanswered is what pathways should be covered by *in vitro* assays to ensure that we do not miss possible toxicities of chemicals using this new paradigm of toxicity testing. Our data suggest that almost every known molecular pathway may be affected by chemical exposures. Should there be evidence connecting these pathways with adverse outcomes, these pathways must be included in the list of targets for *in-vitro* testing.

Unbiased analysis confirms sensitivity of previously recognized mechanisms of toxic response

Modern toxicology recognizes a range of molecular mechanisms responsible for a significant part of toxicity. We asked if these mechanisms are enriched in our analysis. The ability of our approach

to identify these pathways as sensitive to chemical exposures may be considered a quality control for the efficiency of our strategy. Indeed, our analysis demonstrates that mechanisms involved in response to xenobiotics and metabolism of xenobiotics are among the most sensitive to chemical exposures.

Corresponding pathways were enriched in both Metascape (Fig. 3B), and GSEA (Fig. 3C-D) analysis. AhR-regulated CYP1A1 and CYP1B1 were in the list of top 20 genes with highest numbers of chemical-gene interactions (Table 2). Many microsomal and cytosolic oxidoreductases were also among the highest scoring genes for activating chemical-gene interactions (Table 3), while the "*xenobiotic metabolism*" pathway was equally sensitive to activating and suppressive action of chemicals (Table 5). Furthermore, our data highlights the sensitivity of the "*glutathione metabolism*" pathway (Fig. 3D, Table 3, Table 5), which is an indicator of ROS production in response to exposures (Mailloux et al. 2013). Some genes of the oxidative stress response pathway are among the highest scoring for the number of activating chemical-gene interactions (Table 3), although this pathway was highly enriched by both activating and suppressive chemical-gene interactions (Table 5).

Table 5. Top pathways equally sensitive to activating and suppressive effects of chemical exposures.

Enriched category	NES for suppressive chemical-gene interactions	NES for activating chemical-gene interactions
Hallmark		
ESTROGEN RESPONSE LATE	3.06	2.92
INTERFERON GAMMA RESPONSE	3.04	2.93
FATTY ACID METABOLISM	3.07	2.96
UV RESPONSE UP	2.96	2.97
ESTROGEN RESPONSE EARLY	3.01	2.99
XENOBIOTIC METABOLISM	3.21	3.02
EPITHELIAL-MESENCHYMAL TRANSITION	3.17	3.07
HYPOXIA	3.28	3.09
MTORC1 SIGNALING	3.24	3.10
APOPTOSIS	3.17	3.13
TNFA SIGNALING VIA NFKB	3.19	3.26
KEGG		
PANCREATIC CANCER	2.90	2.71
PPAR SIGNALING PATHWAY	2.95	2.73
CELL CYCLE	3.05	2.77
FOCAL ADHESION	2.93	2.78
GLUTATHIONE METABOLISM	2.82	2.81
PATHWAYS IN CANCER	3.06	2.85
P53 SIGNALING PATHWAY	2.93	2.91
Reactome		
FATTY ACID METABOLISM	2.96	2.78
IL4 AND IL13 SIGNALING	3.08	3.03
MITOTIC G1 G1 S PHASES	3.04	2.86
REGULATION OF LIPID METABOLISM BY PPAR ALPHA	2.92	2.75
RESPONSE TO ELEVATED PLATELET CYTOSOLIC Ca2+	2.97	2.79

Additionally, our analysis identified many molecular pathways associated with "*cellular response to stress*" (Fig. 3B-E), including "*response to unfolded protein*" (Fig. 3B) and "*UV response*" (Fig.

3C, Table 5) as sensitive to chemical exposures. Specifically, "*p53 signaling*" was among the most highly enriched pathways in both Metascape and GSEA (Fig. 3B-D, 4E, and 5), and was equally enriched by both activating and suppressive chemical-gene interactions (Table 5). In response to a stress signal, the p53 protein is activated by post-translational modifications, and this leads to either cell cycle arrest and senescence, or cellular apoptosis (Harris and Levine 2005; Jin and Levine 2001). Therefore, it is not surprising that "*apoptosis*", "*cell cycle*" and other related categories were also among the mechanisms most sensitive to chemical exposures (Fig, 3B-E, 4B, F, L). Both *"apoptosois"* and "*cell cycle*" were equally sensitive to activating and suppressive chemical-gene interactions (Table 5). The p53 pathway is reciprocally connected with the hypoxic pathway (Sermeus and Michiels 2011), which is a long recognized response to a broad range of chemical compounds with different toxicities (Lee et al. 2007). "*Hypoxia*" or "*HIF1 TF pathway*" were among the top sensitive molecular mechanisms identified in our analysis (Fig. 3B, C, 4D) and this pathway was equally sensitive to activating and suppressive chemical-gene interactions (Table 5). Given high enrichments of pathways related to cellular stress and the cell cycle, it is not surprising that many cancer-related pathways were also highly enriched in our analysis (see Fig. 4G for example).

Finally, estrogenic signaling, recognized as a pathway affected by a broad range of chemicals (Singleton, Khan, 2003) was among the Hallmark datasets most sensitive to activating and suppressive interaction with chemical compounds (Table 5).

All these findings, taken together, suggest that our approach for the identification of molecular mechanisms sensitive to chemical exposures produced results concordant with the current state of toxicological knowledge.

Lipid metabolism and PPAR signaling

Regulation of lipid metabolism was identified by this study among the molecular mechanisms most sensitive to chemical exposures. This finding was supported by Metascape analysis of the most sensitive genes (Fig. 3B) and by GSEA analysis, where "fatty acid metabolism" was enriched in all 3 analyzed collections of datasets (Fig. 3 C-E, 4K), along with many other relevant categories: "cholesterol homeostasis" (Fig. 3C), "adipogenesis" (Fig. 3C), "PPAR signaling pathway" (Fig. 3D, 4H, 6), "regulation of lipid metabolism by PPAR alpha" (Fig. 3E). "Fatty acid metabolism" and "PPAR signaling pathway" were among the top categories equally sensitive to activating and suppressive chemical-gene interactions (Table 5).

The peroxisome proliferator-activated receptors (PPARs) are a subfamily of nuclear receptors/transcription factors activated by a range of ligands, including free fatty acids, eicosanoids and Vitamin B3, which consists of three isotypes: PPARα, PPARβ/δ and PPARγ encoded by separate genes (Michalik et al. 2006). All three isotypes as well as downstream cascades were among the top scoring genes for chemical-gene interactions (Fig. 5). PPARs modulate genes that regulate energy balance, glucose homeostasis, triglyceride and lipoprotein metabolism, fatty acid synthesis, oxidation, storage, and export, as well as cell proliferation (Berger and Moller 2002; Gross et al. 2017; Liss and Finck 2017; Pawlak et al. 2015; Wang, Y. X. 2010). PPAR mediated dysregulation of these metabolic processes contributes to the pathogenesis of metabolic diseases such as obesity, metabolic syndrome, diabetes, and non-alcoholic fatty liver disease (NAFLD) (Gross et al. 2017; Pawlak et al. 2015; Semple et al. 2006). The ability of some xenobiotics, such as phthalates (Casals-Casas et al. 2008; Ito and Nakajima 2008; Li, Y. et al. 2017), per- and polyfluoroalkyl substances (PFAS) (Behr et al. 2019; Jacobsen et al. 2018; Zhao et al. 2017) and some others (Casals-Casas et al. 2008; Peraza et al. 2006) to interfere with PPAR

signaling is long-recognized (Xi et al. 2020). However, to our knowledge, promiscuous sensitivity of the PPAR regulatory cascade to a broad range of chemical agents has not shown before.

The sensitivity of lipid metabolism pathways to chemical exposures may be relevant to the current epidemic of metabolic disease, one of the biggest public health issues in modern day. According to World Health Organization (WHO), worldwide the prevalence of obesity nearly tripled between 1975 and 2016 (WHO 2016). Other rapidly growing metabolic conditions include Type 2 diabetes, metabolic syndrome and non-alcoholic fatty liver disease (NAFLD). Approximately two out of every five Americans will develop Type 2 diabetes at some point during their adult lives (Gregg et al. 2014). About one third of US adults have metabolic syndrome (National Center for Health Statistics, Division of Health Interview Statistics 2012). Non-alcoholic fatty liver disease (NAFLD) is the most common form of chronic liver disease among adults and children (Kleiner et al. 2005; Patton et al. 2006) with 33% to 88% prevalence (Browning et al. 2004; Ryan et al. 2002; Soejima et al. 2003; Szczepaniak et al. 2005). Increased incidence of obesity and other metabolic conditions cannot be completely explained by poor diet and lack of exercise (Kelly et al. 2008; Wang, Y. and Lobstein 2006; Wang, Y. et al. 2008). Emerging evidence links epidemic of metabolic disease with exposures to environmental xenobiotics (Heindel et al. 2015; Mendrick et al. 2018), but the role of chemical exposure in metabolic health is not well understood. Our data provide an unbiased evidence, supporting sensitivity of lipid metabolism to a broad range of chemical agents. This finding may have significant public health implications and requires additional research.

GH/IGF/mTOR and GH/IGF/FOXO signaling cascades

Another molecular cascade identified in our study as highly sensitive to chemical exposures starts from growth hormone (GH) – insulin-like growth factor (IGF) signaling and includes downstream pathways controlled by the GH-IGF axis. These pathways include mechanistic target of rapamycin (mTOR) and forkhead box O (FOXO) signaling cascades (Fig. 8). Indeed, "*mTOR complex 1 signaling*" was among molecular pathways with the highest NES (Fig 3C, 4C) and it was equally sensitive to both activating and suppressive chemical-gene interactions. "*FOXO mediated transcription*" (Fig 3E) and *"glycolysis"* (Fig. 3C), process activated by FOXO, were also identified as metabolic pathways sensitive to chemical exposures. "*Regulation of IGF transport and uptake by IGFBPs*" was also highly enriched (Fig. 3E). IGF1 was among the top genes with > 200 chemical gene interactions (Table 2). The list of the top 20 genes scoring highest for suppressive interactions included three critical members of the GH-IGF signaling cascade (GHR, IGFBP3, and IGF1) (Table 3). GH regulated insulin-like growth factor binding protein, acid labile subunit (IGFALS) had the second lowest RASIG of all 26,704 genes. Thus, it appears that the whole cascade is highly sensitive to chemical exposures and at the level of GH-IGF it is predominantly suppressed by chemicals.

IGFs are growth factors produced by tissues in response to GH secreted by pituitary. The majority of IGF1 in humans is produced by the liver (LeRoith and Yakar 2007). About 75–80% of serum IGFs are found in a ternary complex consisting of an IGF, IGFBP3 and IGFALS (Allard and Duan 2018). IGF-1 initiates intracellular signaling by binding to receptor tyrosine kinases (RTKs): the IGF-1 receptor (IGF1R), and the insulin receptor (IR). IGF-1 binds with 100-500 times higher affinity to IGF1R than to insulin receptor (IR), while insulin binds with 100 times higher affinity to IR than to IGF1R (Werner et al. 2008). Binding to RTKs activates downstream PI3K/Akt signaling, and IGF-1 is one of the most potent natural activators of this cascade (Peruzzi et al. 1999).
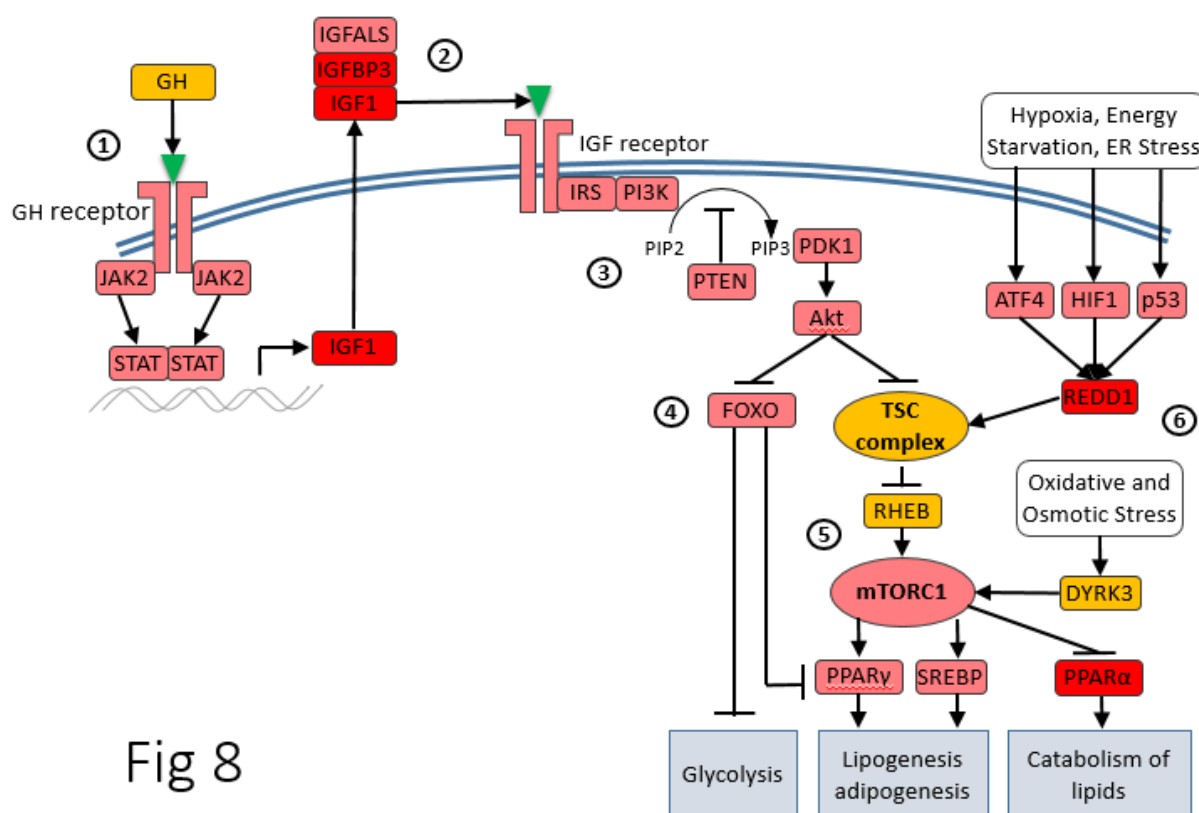
Fig. 8. The growth, longevity, and metabolism cascade identified as highly sensitive to chemical exposures: 1 – GH initiates intracellular signaling via GH receptor/JAK2/STAT pathway and induces production of IGF1; 2 - about 75–80% of serum IGFs is bound in a complex with IGFBP3 and IGFALS; 3 – binding of IGF1 to IGF or insulin receptors activates PI3K/AKT signaling; 4 – activated Akt suppresses FOXO transcription factors and activates downstream lipogenesis, adipogenesis and glycolysis; 5 – activated Akt suppresses TSC complex, which is the major mTORC1 inhibitor; activated mTORC1 suppresses catabolism of lipids via PPARα dependent mechanism and activates lipogenesis and adipogenesis via PPARγ and SREBP transcription factors; 6 – a variety of stress-signaling pathways merge on mTORC1. All nodes shown in red represent the top 1% of genes most sensitive to chemical exposures (≥123 chemical-gene interactions), nodes shown in pink represent the top 10% of genes most sensitive to chemical exposures (≥ 56 chemical-gene interactions). Akt – protein kinase B; ATF4 - activating transcription factor 4; DYRK3 – dual specificity tyrosine phosphorylation regulated kinase 3; FOXO – forkhead box O; GH – growth hormone; HIF1 – hypoxia inducible factor 1; IGF1 – insulin-like growth factor 1; IGFALS – insulin-like growth factor binding protein, acid labile subunit; IGFBP3 – insulin-like growth factor binding protein 3; IRS – insulin receptor substrate; JAK2 – Janus kinase 2; mTORC1 – mechanistic target of rapamycin complex 1; p53 – tumor protein p53; PDK1 – pyruvate dehydrogenase kinase 1; PI3K – phosphatidylinositol-4,5-bisphosphate 3-kinase; PPARγ – peroxisome proliferator activated receptor alpha; PPARγ – peroxisome proliferator activated receptor gamma; PTEN – phosphatase and tensin homolog; REDD1 – regulated in development and DNA damage responses 1; RHEB – RAS homolog enriched in brain; SREBP – sterol regulatory element-binding proteins; STAT – signal transducer and activator of transcription; TSC – tuberous sclerosis complex.

PI3K/Akt activates mTOR, a serine/threonine kinase, which emerged recently as a central master-switch between anabolism and catabolism (Dibble and Manning 2013; Laplante and Sabatini 2012; Laplante and Sabatini 2013). Chronic activation and suppression of the mTOR signaling pathway are associated with obesity and Type 2 diabetes as well as cancer, neurodegeneration, aging and other aging-associated conditions (Khamzina et al. 2005; Laplante and Sabatini 2009; Laplante

and Sabatini 2012; Tremblay et al. 2007; Zoncu et al. 2011). Both fatty acid metabolism (Antikainen et al. 2017; Ben-Sahra and Manning 2017; Lamming and Sabatini 2013; Mao and Zhang 2018), and PPAR signaling (Kim, J. E. and Chen 2004; Sengupta et al. 2010), discussed in the previous section, are regulated downstream of the mTOR pathway. Activation of PI3K/Akt also results in suppression of FOXO signaling via nuclear exclusion and degradation of this transcription factor (Matsuzaki et al. 2003). FOXO positively regulates glucose production via transcriptional activation of gluconeogenesis and glycogenolysis pathways (Daitoku and Fukamizu 2007) and negatively regulates adipogenesis (Farmer 2003) by suppressing transcription of PPARγ (Armoni et al. 2006).

IGF1 is a major regulator of growth, and its secretion patterns in different organisms are linked with different life-history strategies. Species with high IGF concentrations grow, reach reproductive age, reproduce, and age rapidly, while those with low basal levels of IGF have much longer periods of development, reach sexual maturity later, and have a longer lifespan (Swanson and Dantzer 2014). Multiple mouse and other models demonstrate that suppression of different nodes in the GH-IGF1 signaling cascade increase longevity (Bartke 2019; Junnila et al. 2013; Kim, S. S. and Lee 2019; Rincon et al. 2005). Human studies support these findings by revealing greater longevity in people with smaller body size (Chmielewski 2016; Samaras 2014; Samaras 2013). Downstream of GH-IGF1, both mTOR (Blagosklonny 2013a; Blagosklonny 2013b; Papadopoli et al. 2019; Wilkinson et al. 2012) and FOXO (Daitoku and Fukamizu 2007; Jiang et al. 2019) cascades are considered as central mechanisms of longevity regulation.

Ability of diverse chemicals to modulate the GH-IGF axis was noticed previously (Scarth 2006). However, this axis does not attract a lot of attention from researchers despite its relevance for the current epidemic of metabolic disease (see section 4.4) and its importance as a major regulator of growth, metabolism and aging. For example, a PubMed search done on February 14, 2020 for "endocrine disruption" in combination with different hormone names retrieved the following numbers of papers: 1100 for estrogen 502 for "androgen, 365 for testosterone, 353 for thyroid, 134 for progesterone, 71 for cortisol and only 30 for insulin-like growth factor or 24 for IGF. Lack of attention to endocrine disruption of the GH-IGF axis was also reported in our previous study (Suvorov and Takser 2008).

Cytokine mediated signaling

Another group of molecular pathways identified by our study as highly sensitive to chemical exposures consists of immune response pathways. *"Cytokine mediated signaling pathway"* and *"TNF signaling pathway"* were highly enriched in our Metascape analysis (Fig. 3B). Additionally, *"TNF alpha signaling via NF kappa B"* was the top enriched pathway among the Hallmark collection of datasets (Fig. 3C). Other immune-related pathways identified by GSEA were *"Interferon gamma response", "inflammatory response", "IL2 STAT5 signaling"* (Fig. 3C), *"complement and coagulation cascades", "adipocytokine signaling pathway"* (Fig. 3D), and *"IL4 and IL13 signaling"* (Fig. 3E). Among these listed pathways *"Interferon gamma response", "TNF alpha signaling via NF kappa B"* and *"IL4 and IL13 signaling"* were equally sensitive to both activating and suppressive chemical-gene interactions (Table 5).

These findings demonstrate a high sensitivity of cytokine-mediated pathways to chemical exposures. Although having a broad range of functions, TNF alpha and interferon gamma (IFN-γ) are both cytokines involved in systemic inflammation and activation of the acute phase reaction, including the induction of complement and coagulation factors (Gruys et al. 2005). IFN-γ plays an essential role in the development and severity of systemic autoimmunity (Pollard et al. 2013). IL-2 also plays an important role in the maintenance of peripheral self-tolerance. Mice lacking IL-2

or IL-2 receptor subunits develop lymphoproliferative autoimmune syndrome (Malek and Bayer 2004). Similarly, autoimmune disease development is linked to IL-2 deficiencies in humans (Caudy et al. 2007; Roifman 2000; Sharfe et al. 1997). Interleukin-4 and IL-13 are the signature cytokines of the type II inflammatory response. When dysregulated, their activity may trigger asthma, atopic dermatitis, allergic rhinitis, COPD, cancer, inflammatory bowel disease, autoimmune disease and fibrotic disease (May and Fung 2015).

Two major outcomes of a dysfunctional immune system are allergy and autoimmunity. Prevalence for both conditions follows upward trends. For example, among children aged 0–17 years, the prevalence of food allergies in the USA increased from 3.4% in 1997–1999 to 5.1% in 2009–2011. The prevalence of skin allergies increased from 7.4% to 12.5% during the same time-period (Jackson et al. 2013). Resent study of 14,211 participants in the U.S. National Health and Nutrition Examination Survey (NHANES) demonstrated that the prevalence of antinuclear antibodies (ANA), the most common biomarker of autoimmunity, increased from 11% in 1988-1991 to 15.9% in 2011-2012, corresponding to 22 and 41 million affected individuals, respectively (Dinse et al. 2020). Many autoimmune conditions have established links with environmental exposures (Schmidt 2011). Our data demonstrate that immune mechanisms dysfunctional in allergy and autoimmunity are among the molecular mechanisms that are highly sensitive to chemical exposures. This finding indicates the necessity of additional efforts to dissect the connection between chemical exposures and diseases of the immune system.

Many pathways sensitive to chemical exposures are central in cancer biology

A significant portion of the pathways discussed in previous paragraphs 4.3-4.6 and shown in figure 3B-E play important roles in different kinds of cancers. For example, mTOR is the major growth regulator, while p53 is the major growth suppressor and the two have an intricate pattern of cross-talk to determine whether the cell should grow or die in response to damage (Hasty et al. 2013). mTOR inhibitors are a prevalent cancer treatment as mTOR is hyperactivated and/or mutated in many cancers (Hua et al. 2019). Nearly half of human cancers harbor p53 mutations and mutant p53 promotes carcinogenesis, metastasis, tumor recurrence and chemoresistance (Di Agostino et al. 2019). Traditionally, cancer biology is focusing on mutations in proto-oncogenes, which are considered causative to cancer. Our data demonstrate that the same mechanisms that are known to be damaged by mutations in cancers can also undergo suppression or activation as a result of interaction with a broad range of chemical compounds. Given that genetically heterogeneous human populations are ubiquitously exposed to a mix of man-made chemical agents the risks of cancer may be elevated in some susceptible genotypes due to exposures. Identification of these susceptible genotypes results usually in a conclusion, that genetics is the causal factor rather than the environment (Takser and Hunting 2020), while the right interpretation here would be, that specific gene-environment interaction is responsible for increased cancer risks (Takser and Hunting 2020). Identification of molecular mechanisms involved in cancer pathogenesis sensitive to chemical exposures using our or similar approaches could prove useful in advancing our understanding of the biology of gene-environment interactions in heath and disease.

Suppression of pancreatic beta cell genes

Our analysis of molecular pathways that are predominantly enriched by suppressive chemical-gene interactions have identified pancreas as a potentially sensitive target of a broad range of chemical exposures. In fact, *"Pancreas beta cells"* (Hallmark), *"Maturity onset diabetes of the young"* (KEGG), *"Regulation of gene expression in beta cells"* and *"Regulation of beta cell development"*

(Reactome) were datasets with the lowest RASEP values, indicating that genes involved in pancreas physiology and development are highly sensitive to suppressive but not to activating effect of chemical exposures. This phenomenon may be linked to very high susceptibility of pancreatic beta cells to oxidative stress due to a high endogenous production of reactive oxygen species and a low expression of antioxidative enzymes (Lenzen et al. 1996; Li, N. et al. 2008; Wang, D. et al. 2018). Additionally, beta cells development is positively controlled by oxygen in a dose-dependent manner through hypoxia-inducible factor I alpha (HIF-1α) (Heinis et al. 2010; Heinis et al. 2012). Both response to oxidative stress and hypoxia (HIF1a), were among the most sensitive mechanisms affected by chemical exposures in our study. Involvement of these mechanisms cannot completely explain the primarily suppressive effect of chemical compounds on pancreatic genes, however, given that both hypoxia and oxidative stress response ("*glutathione metabolism*") were equally sensitive to suppressive and activating effects of exposures (Table 5).

Insufficient function of beta cells is a critical characteristic of pathogenesis of all three major types of diabetes – type 1 diabetes, type 2 diabetes, and gestational diabetes. According to the WHO, about 422 million people worldwide have diabetes (WHO 2018). Its prevalence has been steadily increasing for the past 3 decades worldwide, and the number of people with diabetes has nearly quadrupled since 1980. In 2012, diabetes was the direct cause of 1.5 million deaths globally. The ability of a broad range of chemicals to suppress expression of genes essential for beta cell development and function may be a significant factor that predisposes the modern population to diabetes development.

Identification of the most sensitive molecular pathways may provide new approaches for the study of mixed exposures

Today, the major approach for studying mixed exposures in toxicology is based on a similarity of mode of action (MOA) within a group of compounds. For example, toxicity of chemical compounds that act as ligands for the aryl hydrocarbon receptor (AhR), is expressed through toxic equivalency (TEQ) to the most potent AhR ligand, 2,3,7,8-tetrachlorodibenzo-p-dioxin. TEQ is established through toxic equivalent factor (TEF) - the ratio of the toxicity of a compound in this category to the toxicity of the most toxic compound (Ahlborg and Hanberg 1994; Van den Berg et al. 2006). TEQ allows for the expression of cumulative toxicity of all AhR ligands in a mixture as a single number. Similar approaches are being developed for other MOA. For example, cumulative effects of compounds with estrogenic activity were successfully predicted using concentration addition (CA) model, which assumes that one agent can be substituted for another in proportion to their relative potencies (Silva et al. 2002; Silva et al. 2011). This approach was further generalized to describe the combination effects of full agonists, partial agonists, and competitive antagonists (Howard and Webster 2009; Howard et al. 2010). In pharmacodynamic modeling, CA provides close approximation even for a mix of agents that affect similar pathways or produce common outcomes via different MOA (Webster 2013). The approach used in this study, identified molecular mechanisms most sensitive to chemical exposures regardless the nature of a chemical compound or its MOA. Use of these pathways and corresponding outcomes to analyze cumulative effects of heterogeneous mixtures may be useful in the future.

Limitations and future directions

The approach used in our study is based on a high level of generalization, which may introduce some bias. For example, it is not clear, how our overall results depend on the composition of target organ/tissues/cells used in toxicological experiments in our database. Also, we do not consider

doses of chemicals used in different studies. It is likely that higher doses of different compounds may trigger different molecular changes (response to damage) than the same compounds at lower doses. In future research we plan to develop annotations of experiments, the sources of data for the database, to elucidate effects of dose, biological model, organ and other factors.

Supplementary Materials

Supplementary Materials 1. Explanation of data preparation for GSEA and results interpretation.

Supplementary Table S1. GSEA enrichment of Hallmark, KEGG and Reactome datasets with genes sensitive to chemical exposures.

Supplementary Data File S1. Chemical-gene interaction database extracted from CTD and manually annotated for chemical use.

References

Ahlborg UG, Hanberg A. 1994. Toxic equivalency factors for dioxin-like PCBs. Environ Sci Pollut Res Int 1(2):67-68; doi: 10.1007/BF02986503 [doi].

Allard JB, Duan C. 2018. IGF-binding proteins: Why do they exist and why are there so many? Front Endocrinol (Lausanne) 9:117; doi: 10.3389/fendo.2018.00117 [doi].

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD et al. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730-741; doi: 10.1002/etc.34 [doi].

Antikainen H, Driscoll M, Haspel G, Dobrowolski R. 2017. TOR-mediated regulation of metabolism in aging. Aging Cell 16(6):1219-1233; doi: 10.1111/acel.12689 [doi].

Armoni M, Harel C, Karni S, Chen H, Bar-Yoseph F, Ver MR et al. 2006. FOXO1 represses peroxisome proliferator-activated receptor-gamma1 and -gamma2 gene promoters in primary adipocytes. A novel paradigm to increase insulin sensitivity. J Biol Chem 281(29):19881-19891; doi: M600320200 [pii].

Bartke A. 2019. Growth hormone and aging: Updated review. World J Mens Health 37(1):19-30; doi: 10.5534/wjmh.180018 [doi].

Behr AC, Plinsch C, Braeuning A, Buhrke T. 2019. Activation of human nuclear receptors by perfluoroalkylated substances (PFAS). Toxicol In Vitro:104700; doi: S0887-2333(19)30628-9 [pii].

Ben-Sahra I, Manning BD. 2017. mTORC1 signaling and the metabolic control of cell growth. Curr Opin Cell Biol 45:72-82; doi: S0955-0674(17)30030-3 [pii].

Berger J, Moller DE. 2002. The mechanisms of action of PPARs. Annu Rev Med 53:409-435; doi: 10.1146/annurev.med.53.082901.104018 [doi].

Blagosklonny MV. 2013a. Aging is not programmed: Genetic pseudo-program is a shadow of developmental growth. Cell Cycle 12(24):3736-3742; doi: 10.4161/cc.27188 [doi].

Blagosklonny MV. 2013b. Rapamycin extends life- and health span because it slows aging. Aging (Albany NY) 5(8):592-598; doi: 100591 [pii].

Boylan JM, Sanders JA, Neretti N, Gruppuso PA. 2015. Profiling of the fetal and adult rat liver transcriptome and translatome reveals discordant regulation by the mechanistic target of rapamycin (mTOR). Am J Physiol Regul Integr Comp Physiol 309(1):R22-35; doi: 10.1152/ajpregu.00114.2015 [doi].

Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC et al. 2004. Prevalence of hepatic steatosis in an urban population in the united states: Impact of ethnicity. Hepatology 40(6):1387-1395; doi: 10.1002/hep.20466 [doi].

Casals-Casas C, Feige JN, Desvergne B. 2008. Interference of pollutants with PPARs: Endocrine disruption meets metabolism. Int J Obes (Lond) 32 Suppl 6:S53-61; doi: 10.1038/ijo.2008.207 [doi].

Caudy AA, Reddy ST, Chatila T, Atkinson JP, Verbsky JW. 2007. CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked-like syndrome, and defective IL-10 expression from CD4 lymphocytes. J Allergy Clin Immunol 119(2):482-487; doi: S0091-6749(06)02119-1 [pii].

Chmielewski P. 2016. The relationship between adult stature and longevity: Tall men are unlikely to outlive their short peers - evidence from a study of all adult deaths in poland in the years 2004-2008. Anthropological Review 79(4):439-460.

Daitoku H, Fukamizu A. 2007. FOXO transcription factors in the regulatory networks of longevity. J Biochem 141(6):769-774; doi: 141/6/769 [pii].

Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J et al. 2019. The comparative toxicogenomics database: Update 2019. Nucleic Acids Res 47(D1):D948-D954; doi: 10.1093/nar/gky868 [doi].

Di Agostino S, Fontemaggi G, Strano S, Blandino G, D'Orazi G. 2019. Targeting mutant p53 in cancer: The latest insights. J Exp Clin Cancer Res 38(1):290-019-1302-0; doi: 10.1186/s13046-019-1302-0 [doi].

Dibble CC, Manning BD. 2013. Signal integration by mTORC1 coordinates nutrient input with biosynthetic output. Nat Cell Biol 15(6):555-564; doi: 10.1038/ncb2763 [doi].

Dinse GE, Parks CG, Weinberg CR, Co CA, Wilkerson J, Zeldin DC et al. 2020. Increasing prevalence of antinuclear antibodies in the united states. Arthritis Rheumatol; doi: 10.1002/art.41214 [doi].

EPA. 2013. Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology. Washington, D.C.:Environmental Protection Agency.

Escher BI, Stapleton HM, Schymanski EL. 2020. Tracking complex mixtures of chemicals in our changing environment. Science 367(6476):388-392; doi: 10.1126/science.aay6636 [doi].

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P et al. 2018. The reactome pathway knowledgebase. Nucleic Acids Res 46(D1):D649-D655; doi: 10.1093/nar/gkx1132 [doi].

Farmer SR. 2003. The forkhead transcription factor Foxo1: A possible link between obesity and insulin resistance. Mol Cell 11(1):6-8; doi: S1097-2765(03)00003-0 [pii].

Grandjean P, Bellanger M. 2017. Calculation of the disease burden associated with environmental chemical exposures: Application of toxicological information in health economic estimation. Environ Health 16(1):123-017-0340-3; doi: 10.1186/s12940-017-0340-3 [doi].

Gregg EW, Zhuo X, Cheng YJ, Albright AL, Narayan KM, Thompson TJ. 2014. Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985-2011: A modelling study. Lancet Diabetes Endocrinol 2(11):867-74; doi: S2213-8587(14)70161-5 [pii].

Gross B, Pawlak M, Lefebvre P, Staels B. 2017. PPARs in obesity-induced T2DM, dyslipidaemia and NAFLD. Nat Rev Endocrinol 13(1):36-49; doi: 10.1038/nrendo.2016.135 [doi].

Gruys E, Toussaint MJ, Niewold TA, Koopmans SJ. 2005. Acute phase reaction and acute phase proteins. J Zhejiang Univ Sci B 6(11):1045-1056; doi: 10.1631/jzus.2005.B1045 [doi].

Harris SL, Levine AJ. 2005. The p53 pathway: Positive and negative feedback loops. Oncogene 24(17):2899-2908; doi: 1208615 [pii].

Hasty P, Sharp ZD, Curiel TJ, Campisi J. 2013. mTORC1 and p53: Clash of the gods? Cell Cycle 12(1):20-25; doi: 22912 [pii].

Haynes RC. 2010. ToxCast on target: In vitro assays and computer modeling show promise for screening chemicals. Environ Health Perspect 118(4):A172-a172a; doi: 10.1289/ehp.118-a172a [doi].

Heindel JJ, vom Saal FS, Blumberg B, Bovolin P, Calamandrei G, Ceresini G et al. 2015. Parma consensus statement on metabolic disruptors. Environ Health 14:54-015-0042-7; doi: 10.1186/s12940-015-0042-7 [doi].

Heinis M, Simon MT, Ilc K, Mazure NM, Pouyssegur J, Scharfmann R et al. 2010. Oxygen tension regulates pancreatic beta-cell differentiation through hypoxia-inducible factor 1alpha. Diabetes 59(3):662-669; doi: 10.2337/db09-0891 [doi].

Heinis M, Soggia A, Bechetoille C, Simon MT, Peyssonnaux C, Rustin P et al. 2012. HIF1alpha and pancreatic beta-cell development. FASEB J 26(7):2734-2742; doi: 10.1096/fj.11-199224 [doi].

Himmelstein MW, Boogaard PJ, Cadet J, Farmer PB, Kim JH, Martin EA et al. 2009. Creating context for the use of DNA adduct data in cancer risk assessment: II. overview of methods of identification and quantitation of DNA damage. Crit Rev Toxicol 39(8):679-694; doi: 10.1080/10408440903164163 [doi].

Howard GJ, Schlezinger JJ, Hahn ME, Webster TF. 2010. Generalized concentration addition predicts joint effects of aryl hydrocarbon receptor agonists with partial agonists and competitive antagonists. Environ Health Perspect 118(5):666-672; doi: 10.1289/ehp.0901312 [doi].

Howard GJ, Webster TF. 2009. Generalized concentration addition: A method for examining mixtures containing partial agonists. J Theor Biol 259(3):469-477; doi: 10.1016/j.jtbi.2009.03.030 [doi].

Hua H, Kong Q, Zhang H, Wang J, Luo T, Jiang Y. 2019. Targeting mTOR for cancer therapy. J Hematol Oncol 12(1):71-019-0754-1; doi: 10.1186/s13045-019-0754-1 [doi].

Ito Y, Nakajima T. 2008. PPARalpha- and DEHP-induced cancers. PPAR Res 2008:759716; doi: 10.1155/2008/759716 [doi].

Jackson KD, Howie LD, Akinbami LJ. 2013. Trends in allergic conditions among children: United states, 1997-2011. NCHS Data Brief (121)(121):1-8.

Jacobsen AV, Norden M, Engwall M, Scherbak N. 2018. Effects of perfluorooctane sulfonate on genes controlling hepatic fatty acid metabolism in livers of chicken embryos. Environ Sci Pollut Res Int 25(23):23074-23081; doi: 10.1007/s11356-018-2358-7 [doi].

Jarabek AM, Pottenger LH, Andrews LS, Casciano D, Embry MR, Kim JH et al. 2009. Creating context for the use of DNA adduct data in cancer risk assessment: I. data organization. Crit Rev Toxicol 39(8):659-678; doi: 10.1080/10408440903164155 [doi].

Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A et al. 2020. The reactome pathway knowledgebase. Nucleic Acids Res 48(D1):D498-D503; doi: 10.1093/nar/gkz1031 [doi].

Jiang Y, Yan F, Feng Z, Lazarovici P, Zheng W. 2019. Signaling network of forkhead family of transcription factors (FOXO) in dietary restriction. Cells 9(1):10.3390/cells9010100; doi: E100 [pii].

Jin S, Levine AJ. 2001. The p53 functional circuit. J Cell Sci 114(Pt 23):4139-4140.

Junnila RK, List EO, Berryman DE, Murrey JW, Kopchick JJ. 2013. The GH/IGF-1 axis in ageing and longevity. Nat Rev Endocrinol 9(6):366-376; doi: 10.1038/nrendo.2013.67 [doi].

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27-30; doi: gkd027 [pii].

Kelly T, Yang W, Chen CS, Reynolds K, He J. 2008. Global burden of obesity in 2005 and projections to 2030. Int J Obes (Lond) 32(9):1431-1437; doi: 10.1038/ijo.2008.102 [doi].

Khamzina L, Veilleux A, Bergeron S, Marette A. 2005. Increased activation of the mammalian target of rapamycin pathway in liver and skeletal muscle of obese rats: Possible involvement in obesity-linked insulin resistance. Endocrinology 146(3):1473-1481; doi: en.2004-0921 [pii].

Kim JE, Chen J. 2004. Regulation of peroxisome proliferator-activated receptor-gamma activity by mammalian target of rapamycin and amino acids in adipogenesis. Diabetes 53(11):2748-2756; doi: 53/11/2748 [pii].

Kim SS, Lee CK. 2019. Growth signaling and longevity in mouse models. BMB Rep 52(1):70-85; doi: 4454 [pii].

Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW et al. 2005. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 41(6):1313-1321; doi: 10.1002/hep.20701 [doi].

Lamming DW, Sabatini DM. 2013. A central role for mTOR in lipid homeostasis. Cell Metab 18(4):465-469; doi: 10.1016/j.cmet.2013.08.002 [doi].

Laplante M, Sabatini DM. 2009. An emerging role of mTOR in lipid biosynthesis. Curr Biol 19(22):R1046-52; doi: 10.1016/j.cub.2009.09.058 [doi].

Laplante M, Sabatini DM. 2012. mTOR signaling in growth control and disease. Cell 149(2):274-293; doi: 10.1016/j.cell.2012.03.017 [doi].

Laplante M, Sabatini DM. 2013. Regulation of mTORC1 and its impact on gene expression at a glance. J Cell Sci 126(Pt 8):1713-1719; doi: 10.1242/jcs.125773 [doi].

Lee KA, Roth RA, LaPres JJ. 2007. Hypoxia, drug therapy and toxicity. Pharmacol Ther 113(2):229-246; doi: S0163-7258(06)00136-7 [pii].

Lenzen S, Drinkgern J, Tiedge M. 1996. Low antioxidant enzyme gene expression in pancreatic islets compared with various other mouse tissues. Free Radic Biol Med 20(3):463-466; doi: 0891-5849(96)02051-5 [pii].

LeRoith D, Yakar S. 2007. Mechanisms of disease: Metabolic effects of growth hormone and insulin-like growth factor 1. Nat Clin Pract Endocrinol Metab 3(3):302-10.

Li N, Frigerio F, Maechler P. 2008. The sensitivity of pancreatic beta-cells to mitochondrial injuries triggered by lipotoxicity and oxidative stress. Biochem Soc Trans 36(Pt 5):930-934; doi: 10.1042/BST0360930 [doi].

Li Y, Liu X, Niu L, Li Q. 2017. Proteomics analysis reveals an important role for the PPAR signaling pathway in DBDCT-induced hepatotoxicity mechanisms. Molecules 22(7):10.3390/molecules22071113; doi: E1113 [pii].

Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 1(6):417-425; doi: 10.1016/j.cels.2015.12.004 [doi].

Liss KH, Finck BN. 2017. PPARs and nonalcoholic fatty liver disease. Biochimie 136:65-74; doi: S0300-9084(16)30292-9 [pii].

Mailloux RJ, McBride SL, Harper ME. 2013. Unearthing the secrets of mitochondrial ROS and glutathione in bioenergetics. Trends Biochem Sci 38(12):592-602; doi: 10.1016/j.tibs.2013.09.001 [doi].

Malek TR, Bayer AL. 2004. Tolerance, not immunity, crucially depends on IL-2. Nat Rev Immunol 4(9):665-674; doi: 10.1038/nri1435 [doi].

Mao Z, Zhang W. 2018. Role of mTOR in glucose and lipid metabolism. Int J Mol Sci 19(7):10.3390/ijms19072043; doi: E2043 [pii].

Matsuzaki H, Daitoku H, Hatta M, Tanaka K, Fukamizu A. 2003. Insulin-induced phosphorylation of FKHR (Foxo1) targets to proteasomal degradation. Proc Natl Acad Sci U S A 100(20):11285-11290; doi: 10.1073/pnas.1934283100 [doi].

May RD, Fung M. 2015. Strategies targeting the IL-4/IL-13 axes in disease. Cytokine 75(1):89-116; doi: S1043-4666(15)00204-5 [pii].

Mendrick DL, Diehl AM, Topor LS, Dietert RR, Will Y, La Merrill MA et al. 2018. Metabolic syndrome and associated diseases: From the bench to the clinic. Toxicol Sci 162(1):36-42; doi: 10.1093/toxsci/kfx233 [doi].

Michalik L, Auwerx J, Berger JP, Chatterjee VK, Glass CK, Gonzalez FJ et al. 2006. International union of pharmacology. LXI. peroxisome proliferator-activated receptors. Pharmacol Rev 58(4):726-741; doi: 58/4/726 [pii].

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J et al. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34(3):267-273; doi: 10.1038/ng1180 [doi].

NASEM. 2017. Using 21st Century Science to Improve Risk-Related Evaluations. National Academies of Sciences, Engineering, and Medicine ed. Washington, DC:The National Academies Press.

National Center for Health Statistics, Division of Health Interview Statistics. 2012. Crude and Age-Adjusted Percentage of Civilian, Noninstitutionalized Adults with Diagnosed Diabetes, United States, 1980–2010. National Center for Chronic Disease Prevention and Health Promotion ed. Atlanta, GA:Centers for Disease Control and Prevention, Division of Diabetes Translation.

NRC. 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, D.C.:National Research Council.

OECD. 2017. Revised Guidance Document on Developing and Assessing Adverse Outcome Pathways. Paris, France:OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 184.

OECD. 2020. Adverse Outcome Pathways, Molecular Screening and Toxicogenomics. http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm ed. .

Papadopoli D, Boulay K, Kazak L, Pollak M, Mallette F, Topisirovic I et al. 2019. mTOR as a central regulator of lifespan and aging. F1000Res 8:10.12688/f1000research.17196.1. eCollection 2019; doi: F1000 Faculty Rev-998 [pii].

Patton HM, Sirlin C, Behling C, Middleton M, Schwimmer JB, Lavine JE. 2006. Pediatric nonalcoholic fatty liver disease: A critical appraisal of current data and implications for future research. J Pediatr Gastroenterol Nutr 43(4):413-427; doi: 10.1097/01.mpg.0000239995.58388.56 [doi].

Pawlak M, Lefebvre P, Staels B. 2015. Molecular mechanism of PPARalpha action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. J Hepatol 62(3):720-733; doi: 10.1016/j.jhep.2014.10.039 [doi].

Peraza MA, Burdick AD, Marin HE, Gonzalez FJ, Peters JM. 2006. The toxicology of ligands for peroxisome proliferator-activated receptors (PPAR). Toxicol Sci 90(2):269-295; doi: kfj062 [pii].

Peruzzi F, Prisco M, Dews M, Salomoni P, Grassilli E, Romano G et al. 1999. Multiple signaling pathways of the insulin-like growth factor 1 receptor in protection from apoptosis. Mol Cell Biol 19(10):7203-7215; doi: 10.1128/mcb.19.10.7203 [doi].

Piller C. 2020. Transparency on trial. Science 367(6475):240-243; doi: 10.1126/science.367.6475.240 [doi].

Pollard KM, Cauvi DM, Toomey CB, Morris KV, Kono DH. 2013. Interferon-gamma and systemic autoimmunity. Discov Med 16(87):123-131.

Rincon M, Rudin E, Barzilai N. 2005. The insulin/IGF-1 signaling in mammals and its relevance to human longevity. Exp Gerontol 40(11):873-877; doi: S0531-5565(05)00160-9 [pii].

Roifman CM. 2000. Human IL-2 receptor alpha chain deficiency. Pediatr Res 48(1):6-11; doi: 10.1203/00006450-200007000-00004 [doi].

Ryan CK, Johnson LA, Germin BI, Marcos A. 2002. One hundred consecutive hepatic biopsies in the workup of living donors for right lobe liver transplantation. Liver Transpl 8(12):1114-1122; doi: 10.1053/jlts.2002.36740 [doi].

Samaras TT. 2014. Evidence from eight different types of studies showing that smaller body size is related to greater longevity. Journal of Scientific Research and & Reports 3(16):2150-2160.

Samaras TT. 2013. Shorter height is related to lower cardiovascular disease risk - a narrative review. Indian Heart J 65(1):66-71; doi: 10.1016/j.ihj.2012.12.016 [doi].

Scarth JP. 2006. Modulation of the growth hormone-insulin-like growth factor (GH-IGF) axis by pharmaceutical, nutraceutical and environmental xenobiotics: An emerging role for xenobiotic-metabolizing enzymes and the transcription factors regulating their expression. A review. Xenobiotica 36(2-3):119-218.

Schmidt CW. 2011. Questions persist: Environmental factors in autoimmune disease. Environ Health Perspect 119(6):A249-53; doi: 10.1289/ehp.119-a248 [doi].

Semple RK, Chatterjee VK, O'Rahilly S. 2006. PPAR gamma and human metabolic disease. J Clin Invest 116(3):581-589; doi: 10.1172/JCI28003 [doi].

Sengupta S, Peterson TR, Laplante M, Oh S, Sabatini DM. 2010. mTORC1 controls fasting-induced ketogenesis and its modulation by ageing. Nature 468(7327):1100-1104; doi: 10.1038/nature09584 [doi].

Sermeus A, Michiels C. 2011. Reciprocal influence of the p53 and the hypoxic pathways. Cell Death Dis 2:e164; doi: 10.1038/cddis.2011.48 [doi].

Sharfe N, Dadi HK, Shahar M, Roifman CM. 1997. Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. Proc Natl Acad Sci U S A 94(7):3168-3171; doi: 10.1073/pnas.94.7.3168 [doi].

Sies H, Berndt C, Jones DP. 2017. Oxidative stress. Annu Rev Biochem 86:715-748; doi: 10.1146/annurev-biochem-061516-045037 [doi].

Silva E, Rajapakse N, Kortenkamp A. 2002. Something from "nothing"--eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. Environ Sci Technol 36(8):1751-1756; doi: 10.1021/es0101227 [doi].

Silva E, Rajapakse N, Scholze M, Backhaus T, Ermler S, Kortenkamp A. 2011. Joint effects of heterogeneous estrogenic chemicals in the E-screen--exploring the applicability of concentration addition. Toxicol Sci 122(2):383-394; doi: 10.1093/toxsci/kfr103 [doi].

Simmons SO. 2009. Hypoxia response: A model toxicity pathway for high-throughput screening. Toxicol Sci 112(1):1-3; doi: 10.1093/toxsci/kfp193 [doi].

Singleton DW, Khan SA. 2003. Xenoestrogen exposure and mechanisms of endocrine disruption. Front Biosci 8:s110-8; doi: 10.2741/1010 [doi].

Soejima Y, Shimada M, Suehiro T, Kishikawa K, Yoshizumi T, Hashimoto K et al. 2003. Use of steatotic graft in living-donor liver transplantation. Transplantation 76(2):344-348; doi: 10.1097/01.TP.0000071205.52835.A4 [doi].

Strapacova S, Brenerova P, Krcmar P, Andersson P, van Ede KI, van Duursen MBM et al. 2018. Relative effective potencies of dioxin-like compounds in rodent and human lung cell models. Toxicology 404-405:33-41; doi: S0300-483X(18)30065-9 [pii].

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102(43):15545-50; doi: 0506580102 [pii] 10.1073/pnas.0506580102.

Suvorov A, Takser L. 2008. Facing the challenge of data transfer from animal models to humans: The case of persistent organohalogens. Environ Health(7):58.

Swanson EM, Dantzer B. 2014. Insulin-like growth factor-1 is associated with life-history variation across mammalia. Proc Biol Sci 281(1782):20132458; doi: 10.1098/rspb.2013.2458 [doi].

Szczepaniak LS, Nurenberg P, Leonard D, Browning JD, Reingold JS, Grundy S et al. 2005. Magnetic resonance spectroscopy to measure hepatic triglyceride content: Prevalence of hepatic steatosis in the general population. Am J Physiol Endocrinol Metab 288(2):E462-8; doi: 10.1152/ajpendo.00064.2004 [doi].

Takser L, Hunting DJ. 2020. Learning from sisyphus: Time to rethink our current, ineffective strategy on neurodevelopmental environmental toxicants. Environ Health 19(1):31-020-00587-w; doi: 10.1186/s12940-020-00587-w [doi].

Tremblay F, Brule S, Hee Um S, Li Y, Masuda K, Roden M et al. 2007. Identification of IRS-1 ser-1101 as a target of S6K1 in nutrient- and obesity-induced insulin resistance. Proc Natl Acad Sci U S A 104(35):14056-14061; doi: 0706517104 [pii].

Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA et al. 2015. Meta- and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. Cell Host Microbe 18(6):723-735; doi: 10.1016/j.chom.2015.11.002 [doi].

Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M et al. 2006. The 2005 world health organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. Toxicol Sci 93(2):223-241; doi: kfl055 [pii].

Vinken M. 2013. The adverse outcome pathway concept: A pragmatic tool in toxicology. Toxicology 312:158-165; doi: 10.1016/j.tox.2013.08.011 [doi].

Wang D, Yan J, Teng M, Yan S, Zhou Z, Zhu W. 2018. In utero and lactational exposure to BDE-47 promotes obesity development in mouse offspring fed a high-fat diet: Impaired lipid metabolism and intestinal dysbiosis. Arch Toxicol; doi: 10.1007/s00204-018-2177-0 [doi].

Wang Y, Beydoun MA, Liang L, Caballero B, Kumanyika SK. 2008. Will all americans become overweight or obese? estimating the progression and cost of the US obesity epidemic. Obesity (Silver Spring) 16(10):2323-30; doi: oby2008351 [pii] 10.1038/oby.2008.351.

Wang Y, Lobstein T. 2006. Worldwide trends in childhood overweight and obesity. Int J Pediatr Obes 1(1):11-25; doi: 10.1080/17477160600586747 [doi].

Wang YX. 2010. PPARs: Diverse regulators in energy metabolism and metabolic diseases. Cell Res 20(2):124-137; doi: 10.1038/cr.2010.13 [doi].

Webster TF. 2013. Mixtures of endocrine disruptors: How similar must mechanisms be for concentration addition to apply? Toxicology 313(2-3):129-133; doi: 10.1016/j.tox.2013.01.009 [doi].

Werner H, Weinstein D, Bentov I. 2008. Similarities and differences between insulin and IGF-I: Structures, receptors, and signalling pathways. Arch Physiol Biochem 114(1):17-22; doi: 10.1080/13813450801900694 [doi].

WHO. 2016. Obesity and Overweight. https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight ed. .

WHO. 2018. Diabetes. https://www.who.int/news-room/facts-in-pictures/detail/diabetes ed. .

Wilkinson JE, Burmeister L, Brooks SV, Chan CC, Friedline S, Harrison DE et al. 2012. Rapamycin slows aging in mice. Aging Cell 11(4):675-682; doi: 10.1111/j.1474-9726.2012.00832.x [doi].

Xi Y, Zhang Y, Zhu S, Luo Y, Xu P, Huang Z. 2020. PPAR-mediated toxicology and applied pharmacology. Cells 9(2):10.3390/cells9020352; doi: E352 [pii].

Zhao M, Jiang Q, Geng M, Zhu L, Xia Y, Khanal A et al. 2017. The role of PPAR alpha in perfluorooctanoic acid induced developmental cardiotoxicity and l-carnitine mediated protection-results of in ovo gene silencing. Environ Toxicol Pharmacol 56:136-144; doi: S1382-6689(17)30264-8 [pii].

Zoncu R, Efeyan A, Sabatini DM. 2011. mTOR: From growth signal integration to cancer, diabetes and ageing. Nat Rev Mol Cell Biol 12(1):21-35; doi: 10.1038/nrm3025 [doi].

Supplemental Material 1. Explanation of data preparation for GSEA and GSEA plots interpretation.

File with pre-ranked gene list was prepared the following way:

1.   All genes were ranked in accordance with numbers of chemical-gene interactions – see figure S1 below.
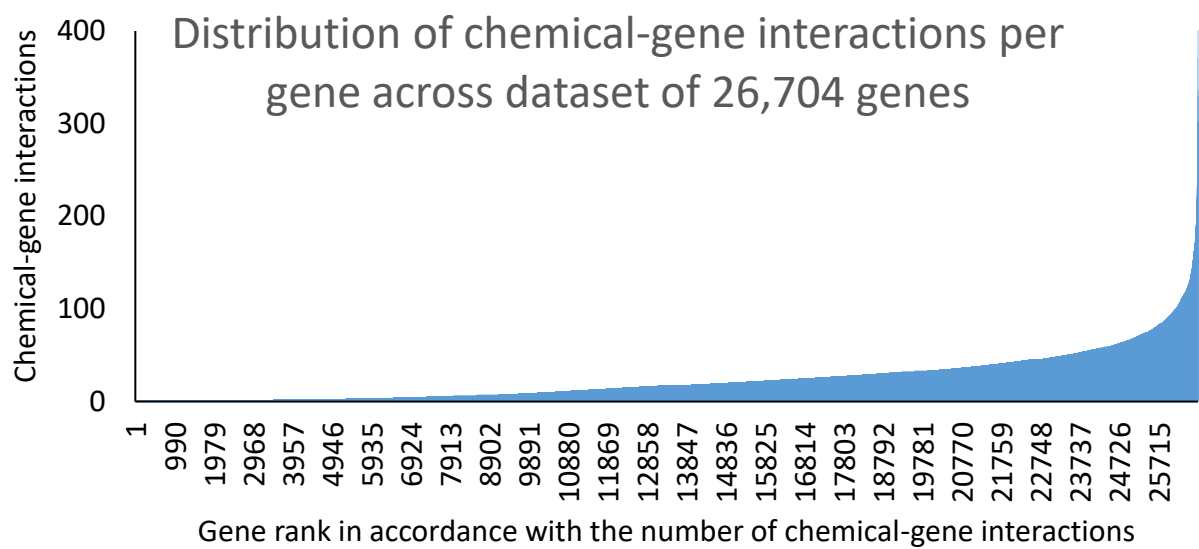

Figure S1. Distribution of chemical-gene interactions values across dataset of 26,704

2.   The same number was subtracted from values of chemical-gene interactions for every gene, to achieve equal negative and positive area under the curve for these values distribution – see figure S2 below.
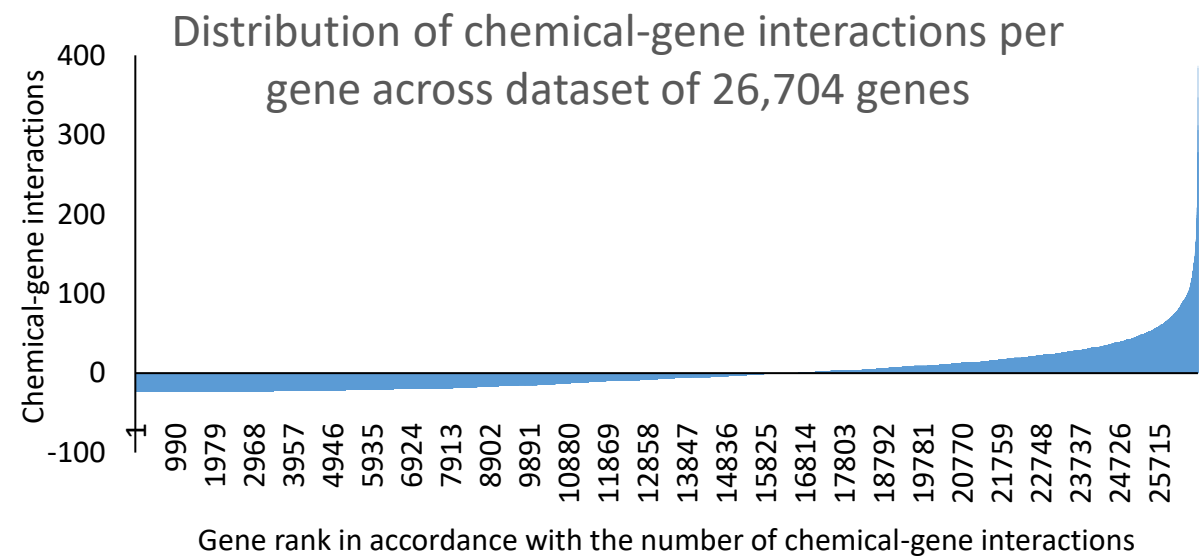

Figure S2. Distribution of chemical-gene interactions values across dataset of 26,704 after subtraction of the same value from all values of chemical-gene interactions to achieve equal negative and positive area under the curve.

Interpretation of GSEA plots based on example of enriched 'PPAR signaling pathway' from KEGG collection of datasets:
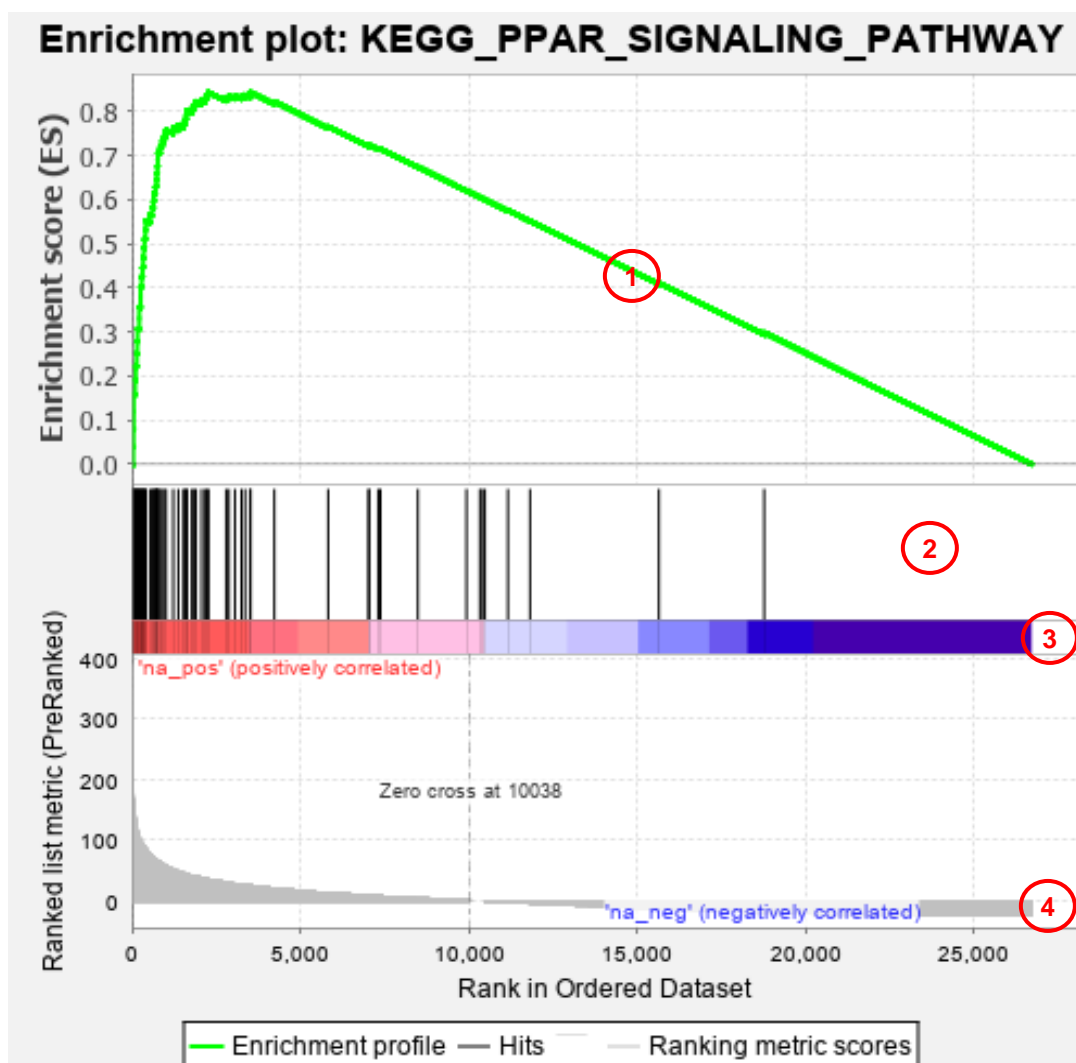
Figure S3:   GSEA plot interpretation: 1 − running enrichment score for the gene set; 2 − vertical lines show where the members of the gene set appear in the ranked list of genes; 3 and 4 – list of genes ranked based on values for chemical-gene interactions from the gene with highest values of interactions (left of the plot) to gene with smallest number of interactions (right of the plot). Number of interactions is shown in the heat-bar (3) and in the bar plot (4). Values of chemical-gene interactions were adjusted to achieve equal negative and positive area under the curve for these values distribution – see figures S1 and S2..