

Supplementary data for “diXa: a Data Infrastructure for Chemical Safety Assessment”

Diana M. Hendrickx^{1,*}, Hugo J.W.L. Aerts², Florian Caiment¹, Dominic Clark³, Timothy M.D. Ebbels⁴, Chris T. Evelo⁵, Hans Gmuender⁶, Dennie G.A.J. Hebels¹, Ralf Herwig⁷, Jürgen Hescheler⁸, Danyel G.J. Jennen¹, Marlon J.A. Jetten¹, Stathis Kanterakis³, Hector C. Keun⁴, Vera Matser³, John P. Overington³, Ekaterina Pilicheva³, Ugis Sarkans³, Marcelo P. Segura-Lepe⁴, Isaia Sotiriadou⁸, Timo Wittenberger⁶, Clemens Wittwehr⁹, Antonella Zanzi⁹, and Jos C.S. Kleinjans¹

¹Department of Toxicogenomics, School of Oncology and Developmental Biology (GROW), Maastricht University, 6200 MD Maastricht, The Netherlands. ² Dana-Farber Cancer Institute, Brigham and Women’s Hospital, Harvard Medical School, Boston, 02215, MA, USA. ³European Molecular Biology Laboratory – European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, CB10 1SD, United Kingdom. ⁴ Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, South Kensington, London SW7 2AZ, UK. ⁵ Department of Bioinformatics – BiGCaT Bioinformatics and Systems Biology Research Group, Maastricht University, 6200 MD Maastricht, The Netherlands. ⁶Genedata AG, CH-4053 Basel, Switzerland. ⁷Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. ⁸Center of Physiology and Pathophysiology, Institute of Neurophysiology, University of Cologne, Cologne 50931, Germany. ⁹ European Commission , Joint Research Centre, Via E. Fermi 2749, 21027 Ispra VA, Italy.

Table S1 Overview of projects participating in diXa

Project	Description	Weblink	Organism(s)	Organ(s) / tissue(s)	Measurement type	Technology type
New Generis (1 study)	Links between exposure to environmental/food carcinogens during pregnancy and childhood cancer	http://www.newgeneris.org/	Homo Sapiens	Human PBMC	Transcription profiling	DNA microarray (Agilent)
Carcinogenomics (3 studies)	High throughput genomics-based tests for assessing genotoxic and carcinogenic properties of chemical compounds <i>in vitro</i>	http://www.carcinogenomics.eu/	Homo Sapiens Rattus norvegicus	Cell lines: liver (human and rat), kidney (human and rat), lung (human)	Metabolite profiling (only liver and kidney) Transcription profiling (liver, kidney, lung)	NMR spectroscopy (Bruker) DNA microarray (Affymetrix)
Envirogenomarkers (2 studies)	Role of environmental agents in human disease	http://www.envirogenomarkers.net/	Homo Sapiens	Blood specimen	Transcription profiling	DNA microarray (Agilent)
InnoMed – PredTox (16 studies)	Pre-clinical safety evaluation of compounds	https://www.genedata.com/lp/innomed-predtox.html	Rattus norvegicus	Urine, liver, kidney, serum, blood, plasma	Transcription profiling Metabolite profiling Protein expression profiling	DNA microarray (Affymetrix GeneChip) Mass spectrometry (LC/MS) NMR spectroscopy (1D NMR) Mass spectrometry (SELDI-TOF-MS, 2D-PAGE/MALDI-TOF/MS, 2D-DIGE/MALDI-TOF-MS)
Esnats (2 studies)	Embryonic stem cell-based alternative novel testing strategies	http://www.esnats.eu/	Homo Sapiens	Embryonic stem cells, skin-derived stem cells	Transcription profiling	DNA microarray (Affymetrix)
Predictomics (1 study)	Short-term <i>in vitro</i> assays for long-term toxicity	http://www.ist-world.org/ProjectDetails.aspx?ProjectId=d62e81aa1bf442559aa3d2eea5b20c4f&SourceDatabaseId=7cff9226e582440894200b751bab883f	Homo Sapiens	Liver cell lines	Transcription profiling	DNA microarray (Affymetrix)
Open TG-GATEs (4 studies)	Toxicity information for 150 compounds on rat <i>in vivo</i> , liver and kidney cells, and human hepatocytes	http://toxico.nibio.go.jp/english/index.html	Homo Sapiens Rattus norvegicus	Liver cell lines (rat and human) Liver (only rat) Kidney (only rat)	Transcription profiling	DNA microarray (Affymetrix)
DrugMatrix (5 studies)	Toxicogenomic profiles for 638 different compounds	https://ntp.niehs.nih.gov/drugmatrix/index.html	Rattus norvegicus	Kidney, liver, heart, liver cell lines, skeletal muscle tissue	Transcription profiling	DNA microarray (Affymetrix)

Table S2 Current external sources linked to the diXa Data Warehouse through the ChemAgora Portal.

Database	Weblink	Reference
ChEMBL	https://www.ebi.ac.uk/chembl/	(Warr, 2009)
ChEBI	http://www.ebi.ac.uk/chebi/	(Degtyarenko, et al., 2008)
ChemSpider	http://www.chemspider.com/	(Pence and Williams, 2010)
PubChem	https://pubchem.ncbi.nlm.nih.gov/	(Wang, et al., 2009)
eChemPortal	http://www.echemportal.org	(Wittwehr, 2011)
ConsensusPathDB	http://cpdb.molgen.mpg.de/	(Kamburov, et al., 2013)
CTD	http://ctdbase.org/	(Davis, et al., 2013)
DrugBank	http://www.drugbank.ca/	(Law, et al., 2014)
ToxNet	http://toxnet.nlm.nih.gov/	(Wexler, 2001)
CommonChemistry	http://www.commonchemistry.org/	-
AOP-Wiki	https://aopkb.org/aopwiki/	-
CheList	http://chelist.jrc.ec.europa.eu/	-

Uploading data

New data are added to the diXa repository in ISA-Tab format (Rocca-Serra, et al., 2010) (isatab.sourceforge.net/tools.html). ISA-Tab is a metadata format that describes investigations, studies and assays, and ties together this information using links to the experimental data. The data catalogues used by the diXa studies are available in ArrayExpress (Rustici, et al., 2013).

The diXa ISA template can be downloaded from bitbucket.org/kanterae/isaconfig-dixa/src. The template has additional fields related to toxicology, such as 'dose' and 'treatment' and uses several ontologies (see Supplementary Table S3).

diXa Data Warehouse
 A collection of European Toxicogenomics experiments with cross-links to chemical and molecular medicine databases.

Studies 34

Samples 3142

Compounds 463

Analysis 15

Disease Data 188

News

- 26.09.2014 Tools catalogue added
- 23.09.2014 Disease Data collection available
- 23.09.2014 Analysis results for DrugMatrix, ESNATS and carcinoGENOMICS loaded

Links

- Submit your data to the [DDW](#)
- Training Resources
- Tools Catalogue
- diXa Main Website

Fig. S1 diXa Data Warehouse homepage

diXa Home / Studies
 34 Studies found

ID	Project	Title	Description
DIXA-001	New-Generis	Transcriptomic fingerprints in human peripheral blood mononuclear cells indicative of genotoxic and non-genotoxic carcinogenic exposure	For evaluating genotoxic exposure in human populations a number of biomarkers has been successfully applied over the last 30 years to determine early biological effects due to exposure to carcinogens. Despite their success, these early biological effects markers provide limited mechanistic insight, and are unable to detect exposure to non-genotoxic carcinogens. Gene expression profiling forms a promising tool for the
DIXA-002	CarcinoGenomics	Liver models - a successful tool for mechanism-based in vitro detection of genotoxicants	Among the different in vitro liver models used in the carcinogenomics project, the human HepaRG cell line generated the most robust classifier that discriminated the GTX from the NGTX carcinogens and NC. This mode could now be taken forward to develop a novel in vitro test to detect genotoxic compounds
DIXA-003	CarcinoGenomics	Kidney Models - A success Story	Two renal cell-lines (human and rat) were assayed for gene changes resulting from carcinogen exposure in order to develop models for detecting carcinogens. The novel RPTEC/TERT1 cell was found to maintain excellent characteristics of the proximal tubule including transport capabilities and maintenance of a primary cilium. It also showed normal chromosomes and nuclear stability. It was therefore chosen as the

Fig. S2 The Study list view shows all studies matching a search query or the complete list of studies when browsing.

ID	Study	Species	Cell Type	ChEMBL	Archive data file
DIXAS-5038011LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓
DIXAS-5038012LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓
DIXAS-5038021LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓
DIXAS-5038022LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓
DIXAS-5038031LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓
DIXAS-5038032LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓
DIXAS-5038041LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓
DIXAS-5038042LRA	DIXA-005	Rattus norvegicus	hepatocyte	CHEMBL295416	↓

Fig. S3 The Sample list view shows all samples matching a search query or the complete list of samples when browsing.

ID	Image	Compound Name	Synonyms	Known Drug	Std InChI Key	Smiles	ChemAgora Link	ChEMBL Link
CHEMBL101		PHENYLBUTAZONE	Diphenylbutazone; SID90341773; Antadol; Butazolidin; Phenylbutazone; Pro-Bute; Equiphar; SID855958; Elmedal;	Yes	VYMDGNCVAMGZFE-UHFFFAOYSA-N	<chem>CCCC1C(=O)N(N(C1=O)c2ccccc2)c3ccccc3</chem>	ChemAgora Portal	ChEMBL
CHEMBL101283		4-NITROBENZOIC ACID	SID26753070; SID17389791; 4-Nitro-Benzoic Acid; 4-Nitrobenzoic Acid; SID85147167	No	OTLNPYUJJOZPPA-UHFFFAOYSA-N	<chem>OC(=O)c1ccc(cc1)[N+](=O)[O-]</chem>	ChemAgora Portal	ChEMBL
CHEMBL1016		CANDESARTAN	Candesartan; SID50112728; CV-11974	No	HTQMVOVFRQIKW-UHFFFAOYSA-N	<chem>CCOc1nc2ccc(C(=O)O)c2n1Cc3ccc(cc3)c4ccccc4c5nn[nH]5</chem>	ChemAgora Portal	ChEMBL
CHEMBL101740		CIDOXEPIN	Sinequan; Zonalon; Doxepin; Cidoxepin	No	ODQWQRAPPTVAG-BOPFTXTBSA-N	<chem>CN(C)CC(C=C/11c2ccccc2Cc3ccccc13</chem>	ChemAgora Portal	ChEMBL
CHEMBL1024		IFOSFAMIDE	Ifex; SID11112548; SID56463644; Z-4942; SID301170; Ifosfamide; MJF-9325; Ifosamide	Yes	HOMGKSMUEGBAAB-UHFFFAOYSA-N	<chem>C1CCN1(=O)OCCC1N1CC1</chem>	ChemAgora Portal	ChEMBL

Fig. S4 The Compound list view shows all compounds matching a search query or the complete list of compounds when browsing.

The screenshot shows the 'Analysis' section of the diXa Warehouse. It features a search bar and a list of 15 analyses. The table below summarizes the visible entries:

ID	Title	Summary
AN-DIXA-001	Project Envirogenomarkers: QC, normalization, transformation and analyses of transcriptomics data of cases and controls of breast cancer and Non-Hodgkins lymphoma	The project concerns the large-scale application of omics technologies in a population study aiming at:
AN-DIXA-002	Project NewGeneris: QC, normalization, transformation and analyses of transcriptomics data from human peripheral blood mononuclear cells (PBMC)	Human peripheral blood mononuclear cells (PBMC) from healthy, non-smoking 2530 years old Norwegian volunteers (25 males and 35 females) were treated with 12 genotoxic or non-genotoxic compounds, respectively. Volunteers reported to be free of medication at the time of sampling. Cells were exposed in the presence of S9-mix for in vitro metabolic activation to a dose range of 3 concentrations and an untreated control. The platform used for gene expression profiling was Agilent 44k oligonucleotide microarrays. Exposures were competitively hybridized against the
AN-DIXA-003	Project PredTox: QC, normalization, transformation and analyses of transcriptomics data	The main goal of the EU InnoMed PredTox Project was to assess the value of combining results from omics technologies (proteomics, transcriptomics and metabonomics) together with the results from more conventional toxicology methods for more informed decision making earlier in preclinical safety evaluation. The objective was to better understand the mechanisms of liver and kidney toxicity and to find novel biomarkers of toxicity more sensitive and predictive than traditional toxicological parameters. Towards this goal, in vivo studies were conducted in rats (male Wistar
AN-DIXA-004	Project TG-GATEs Human in vitro hepatocytes: QC, normalization, transformation and analyses of transcriptomics data	The Toxicogenomics Project was a 5-year collaborative project (2002-2007) by a consortium comprising the Japanese government and several pharmaceutical companies. The project produced the Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (TG-GATEs), a large-scale database of transcriptomics and pathology data potentially useful for predicting the toxicity of new chemical entities. Conventional in vivo toxicology data was collected from single dose and repeat dosing studies on rats, and gene expression measured for the liver (and kidney in some

Fig. S5 The Analysis list view shows all analyses matching a search query or the complete list of analyses when browsing.

The screenshot shows the 'Disease Data' section of the diXa Warehouse. It features a search bar and a list of 188 disease data entries, displaying 1 to 50. The table below summarizes the visible entries:

ID	Project	Title	Description
DIXAD-1001	Liver disease collection	Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma	Background: It is a challenge to identify those patients who, after undergoing potentially curative treatments for hepatocellular carcinoma, are at greatest risk of recurrence. Such high-risk patients could receive novel
DIXAD-1002	Liver disease collection	Integrative Transcriptome Analysis Reveals Common Molecular Subtypes of Human Hepatocellular Carcinoma	Hepatocellular carcinoma (HCC) is a highly heterogeneous disease, and prior attempts to develop genomic-based classification for HCC have yielded highly divergent results, indicating difficulty in identifying unified molecular

Fig. S6 The Disease list view shows all diseases matching a search query or the complete list of diseases when browsing.

Study Details:

Internal ID: DIXA-001
 Source ID: E-GEOD-24891
 Project: New-Genes
 Title: Transcriptional fingerprints in human peripheral blood mononuclear cells indicative of genotoxic and non-genotoxic carcinogenic exposure

Description:
 For evaluating genotoxic exposure in human populations a number of biomarkers has been successfully applied over the last 30 years to determine early biological effects due to exposure to carcinogens. Despite their success, these early biological effects markers provide limited mechanistic insight, and are unable to detect exposure to non-genotoxic carcinogens. Gene expression profiling forms a promising tool for the development of new biomarkers in blood cells to overcome these limitations. The aim of our research was to identify novel genomics-based candidate markers for genotoxic and non-genotoxic carcinogenic exposure. Whole genome gene expression changes were investigated in human blood cells following in vivo exposure to a range of genotoxic and non-genotoxic carcinogenic compounds using whole genome microarrays. Sets of genes, as well as biological pathways indicative of genotoxic exposure and of non-genotoxic carcinogenic exposure were identified. Furthermore, networks were built using the genotoxic and non-genotoxic genes sets, showing the majority of the genes to be interlinked and revealing cooperative transcription factors for both classes. The identification of these potential candidate marker genes might contribute to the development of genomic based biomarkers of genotoxic exposure, and possibly even more importantly biomarkers of exposure to non-genotoxic carcinogens since presently no biomarkers are available. **Keywords:** Genome wide gene expression analysis, Transcriptional profile indicative of immunotoxic exposure For analysis of whole genome gene expression by microarray; PBMC from five independent donors per compound were exposed for 20 hours to three concentrations, i.e. the 100% and two serial ten-fold dilutions (10% and 1%), and a DMSO or PBS vehicle control. Exposed samples were always labelled with Cy3, whereas the vehicle control samples were labelled with Cy5, and were competitively hybridized on 444K Agilent microarrays.

Submission Date: 2007-01-01
 Factor Name: COMPOUND TOX CLASS CONCENTRATION DOSE
 Organism: HOMO sapiens
 Tissue: HUMAN PBMC
 Compound: 2,2',4,4'-TETRACHLOROBIPHENYL; 2,3,7,8-TETRACHLORODIBENZO-P-DIOXIN; 2-AMINO-1-METHYL-6-PHENYLMIDAZOLE-5-PYRROLE; 4-HYDROXYNINOSIN; ACRYLAMIDE; AFLATOXIN; ALCOHOL; BENZOPHENANTHRENE; IMIDAZOQUINOLINE; MALONALDEHYDE; N-NITROSODIMETHYLAMINE
 Measurement Type: Transcription profiling
 Technology Type: DNA microarray
 Technology Platform: Agilent
 Link to Sample Information: [click](#)
 Link to Data Files: [click](#)
 Link to ArrayExpress: [click](#)

Analysis with this Study:

id	Title
AN-DIXA-002	Project NewGenes: QC, normalization, transformation and analyses of transcriptomics data from human peripheral blood mononuclear cells (PBMC)

Fig. S7 The Study detail view shows additional information on a selected study. Furthermore, the study detail view provides links to the repositories where the experimental data are stored and links to detailed information on the compounds in the study.

Sample Details:

Sample ID: 5036011LRA
 Study: DIXA-005
 Species: Rattus norvegicus
 Cell Type: hepatocyte
 Sex: male
 ChEMBL: CHEMBL295416
 Treatment: Control
 Dose: 0
 Concentration Unit: micromolar
 Archive data file: [click](#)

© European Bioinformatics Institute 2013

Fig. S8 The Sample detail view shows additional information on a selected sample. Furthermore, the sample detail view provides links to ChEMBL and to the data file for the selected sample.

Compound Details:

ChEMBL ID: CHEMBL101

Compound Name: PHENYLBUTAZONE

Synonyms: Diphenylbutazone, SID90341773, Butazolidin, Phenylbutazone, Antadol, Pro-Bute, Equiphar, SID855958, Elmédal, SID85231197, Butaject, Azolid

Known Drug: Yes

Smiles: CCCCC1C(=O)N(N(C1=O)c2ccccc2)c3ccccc3

Standard InChI Key: VYMDGNCVAMGZFE-UHFFFAOYSA-N

Link to ChemAgora Portal: [click](#)

Image:

Link to ChEMBL: [click](#)

Studies with this Compound:

Id	Project	Title
DIXA-005	TG-GATES	Open TG-GATES (in vitro rat)
DIXA-006	TG-GATES	Open TG-GATES (in vitro, human)
DIXA-007	TG-GATES	Open TG-GATES (in vivo single dose)
DIXA-008	TG-GATES	Open TG-GATES (in vivo repeat dosing)
DIXA-030	Drug matrix	Rat toxicological microarrays kidney

Fig. S9 The Compound detail view shows additional information on a selected compound. Moreover, the compound detail view links to ChemAgora Portal, ChEMBL and studies with the selected compound.

Analysis Details:

Title: Project NewGeneris: QC, normalization, transformation and analyses of transcriptomics data from human peripheral blood mononuclear cells (PBMC)

Summary: Human peripheral blood mononuclear cells (PBMC) from healthy, non-smoking 20-30 years old Norwegian volunteers (25 males and 35 females) were treated with 12 genotoxic or non-genotoxic compounds, respectively. Volunteers reported to be free of medication at the time of sampling. Cells were exposed in the presence of 50-ns for in vitro metabolic activation to a dose range of 3 concentrations and an untreated control. The platform used for gene expression profiling was Agilent 44 oligonucleotide microarray. Exposures were competitive) replicates against the control for each donor. Initially, 2 doses of the 12 chemicals were analyzed but since proliferation data showed for some compounds very high cytotoxicity, exposures with lower concentration were included and the high concentrations were omitted.

Attachments:

1. Probes annotations AgilentList.txt
2. log2 signals median normalized filtered.txt
3. Data overview figures NewGeneris.pdf
4. log2 ratios.txt
5. BH q-values.txt
6. Exp annotations.txt
7. log2 signals median normalized not filtered.txt
8. average log2 ratios.txt

Study: [DIXA-001](#)

Studies in this Analysis:

Id	Project	Title
DIXA-001	New-Generis	Transcriptomic fingerprints in human peripheral blood mononuclear cells indicative of genotoxic and non-genotoxic carcinogenic exposure

© European Bioinformatics Institute 2013

Fig. S10 The Analysis detail view shows additional information on a selected analysis. Furthermore, the analysis detail view provides access to the files containing the results and studies related to a selected analysis.

Disease Data Details:	
Internal Id :	DIXAC-1001
Source Id :	E-GEOD-10143
Project :	Liver disease collection
Title :	Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma
Description :	<p>Background: It is a challenge to identify those patients who, after undergoing potentially curative treatments for hepatocellular carcinoma, are at greatest risk of recurrence. Such high-risk patients could receive novel interventional measures. An obstacle to the development of genome-based predictors of outcome in patients with hepatocellular carcinoma has been the lack of a means to carry out genome-wide expression profiling of fixed, as opposed to frozen, tissues. Methods: We aimed to demonstrate the feasibility of gene-expression profiling of more than 6000 human genes in formalin-fixed paraffin-embedded tissues. We applied the method to tissues from 307 patients with hepatocellular carcinoma, from four series of patients, to discover and validate a gene-expression signature associated with survival. Results: The expression-profiling method for formalin-fixed, paraffin-embedded tissue was highly effective: samples from 92% of the patients yielded data of high quality, including samples that had been archived for more than 24 years. Gene-expression profiles of tumor tissue failed to yield a significant association with survival. In contrast, profiles of the surrounding nontumoral liver tissue were highly correlated with survival in a training set of 82 Japanese patients, and the signature was validated in tissues from an independent group of 225 patients from the United States and Europe ($p = 0.04$). Conclusions: We have demonstrated the feasibility of genome-wide expression profiling of formalin-fixed, paraffin-embedded tissues and have shown that a reproducible gene-expression signature correlating with survival is present in liver tissue adjacent to the tumor in patients with hepatocellular carcinoma. This SuperSeries is composed of the following subset Series: GSE10140: Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma (Training Set, Liver) GSE10141: Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma (Training Set, HCC) GSE10142: Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma (Validation Set) Keywords: Hepatocellular carcinoma, Expression array, Illumina, Signatures, Outcome prediction Training cohort: 80 tumor and 82 non-tumor liver tissues surgically resected from patients with hepatocellular carcinoma (HCC); Validation cohort: 225 non-tumor liver tissues surgically resected from patients with HCC. Clinical data has been withheld from GEO due to privacy concerns.</p>
Submission Date :	10/01/2008
Factor Name :	ETHNICITY CORRESPONDING TUMOR SAMPLE IN GSE19977 USED FOR PROGNOSTIC PREDICTION EGF LEVEL
Organism :	Homo sapiens
ICD10 accession :	C22.0 B18.1 B18.2 K70.9
Disease name :	Liver cell carcinoma Chronic viral hepatitis B without delta-agent Chronic viral hepatitis C Alcoholic liver disease, unspecified
Origin :	In vivo
Assay type :	mRNA
Measurement Type :	transcription profiling
Technology Type :	array
Link to Sample Information :	click
Link to Data Files :	click
Link to ArrayExpress :	click

Fig. S11 The Disease detail view shows additional information on a selected disease. Furthermore, the disease detail view provides links to the repositories where the experimental data are stored.

Table S3 Overview of ontologies used by diXa ISA template.

Ontology	Weblink
NCBI Organismal Classification (NCBITAXON)	http://bioportal.bioontology.org/ontologies/NCBITAXON
Units of Measurements Ontology (UO)	http://bioportal.bioontology.org/ontologies/UO
Ontology for Biomedical Investigations (OBI)	http://bioportal.bioontology.org/ontologies/OBI
Chemical Entities of Biological Interest (ChEBI)	http://bioportal.bioontology.org/ontologies/CHEBI
NCI Thesaurus (NCIT)	http://bioportal.bioontology.org/ontologies/NCIT

Table S4 Overview of currently available analysis descriptions, together with their location on the diXa website.

Analysis description	Weblink
Project Envirogenomarkers: QC, normalization, transformation and analyses of transcriptomics data of cases and controls of breast cancer and Non-Hodgkins lymphoma	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-001
Project NewGeneris: QC, normalization, transformation and analyses of transcriptomics data from human peripheral blood mononuclear cells (PBMC)	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-002
Project PredTox: QC, normalization, transformation and analyses of transcriptomics data	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-003
Project TG-GATEs Human in vitro hepatocytes: QC, normalization, transformation and analyses of transcriptomics data	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-004
Project TG-GATEs Rat in vitro hepatocytes: QC, normalization, transformation and analyses of transcriptomics data	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-005
Project TG-GATEs Rat in vivo Kidney: QC, normalization, transformation and analyses of transcriptomics data	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-006
Project TG-GATEs Rat in vivo Liver: QC, normalization, transformation and analyses of transcriptomics data	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-007
Project Predictomics: QC, normalization, transformation and analyses of transcriptomics data of test substances (hepatotoxins) in human primary hepatocytes and HepG2 cells	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-008
Project DrugMatrix: QC, normalization, transformation and analyses of transcriptomics data from rat kidney, heart, liver, skeletal muscle tissue, and rat hepatocytes	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-009
Project ESNATS: QC, normalization, transformation and analyses of transcriptomics data from human embryonic stem cells	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-010
Project PredTox: QC, normalization, transformation and analyses of serum metabolites by LC/MS	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-011
Project carcinoGENOMICS: QC, normalization, transformation and analyses of transcriptomics data of the kidney (RPTEC/TERT1) and liver (HepaRG) cell lines	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-012
Project carcinoGENOMICS: QC, normalization, transformation and analyses of transcriptomics data of the liver cell line HepG2	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-013
Project carcinoGENOMICS: QC, normalization, transformation and analyses of transcriptomics data of Human Embryonic Stem Cell Derived Hepatocyte-Like Cells (hESC DE-Hep)	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-014
Project carcinoGENOMICS: QC, normalization, transformation and analyses of transcriptomics data in a lung epithelial cell-based test system	http://wwwdev.ebi.ac.uk/fg/dixa/group/AN-DIXA-015

References for Table S2 and the text under Table S2.

- Davis, A.P., *et al.* (2013) The Comparative Toxicogenomics Database: update 2013, *Nucleic Acids Res.*, **41**, D1104-1114.
- Degtyarenko, K., *et al.* (2008) ChEBI: a database and ontology for chemical entities of biological interest, *Nucleic Acids Res.*, **36**, D344-D350.
- Kamburov, A., *et al.* (2013) The ConsensusPathDB interaction database: 2013 update, *Nucleic Acids Res.*, **41**, D793-D800.
- Law, V., *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism, *Nucleic acids research*, **42**, D1091-1097.
- Pence, H.E. and Williams, A. (2010) ChemSpider: an online chemical information resource, *J. Chem. Educ.*, **87**, 1123-1124.
- Rocca-Serra, P., *et al.* (2010) ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level, *Bioinformatics*, **26**, 2354-2356.
- Rustici, G., *et al.* (2013) ArrayExpress update - trends in database growth and links to data analysis tools, *Nucl. Acids Res*, **41**, D987-D990.
- Wang, Y., *et al.* (2009) PubChem: a public information system for analyzing bioactivities of small molecules, *Nucleic acids research*, **37**, W623-633.
- Warr, W. (2009) ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI), *J Comput Aid Mol Des*, **23**, 195-198.
- Wexler, P. (2001) TOXNET: an evolving web resource for toxicology and environmental health information, *Toxicology*, **157**, 3-10.
- Wittwehr, C. (2011) Neuer Zugang zu Chemikalien-Daten. eChemPortal, *Chemie in unserer Zeit*, **45**, 122-125.