



An innovative index to incorporate transcriptomic data into weight of evidence approaches for environmental risk assessment

Martina Cecchetto^{a,1}, Luca Peruzza^{b,1}, Elisa Giubilato^a, Iliaria Bernardini^b,
Giulia Dalla Rovere^b, Antonio Marcomini^a, Francesco Regoli^{c,d}, Luca Bargelloni^b,
Tomaso Patarnello^b, Elena Semenzin^{a,*}, Massimo Milan^{b,d}

^a Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, via Torino 155, 30172, Mestre-Venezia, Italy

^b Department of Comparative Biomedicine and Food Science, University of Padova, Viale dell'Università 16, 35020, Legnaro, Padova, Italy

^c Department of Life and Environmental Sciences, Marche Polytechnic University, Via Breccia Bianche, 60131, Ancona, Italy

^d NFBC, National Future Biodiversity Center, Palermo, Italy

ARTICLE INFO

Handling Editor: Dr Robert Letcher

Keywords:

Transcriptomic hazard index
Gene expression profiling
Emerging contaminants
Eco-toxicogenomics
Weight of evidence
Environmental risk assessment

ABSTRACT

The sharp decrease in the cost of RNA-sequencing and the rapid improvement in computational analysis of ecotoxicogenomic data have brought new insights into the adverse effects of chemicals on aquatic organisms. Yet, transcriptomics is generally applied qualitatively in environmental risk assessments, hampering more effective exploitation of this evidence through multidisciplinary studies. In view of this limitation, a methodology is here presented to quantitatively elaborate transcriptional data in support to environmental risk assessment. The proposed methodology makes use of results from the application of Gene Set Enrichment Analysis to recent studies investigating the response of *Mytilus galloprovincialis* and *Ruditapes philippinarum* exposed to contaminants of emerging concern. The degree of changes in gene sets and the relevance of physiological reactions are integrated in the calculation of a hazard index. The outcome is then classified according to five hazard classes (from absent to severe), providing an evaluation of whole-transcriptome effects of chemical exposure. The application to experimental and simulated datasets proved that the method can effectively discriminate different levels of altered transcriptomic responses when compared to expert judgement (Spearman correlation coefficient of 0.96). A further application to data collected in two independent studies of *Salmo trutta* and *Xenopus tropicalis* exposed to contaminants confirmed the potential extension of the methodology to other aquatic species. This methodology can serve as a proof of concept for the integration of "genomic tools" in environmental risk assessment based on multidisciplinary investigations. To this end, the proposed transcriptomic hazard index can now be incorporated into quantitative Weight of Evidence approaches and weighed, with results from other types of analysis, to elucidate the role of chemicals in adverse ecological effects.

1. Introduction

The functioning of an aquatic ecosystem is the result of complex interactions often under growing anthropogenic pressure (Borgwardt et al., 2019), with chemical contamination being one of the most alarming threats. The classification of ecosystem health status should be based on a broad range of indicators that are able to evaluate not only the severity of the chemical contamination affecting its different compartments, but also the biological outcomes on aquatic organisms inhabiting the ecosystem (WFD 2000/60/EC).

As recently debated in Suter (2021) and in Johnson et al. (2021), when dealing with multiple and heterogeneous bodies of evidence, the use of an integrated framework should be encouraged to comprehend causal relations and correctly estimate the likelihood of hazard posed by chemical contaminants of concern. The integration of different studies can be pursued by means of a quantitative Weight of Evidence (QWoE) approach (US Environmental Protection Agency, 2016), with several examples of methods and applications available in the literature (e.g., Gottardo et al., 2011; Micheletti et al., 2011; Piva et al., 2011; Benedetti et al., 2012; Bebianno et al., 2015; Fan et al., 2015; Barjhoux et al., 2018;

* Corresponding author.

E-mail address: semenzin@unive.it (E. Semenzin).

¹ M.C. and L.P. contributed equally to this paper.

Lehtonen et al., 2019). In a QWoE approach, multiple and heterogeneous data are organized into distinct lines of evidence (LoEs) (Chapman et al., 2002), to be then denoted by synthetic indices specifying the class of hazard, and eventually harmonized together using weighing or ranking (Linkov et al., 2009). Because of the modularity and flexibility of such models, the possibility of adding other relevant or site-specific LoEs should be considered as long as they can be quantitatively integrated with other lines (Suter and Cormier, 2011). Yet, in QWoE approaches, questions remain over the feasibility of bringing together the most validated types of analyses with evidence from effect-based tools (Solimini et al., 2009), such as transcriptomics (Wernersson et al., 2015).

In recent years, the sharp decrease in RNA-sequencing costs and the rapid improvement in computational analysis encouraged the application of transcriptomics in ecotoxicology (Sauer et al., 2017). By revealing temporary and/or persistent alterations in gene expression at the whole genome level, transcriptomics can help clarify contaminants mechanisms or mode of action (MoA) (Ankley et al., 2006), identify specific molecular fingerprints of chemical stress (Milan et al., 2015), contribute to the delineation of Adverse Outcome Pathways (AOPs; Bajard et al., 2023), or derive a no effect threshold for transcriptional response (Villeneuve et al., 2012; Pagé-Larivière et al., 2019). Being sensitive in the recognition of exposure makes transcriptomics an effective complementary investigation to biomarkers and bioassays analyses.

However, mostly because of difficulties in generating and interpreting transcriptomic data consistently under a common framework (Verheijen et al., 2020), transcriptomics has served primarily as a qualitative tool. Also in recent WoE applications, transcriptional evidence was evaluated qualitatively without taking part in the weighted integrative process (e.g., Mezzelani et al., 2021; Schmitz et al., 2022; Lucia et al., 2023). This represents a missed opportunity for transcriptomics, and -omics data in general, to be considered within QWoE approaches for environmental risk assessment (ERA) and regulatory purposes (Soufan et al., 2022).

In this respect, a novel methodology is here proposed with the aim of synthesizing transcriptomic data into a hazard index so that transcriptomics can in future constitute an additional piece of information in QWoE schemes. To this end, transcriptomic alterations measured in two sentinel species, the Manila clam *Ruditapes philippinarum* (hereafter *Rp*) and the Mediterranean mussel *Mytilus galloprovincialis* (hereafter *Mg*), were elicited with Gene Set Enrichment Analysis (GSEA, Subramanian et al., 2005) and used to develop a transcriptomics hazard index (section Materials and methods). Both species have high ecological and economic importance and are widely used to characterize the effects of emerging contaminants (Bernardini et al., 2021) and the impact of anthropogenic activities in marine ecosystems (d'Errico et al., 2021). The ability of the synthetic index to mirror the expert judgment of transcriptional alterations was tested with both simulated and experimental transcriptomics profiles of these species (section Results and discussion). The methodology was further tested on two independent studies, one of brown trout *Salmo trutta* (hereafter *St*) exposed to herbicides (Webster and Santos, 2015) and a second one of western clawed frog *Xenopus tropicalis* (hereafter *Xt*) exposed to benzo[a]pyrene (Regnault et al., 2014). The index was able to disentangle the transcriptional changes between the different concentrations of chemicals and sampling times used in the studies, demonstrating its potential applicability to other species. Challenges, opportunities and future developments of the methodology are discussed in the last section of the manuscript.

The development of such hazard index based on transcriptomic data should facilitate the integration of -omics tools in environmental risk assessment. The ultimate goal is to broaden the range of modules available in QWoE studies for fostering the understanding of the ecological status of aquatic environments.

2. Materials and methods

2.1. Rationale for transcriptomic data processing

One of the main challenges in transcriptomics is dealing with high dimensional data. In particular, gaining biological insight from the generated data. Single-gene analysis, implemented to draw a list of differentially expressed genes (DEGs) between a sample and a control condition, can lead to a misinterpretation of the biological outcomes because of the many false negatives, the arbitrary choice of a cut-off value, and the disregarding of multiple functions operating by a single gene in different cellular processes (Maleki et al., 2020). In light of these shortcomings, the present work makes use of gene set enrichment analysis (GSEA) in its revised form proposed by Subramanian et al. (2005). GSEA is a non-parametric statistical approach that ranks all genes in the dataset on the basis of expression criteria (typically the Fold Change) but allows to consider all information from an expression matrix without an *a priori* definition of DEGs lists. GSEA has found applications in different fields, from drug response to complex diseases studies, becoming one of the most widely used gene set analysis methods (i.e. about 34.000 citations in Google Scholar).

As pointed out by Maleki et al. (2020), there is no general consensus on which methods outperforms the others and GSEA is not without flaws either. However, GSEA is here reckoned as a good compromise when it comes to more computationally demanding analysis, methods with stringent assumptions, or analyses characterized by high sensitivity but low specificity.

GSEA reveals, within the same biological pathway/process, coordinated shifts in gene expression across multiple genes and associates to such coordinated shifts an "intensity index", the Enrichment Score (ES), and a *p*-value for the significance of this effect. The ES embodies the degree to which a gene set is over or under-represented. To calculate the ES, a running sum increases when walking down the ranked list of genes, ordered for decreasing values of gene score, a gene belonging to the gene set is encountered, while it decreases if meeting a gene not part of the gene set. Given its definition, the ES can vary between +1, when all genes belonging to the gene set are displayed at the top of the list (up-regulated compared to the control group), and -1 when genes are all located at the bottom of the list (down-regulated compared to the control group). Here, the ES has been considered to translate the significant pathways (i.e. False Discovery Rate, FDR, <0.2) identified by the GSEA into a quantitative hazard index as it is explained in the next section.

The choice to deem pathways with FDR <0.2 as significant stems from two considerations: *i*) the default FDR threshold suggested by the authors of the GSEA method is 0.25; *ii*) the functional database used to build the hazard index is composed of only 48 pathways (see below for more details), thus implying a limited number of multiple comparisons. For these reasons, the 0.2 FDR threshold represents a good compromise for the purpose of the following analysis.

GSEA requires the *a priori* definition of gene sets (e.g., KEGG pathways, Reactomes, Gene Ontology) that are responsible for driving a specific biological function; hence, a key aspect in implementing the analysis is the identification of an appropriate collection of gene sets. Considering that a single gene can be accountable for multiple functions, redundancy in gene sets may lead to over-representation of some biological responses hindering the identification of others in the following statistical analysis. To overcome redundancy issues, the present study relies on the "Hallmark" gene set collection presented in Liberzon et al. (2015). Each hallmark gene set represents a distinct biological process that is then assigned to one of eight higher biological categories following Dean et al. (2017): cellular component, development, DNA damage, immune, metabolism, stress response, proliferation, and signaling (see detail in Supplementary information, Table S1).

To further reduce the redundancy between gene sets compositions, before application, the Hallmark dataset was cross checked to identify gene sets overlaps. The immune response gene sets "HALLMARK

INTERFERON GAMMA RESPONSE” and “HALLMARK INTERFERON ALPHA RESPONSE”, which shared 70 genes (corresponding to 36% and 75% of the entire “HALLMARK INTERFERON GAMMA RESPONSE” and “HALLMARK INTERFERON ALPHA RESPONSE”, respectively), were merged creating a unique inclusive hallmark representing both hallmark sets. The same approach was applied to “HALLMARK ESTROGEN RESPONSE EARLY” and “HALLMARK ESTROGEN RESPONSE LATE”, characterized originally by an overlap of 50%, eventually leaving the hallmark database with 48 gene sets out of the original 50 sets.

The choice to rely on functional information based on human despite the current work being conducted on very distant non-model species, mostly stems from the strengths of the Hallmark database: *i*) it is generated from one of the most widely used and comprehensive functional databases of gene sets; *ii*) it was generated by a combination of automated approaches and expert curation, assuring refinement and high accuracy and reducing variation and redundancy in gene sets; *iii*) it can be easily “customised” by removing pathways that are not known/described in the studied species. In this regard, when applied to bivalves, the Hallmark database was further refined eliminating four gene sets that were never described in these taxa i.e., “HALLMARK ALLOGRAFT REJECTION”, “HALLMARK HEME METABOLISM”, “HALLMARK KRAS SIGNALING DN”, and “HALLMARK KRAS SIGNALING UP”.

The retained Hallmark gene sets remained well populated with an average of about 70 genes ascribed to each gene set. Clam and mussel transcriptomes shared an average of 50% of genes with the human Hallmark gene sets. The distribution of biological categories, along with

annotated Hallmark gene sets, in terms of gene size, are reported in Fig. 1 for the bivalve species of interest, and listed in the SI (Table S1).

2.2. Elaboration of a transcriptomic quantitative index

With the objective of integrating GSEA results into environmental risk assessment, a new index was developed to translate transcriptomic data into a quantitative evaluation of risk.

The statistically significant gene sets indicate which categories are subject to transcriptional alteration after exposure to chemical contaminants. To elucidate the severity of such alteration, specific algorithms were developed to account for both the relevance of each category in terms of physiological reactions and the degree of its differential regulation. The conceptual framework finds inspiration in the mathematical structure of the LoE outlined for biomarkers in Piva et al. (2011) and, as subsequently modified, in Regoli et al. (2019).

First, to each of the eight biological categories a “weight” between 1 and 3 was assigned based on the relevance of the biological endpoint triggered by the change in transcripts, reflecting, where possible, the indication proposed by Piva et al. (2011) for biomarkers LoE. The weight 1 was given to the category “cellular components” that includes gene sets for which changes at transcriptional level are less evidently associated to the onset of adverse effects at higher biological levels. The weight 1.5 was assigned to the categories “immune” and “metabolism” because they may prelude adverse effects at a higher biological level but for which transcriptional/biological changes could be reversible. The

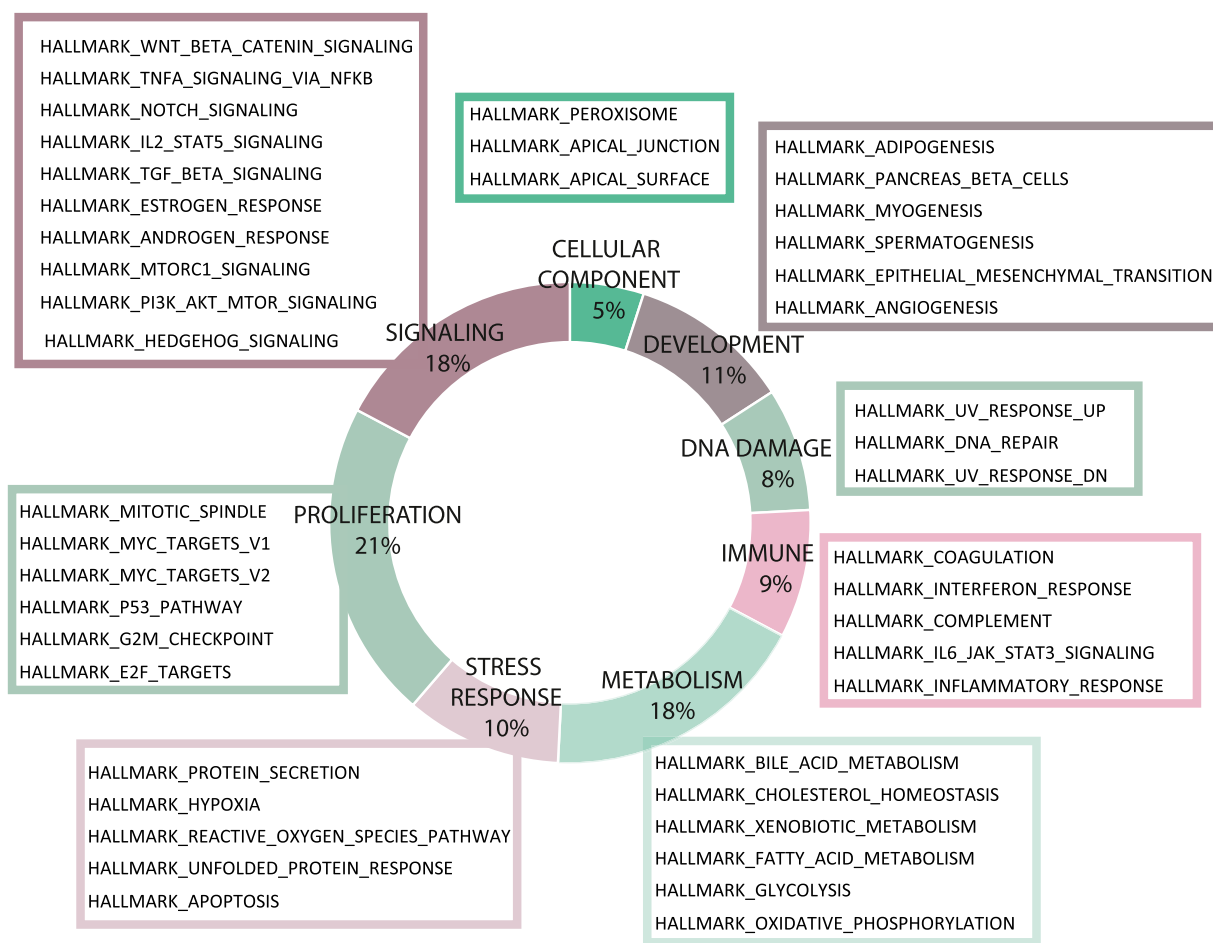


Fig. 1. Plot depicting the Hallmark gene sets selected for this study, grouped together in eight higher categories. Percentages reflect the number of genes involved in each category for bivalves. The list of gene sets is reported also in SI (Table S1).

weight 2 was attributed to the category “stress response” that includes gene sets prognostic of impairment at higher levels of biological organization (e.g. “apoptosis” among others). The weight 2.5 was given to “DNA damage”, “development” and “signaling” because they include pathways potentially reflecting dysfunctions at hormonal, reproductive and developmental level. Weight 3 was assigned to “proliferation” as it reflects alterations related to cell cycle regulation, cell death and cancer. The full list of gene sets with assigned weights is reported in [Table S1 of Supplementary Information](#). It should be highlighted that the weighting system can be revised if additional mechanistic knowledge of biological functioning triggered by alterations in the transcriptome becomes available. As showed here, it can also be tailored to the tested species.

The relevance of differential regulation is represented by the ratio between the observed cumulative ES of significant gene sets and the maximum ES attainable within a specific category. The maximum ES is obtained by adding together the highest level of differential regulation potentially achievable in each molecular pathway, i.e. $|ES| = 1$, contributing to the same category. For example, the category “metabolism” consists of six gene sets, whose maximum attainable ES in absolute value is 1, hence the highest level of differential regulation is equal to 6.

In the proposed approach, the transcriptional effect observed for the category i is defined according to Equation (1), where the category weight w_i , is normalized by the maximum value $w_{max} = 3$ (maximum assigned weight value) and the sum of the statistically significant ESs is divided by total ES attainable for the biological category, i.e. $ES_{max,i}$.

$$E_i = \frac{w_i}{w_{max}} \cdot \frac{\sum_{gene\ sets \in i} |ES|}{ES_{max,i}} \cdot 100 \quad (1)$$

So defined, the transcriptomic effect for each category can vary from 0% to 100%, with the latter value representing the case of maximum differential regulation achieved simultaneously by all the pathways contributing to a category whose alterations can lead to cell death, apoptosis, and cancer (i.e. for which $w = 3$). Depending on the value of E_i , five classes of effect have been defined, namely: A - Absent ($E_i = 0\%$), B - Slight ($0\% < E_i \leq 7\%$), C - Moderate ($7\% < E_i \leq 25\%$), D - Major ($25\% < E_i \leq 40\%$), E - Severe ($40\% < E_i \leq 100\%$) ([Fig. 2](#)).

Finally, the Transcriptomics Hazard Index (THI) associated with the overall transcriptomics results is calculated based on the percentages of categories included in each class of effect, as described in Equation (2).

$$THI = \%categories_A \cdot a + \%categories_B \cdot b + \%categories_C \cdot c + \%categories_D \cdot d + \%categories_E \cdot e \quad (2)$$

With $a = 0$, $b = 0.7$, $c = 2.5$, $d = 4$ and $e = 10$. Factors a to e find correspondence with the upper bound of each class of hazard, and the increasing values make so that larger importance is given to categories that are progressively more affected by transcriptomics variations ([Fig. 2](#)).

Five classes for the final THI have been defined, namely: Absent, Slight, Moderate, Major, and Severe. As shown in [Fig. 2](#), the class of hazard results Absent when THI is equal to zero. When THI varies between 0 and 70 (70 corresponding to all categories with slight effects), the hazard is Slight; THI from 70 to 250 (the maximum value when all categories have moderate effect) classifies the hazard as Moderate, while from 250 to 400 it is considered Major. Severe hazard index ranges between 400 and 1000, the latter caused by all categories with severe effects.

The logical steps of the above procedure as well as classes boundary values were calibrated and validated using data, reported in the following sections, from experimental studies along with transcriptional situations made up to simulate a vast range of genomic alterations.

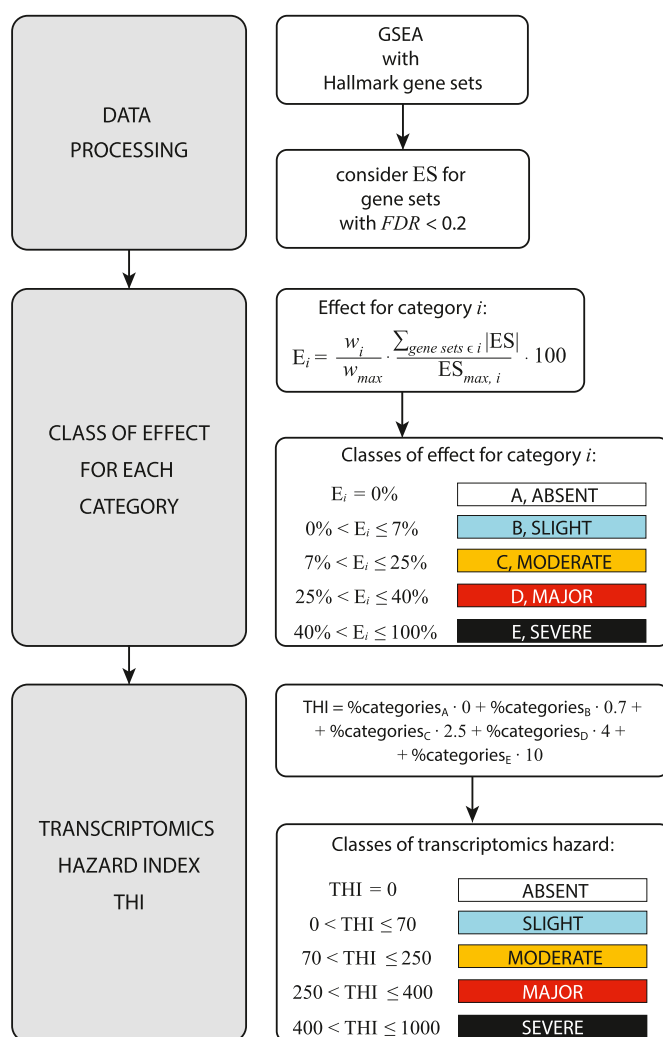


Fig. 2. Flowchart for the calculations of the Transcriptomics Hazard Index.

2.3. Calibration, validation and testing of the methodology

Calibration and validation of the proposed methodology for THI estimation were carried out using experimental and simulated datasets. Detailed information about the experimental datasets and the bioinformatic pipeline used to produce the GSEA input files is reported in the following section.

A sub-set of experimental and simulated datasets was used to calibrate the boundaries between classes of effect and, consequently, for the determination of coefficients to derive the five classes of hazard. For both experimental and simulated datasets, calibration was performed by means of expert judgment in order to align the THI output with the expert qualitative interpretation of the transcriptomic altered profiles. Interpretations were based on the number of differentially expressed pathways, their degree of differential regulation, and the significance of the alteration in terms of physiological outcomes. All together, these factors supported the assignment of a semi-quantitative hazard class, from absent to severe, to each condition.

Once thresholds and coefficients were calibrated, the index was calculated for the second set of experimental and simulated datasets to assess its performance. For this purpose, similar to calibration, the THI output was compared with the expert qualitative evaluation. The versatility of the methodology was assessed on two independent studies of aquatic vertebrate species exposed to contaminants under laboratory conditions.

To further assess the reliability of the index, the THI outputs were

evaluated against different quantitative summaries of transcriptomics results, i.e., DEGs, number of GSEA statistically significant pathways and average ESs.

Finally, the fitness of THI was tested by independent peer-evaluation of ten datasets of GSEA outputs picked from both experimental and simulated datasets so that the risk ranged from slight to severe (datasets are reported in Table S2, section S5, of the SI). To this end, an independent panel of eleven experts from different institutions, selected for their expertise in transcriptomics, were presented with the list of statistically significant Hallmark gene sets generated by the GSEA and the respective enrichment score as they appear in the SI for each dataset. After a debrief on gene sets grouped into biological categories with assigned weights and the possible range of variation of the ES, experts were asked to assign a class of hazard, from *absent* to *severe*, to each dataset. The evaluations of the independent panel were compared with the THI results gained on the same datasets and the THI performance was assessed by means of a non-parametric Spearman correlation test.

2.4. Experimental and simulated data

Data from previous experiments consisted of transcriptomic analysis

Table 1

Datasets with relative information about chemical/stressor, exposure dose and time, tested species, and additional information on biological replicates and technical procedures. Reference studies are cited in progressive roman numbers next to the tested chemical/stressor in the first column.

Chemical/Stressor	Dose	Exposure time	Species	Number of tested biological replicates	Development stage	Methodology (RNA extraction and library preparation)	Sequencing approach
C6O4 ^(I)	0.1 µg/l	7 days	<i>Rp</i>	5 pools (each composed by 4 individuals) for condition/sampling time	Adults (3.64 ± 0.32 cm shell length)	Total RNA extraction: RNeasy Mini Kit Qiagen with additional DNase treatment; Libraries Preparation: Illumina TruSeq RNA Library Prep Kit.	Illumina Novaseq 6000 with a paired-end 2 × 100 bp setup
C6O4 ^(I)	1 µg/l	7 days					
PFOA ^(I)	1 µg/l	7 days					
C6O4 ^(I)	0.1 µg/l	21 days					
C6O4 ^(I)	1 µg/l	21 days					
PFOA ^(I)	1 µg/l	21 days	<i>Mg</i>	5 pools (each composed by 5 individuals) for each condition	Adults (6.0 ± 0.5 cm shell length)	Total RNA extraction: RNeasy Mini Kit Qiagen with additional DNase treatment; Libraries Preparation: Agilent Sure Select Strand-Specific mRNA Library	Illumina HighSeq 4000 with a single 1*100 bp setup
Acidification and Carbamazepine ^(II)	pH 7.6, 1 µg/l	28 days					
Acidification ^(III)	pH 7.6	28 days					
Carbamazepine ^(III)	1 µg/l	28 days	<i>Mg</i>	6 individuals for each condition/sampling time	Adults (5–6 cm shell length)	Total RNA extraction: RNeasy Mini Kit Qiagen with additional DNase treatment; Libraries Preparation: Agilent Sure Select Strand-Specific mRNA Library	Illumina HighSeq 4000 with a single 1*100 bp setup
Glyphosate ^(III)	100 µg/l	7 days					
AMPA ^(III)	100 µg/l	7 days					
Glyphosate and AMPA ^(III)	100 µg/l	7 days					
Glyphosate ^(III)	100 µg/l	21 days					
AMPA ^(III)	100 µg/l	21 days	<i>Mg</i>	4 pools (each composed by 5 individuals) for each condition	Adults (6.5–7 cm shell length)	Total RNA extraction: RNeasy Mini Kit Qiagen with additional DNase treatment; Libraries Preparation: Agilent Sure Select Strand-Specific mRNA Library	Illumina HighSeq 4000 with a single 1*50 bp setup
Glyphosate and AMPA ^(III)	100 µg/l	21 days					
Glyphosate ^(IV)	10 µg/l	21 days					
Glyphosate ^(IV)	100 µg/l	21 days					
Glyphosate ^(IV)	1000 µg/l	21 days					
Amyl Salicylate ^(V)	0.1 µg/l	7 days	<i>Mg</i>	5 pools (each composed by 4 individuals) for each condition/sampling time	Adults (5.4 ± 0.5 cm shell length)	Total RNA extraction: RNeasy Mini Kit Qiagen with additional DNase treatment; Libraries preparation: 3' QuantSeq kit (Lexogen)	NextSeq 500 Illumina with a single 1*75 bp setup
Amyl Salicylate ^(V)	0.5 µg/l	7 days					
Amyl Salicylate ^(V)	0.1 µg/l	14 days					
Amyl Salicylate ^(V)	0.5 µg/l	14 days					
Glyphosate ^(VI)	10 µg/l	14 days					
Glyphosate ^(VI)	10000 µg/l	14 days	<i>St</i>	6 individuals for each condition	Juvenile female brown trout (six months old)	Total RNA extraction: RNeasy Mini kit with additional DNase treatment; Libraries preparation: Illumina TruSeq RNA Sample Preparation kit.	Illumina HiSeq 2500 (100 bp paired reads)
Roundup ^(VI)	10 µg/l	14 days					
Roundup ^(VI)	500 µg/l	14 days					
Roundup ^(VI)	10000 µg/l	14 days	<i>Xt</i>	3 individuals for each condition/sampling time	Adults <i>Xenopus tropicalis</i> female	RNA extraction using the RNAqueous®-4PCR Kit; Library preparation; double-strand fragmented cDNA using superscript II	Illumina Genome Analyzer II (75 bp single-end reads)
Benzo[a]pyrene ^(VII)	10 µg/l	6 h					
Benzo[a]pyrene ^(VII)	10 µg/l	12 h					
Benzo[a]pyrene ^(VII)	10 µg/l	18 h					
Benzo[a]pyrene ^(VII)	10 µg/l	24 h					

I: Bernardini et al. (2021); II: Mezzelani et al. (2021); III: Iori et al. (2020); IV: Milan et al. (2018); V: Bernardini et al. (2022); VI: Webster and Santos (2015); VII: Regnault et al. (2014).

tested doses, exposure times, target species, biological replicates and development stage, technical procedures), while more details are reported in the Supplementary Information (section S2) and in the cited studies.

Raw transcriptomics data were generated using the protocols detailed in the respective published studies. Once the raw count tables were obtained, data were normalized using the “RUVs” function from the RUVseq library protocol by [Risso et al. \(2014\)](#). After that, normalized counts were imported in edgeR and subjected to Differential Expression Analysis, following the steps detailed in section 2.3 of the RUVseq package vignette. The differential expression table containing the *p*-value and log fold change for each tested gene in the transcriptome was saved.

As previously mentioned, Hallmark gene sets are named following the HUGO (Human Genome Organization) nomenclature, which uses human gene symbols. For the mollusc datasets, sequence similarities between the predicted protein set of bivalve species and the Ensembl human proteome were obtained using protein BLAST (Basic Local Alignment Search Tool, [Altschul et al., 1990](#)) with an *e*-value cut-off $<10^{-6}$. BLAST was chosen because, due to the taxonomic distance between human and mollusc, a homology-based approach (e.g. OrthoFinder) would have been too strict, leaving the majority of the genes in transcriptomes without an orthologous gene and hence without a functional annotation. To overcome this issue and, at the same time, trying to be as strict as possible, protein BLAST was performed with a conservative threshold *E*-value of 10^{-6} instead of the threshold commonly accepted for protein-BLAST (i.e. *E*-value threshold of 10^{-3}).

The differential expression table with the updated gene ids and respective logFC was then used as input for GSEA, which was run inside the R software (via the “clusterProfiler” R library, [Wu et al., 2021](#)). The scripts to run the GSEA pipeline can be found at <https://github.com/GEMMA-BCA/TranscriptomicsWeightOfEvidence>.

Datasets used for calibration and validation consisted of statistically significant gene sets with their relative ES. To create a broader variety of scenarios, simulated datasets complemented the experimental ones, and were created *ad-hoc* to resemble plausible GSEA outputs. To characterize each scenario in terms of transcriptional alterations, a list of statistically significant gene sets with assigned ESs were fictively assembled by the authors as also presented in the Supplementary Information at Chapter S4.

3. Results and discussion

3.1. THI application

Results are presented in [Table 2](#) (and in chapter S3 in the SI), where priority has been given to results obtained for the experimental datasets, whereas the expected hazard and THI outcome for the simulated scenarios are included in the Supplementary Information (Chapter S4). In [Table 2](#), for each dataset a class of transcriptomic hazard is provided based on the THI value obtained. To facilitate the visualization of the hazard magnitude, classes' intervals were scaled to a common width and graphically represented by a hazard-meter. THI was converted accordingly and its value is indicated by a hazard pointer.

If only classes were to be considered as hazard indication, results for the experimental datasets would covered most of the outcomes available in the method, spanning from “slight” to “major”. Although, a closer look at the hazard-meters provides additional information to discriminate different situations between or within classes of hazard. This is particularly relevant for transcriptional alterations that fall close to the threshold between classes and for which the level of hazard could be misinterpreted if the hazard class was to be considered alone (e.g., *Mg* exposed for 7 days to 100 µg/l of AMPA, hazard class “moderate” but THI = 71, very close to the boundary with the “slight” class). Another example of hazard-meters assisting in the interpretation of the chemical risk is provided by experiments with clams exposed to two C6O4

concentrations (0.1 µg/l and 1 µg/l) and PFOA. Despite the resulting class indicates “moderate” risk in all three treatments (at both 7 and 21 days), the highest THI was obtained for clams exposed to the lower C6O4 concentration. This finding echoes the most important transcriptional changes observed in correspondence to 0.1 µg/l of C6O4 by [Bernardini et al., \(2021\)](#).

Given the large number of experimental datasets presented, it is not within the intent of this study to discuss the transcriptional responses of organisms under the tested exposure conditions, which are extensively reviewed in the referenced manuscripts. Exception is made for the two independent studies of *St* exposed to glyphosate and Roundup ([Webster and Santos, 2015](#)) and *Xt* exposed to benzo(a)pyrene ([Regnault et al., 2014](#)). A comparison between the findings reported by the authors of the studies and the THI outputs would elucidate the applicability of the proposed methodology to other species.

[Webster and Santos \(2015\)](#) reported that most transcripts displayed alike expression trends with an overlap in gene ontologies and signaling pathways, suggesting similar risks and shared mechanisms of toxicity for glyphosate and Roundup. These results agree with the findings obtained here, where glyphosate and Roundup treatments share 59% and 53% of the significant Hallmark gene sets at the lowest and highest concentrations, respectively ([Supplementary information S3](#)). This would explain the similar THI values obtained across different treatments.

The significant Hallmark gene sets identified by GSEA, and reported in the SI, also reflect evidence described by the authors of the study. Gene sets contributing to the THI values were responsible for transcriptional changes of signaling processes (“TNF signaling via NFKB”), apoptosis (mainly in trout exposed to low glyphosate and low/medium Roundup concentrations), cell proliferation and disruption of several immune pathways. In trouts exposed to the highest glyphosate concentration, a depressed stress response was also present. For both chemicals, the highest THIs were obtained at the lowest concentrations, confirming the observation of Webster and Santos that fewer transcriptional changes were found at the highest treatment concentrations. In agreement with the conclusions reached in the study of brown trout, THI ranging from 156 (moderate hazard) to 284 (major hazard) expresses a considerable degree of similarity between treatments, as well as more adverse effects at low environmentally-relevant concentrations.

Concerning the second study, [Regnault et al. \(2014\)](#) investigated short-term transcriptomic changes occurring on clawed frogs at 6-, 12-, 18- and 24-h exposure to benzo(a)pyrene. The authors observed maximum transcriptional changes at 12 h post exposure, with the disruption of glucose and fatty acid metabolism, apoptosis regulation and tight junctions. These findings are well summarized by the highest THI (THI = 265) and class of hazard *major* calculated after 12 h exposure ([Table 2](#)). The increase in THI originates from the disruption of several Hallmark gene sets involved in the same biological processes described by the authors of the study as also reported in SI (Chapter S3).

It is evident that in the process of summarizing into a synthetic index the size and complexity of data generated via transcriptional analysis, functional genomics information might be lost. However, THI proved to be able to complement functional information by capturing the severity of the perturbation occurring in the whole transcriptome and by trying to convey it in a way that can be used in ERA by a wider range of users.

3.2. THI output: testing and validation

The performance of the novel methodology is addressed here by comparing the THI outputs with other quantitative summaries of transcriptomics metrics and with the judgment of a panel of external experts. To understand whether the THI outperformed alternative analysis results, the index values obtained from application to 31 experimental datasets (on *Rp*, *Mg* and *St* species) were statistically compared with *i*) the number of DEGs as presented in the reference studies, *ii*) the number of GSEA pathways considered as statistically significant, and *iii*) the average ES of all significant GSEA pathways. [Fig. 3](#) presents results of the







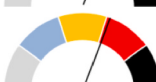
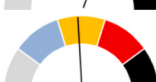

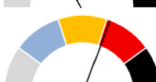


Table 2

Estimated THI as a numerical value and relative hazard class for the experimental datasets considered in this study. THI values are visualized in a hazard-meter plot where classes of hazard are converted to a common width. Colours in plots identify the five classes of hazard: grey/absent, light blue/slight, yellow/moderate, red/major and black/severe.

Chemical/Stressor	Species	Dose	Transcriptomics Hazard Index (THI): class and numerical value	Class of hazard	References
C6O4 (7 days)	<i>Rp</i>	0.1 µg/l	Moderate, 111		Bernardini et al., (2021)
C6O4 (7 days)	<i>Rp</i>	1 µg/l	Moderate, 80		
PFOA (7 days)	<i>Rp</i>	1 µg/l	Moderate, 89		
C6O4 (21 days)	<i>Rp</i>	0.1 µg/l	Moderate, 134		
C6O4 (21 days)	<i>Rp</i>	1 µg/l	Moderate, 71		
PFOA (21 days)	<i>Rp</i>	1 µg/l	Moderate, 130		
Acidification and Carbamazepine	<i>Mg</i>	pH 7.6, 1 µg/l	Slight, 63		Mezzelani et al., (2021)
Acidification	<i>Mg</i>	pH 7.6	Moderate, 144		
Carbamazepine	<i>Mg</i>	1 µg/l	Moderate, 194		
Glyphosate (7 days)	<i>Mg</i>	100 µg/l	Moderate, 90		Iori et al. (2020)
AMPA (7 days)	<i>Mg</i>	100 µg/l	Moderate, 71		
Glyphosate and AMPA (7 days)	<i>Mg</i>	100 µg/l	Moderate, 89		
Glyphosate (21 days)	<i>Mg</i>	100 µg/l	Slight, 49		
AMPA (21 days)	<i>Mg</i>	100 µg/l	Slight, 26		
Glyphosate and AMPA (21 days)	<i>Mg</i>	100 µg/l	Slight, 40		
Glyphosate (21 days)	<i>Mg</i>	10 µg/l	Moderate, 163		Milan et al., (2018)
Glyphosate (21 days)	<i>Mg</i>	100 µg/l	Moderate, 121		
Glyphosate (21 days)	<i>Mg</i>	1000 µg/l	Moderate, 103		
Amyl Salicylate (7 days)	<i>Mg</i>	0.3 µg/l	Moderate, 156		Bernardini et al., (2022)

(continued on next page)

Table 2 (continued)

Chemical/Stressor	Species	Dose	Transcriptomics Hazard Index (THI): class and numerical value	Class of hazard	References
Amyl Salicylate (7 days)	Mg	3 µg/l	Moderate, 144		
Amyl Salicylate (14 days)	Mg	0.3 µg/l	Slight, 40		
Amyl Salicylate (14 days)	Mg	3 µg/l	Moderate, 125		
Glyphosate (14 days)	St	10 µg/l	Major, 284		Webster and Santos (2015)
Glyphosate (14 days)	St	10000 µg/l	Major, 265		
Roundup (14 days)	St	10 µg/l	Major, 256		
Roundup (14 days)	St	500 µg/l	Major, 256		
Roundup (14 days)	St	10000 µg/l	Moderate, 156		
Benzo[a]pyrene (6 h)	Xt	10 µg/l	Slight, 49		Regnault et al. (2014)
Benzo[a]pyrene (12 h)	Xt	10 µg/l	Major, 265		
Benzo[a]pyrene (18 h)	Xt	10 µg/l	Slight, 40		
Benzo[a]pyrene (24 h)	Xt	10 µg/l	Slight 58		

correlation tests conducted by means of Spearman's rank correlation coefficient (S) with the function `ggscatterstats` of the R package `ggstatsplot`.

No significant correlation denoted the relationship either between THI and the number of DEGs (p -value = 0.75, $\rho_{\text{Spearman}} = 0.06$, Fig. 3a) or between the number of GSEA pathways and the number of DEGs (p -value = 0.61, $\rho_{\text{Spearman}} = 0.10$, Fig. 3b). The absence of correlation remarks one of the shortcomings of single-gene analysis: in some cases, few DEGs are found but the associated THI value is high. The use of DEGs creates a situation where, due to the cut-off threshold for defining DEGs, the output might miss many subtle (i.e. not "strong" enough to call the gene as differentially expressed) but coordinated changes in gene expression which still have a great biological relevance.

When compared to GSEA summative information, THI well correlates with the number of statistically significant pathways (p -value = 5.21×10^{-14} , $\rho_{\text{Spearman}} = 0.95$, Fig. 3c) but poorly with the average ES (p -value = 0.33, $\rho_{\text{Spearman}} = -0.19$, Fig. 3d). The first result (Fig. 3c) is in agreement with the conceptual framework at the basis of the THI derivation, where the number of statistically significant pathways contributes to define the class of effect of each category. It should, however, be highlighted that on top of the number of significant pathways, THI adds an extra layer of information, which consists of the relevance of the physiological reactions (i.e., included through the weighing) triggered by a change in pathways' expression.

On the other hand, the "simpler" metrics that considers the average ES disregards the severity of the transcriptomic alteration because, due to the averaging process, extreme responses (i.e. extreme ES values) are attenuated. Thereby, using the average ES instead of THI would lead to an underestimation of the hazard. Overall, the analysis proved that THI can summarize whole-transcriptome changes more accurately than single-gene analysis results, i.e. the number of DEGs, while performing better than simpler gene set analysis metrics, e.g. average ES.

For ten datasets selected from experimental and simulated profiles to cover the entire spectrum of hazard classes, the THI output was then compared with the judgment of a panel of eleven external experts (Supplementary information S5). By converting experts' classes into a nominal scale, ranging from 1 – for absent hazard, to 5 – in case of severe hazard – an averaged judgment accompanied by its variation in terms of standard deviation could be derived for each dataset. To allow comparison, the respective THI value was proportionally adjusted to follow the same scale. Fig. 4 shows the correlation between averaged expert evaluations and THI outputs for each data set.

Overall, THI proved to be a good indicator of the different levels of adverse transcriptomic response across the entire spectrum of hazard classes. The correlation between the two variables was indeed significant, with a Spearman correlation coefficient of 0.96 (p -value < 0.0001) and with data falling closely to the identity ($y = x$) especially for situations of slight and major hazards (Fig. 4). Again, the quantitative

Experimental datasets

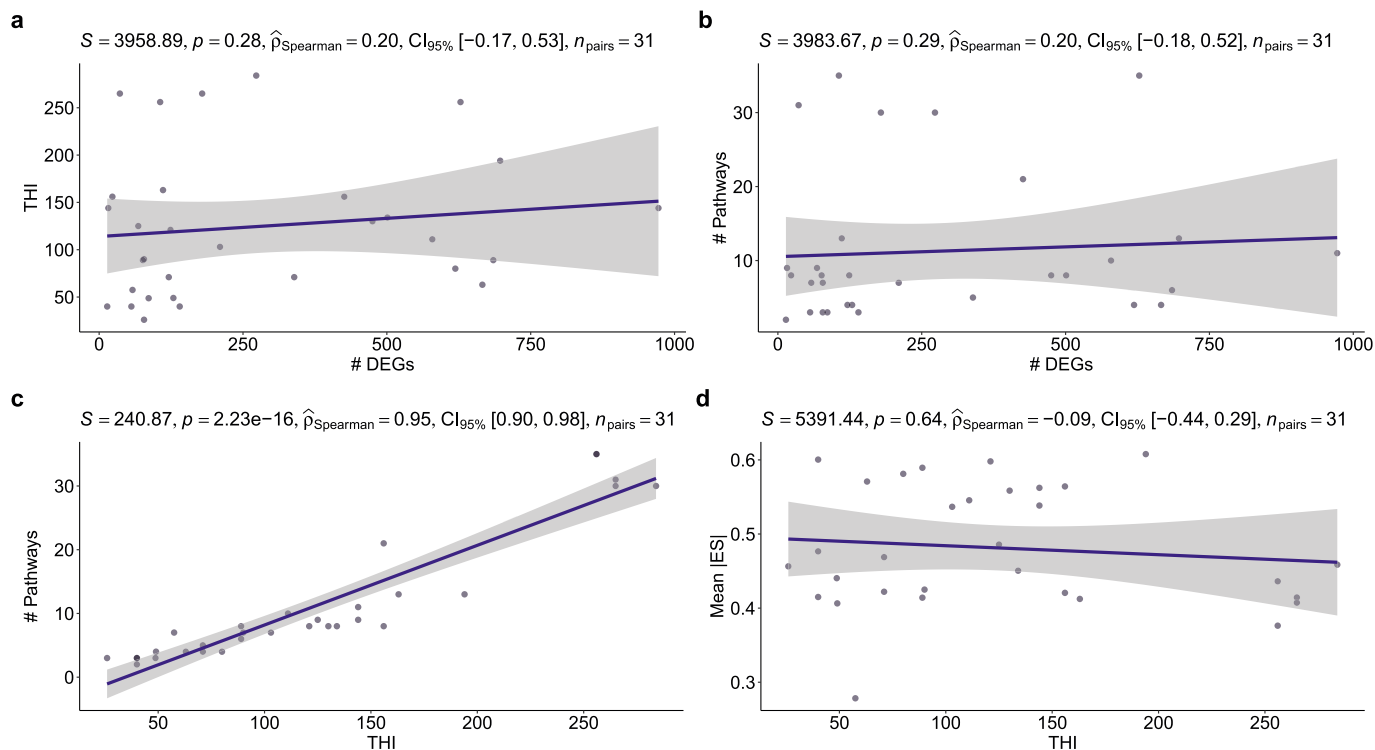


Fig. 3. Correlations between the THI index and different transcriptomic metrics from the experimental datasets presented in Table 2: a) THI and number of DEGs; b) number of significant GSEA pathways and number of DEGs; c) number of significant GSEA pathways and THI; d) mean ES and THI. Solid blue line represents the regression line with its confidence interval as shaded grey area. Correlation test was assessed by means of Spearman’s rank correlation coefficient (S). Additional statistical details are depicted in the subtitle of each plot: “p” denotes the significance of the correlation, the effect size and confidence interval are identified by “rho” and “CI95%”, while “n” indicates the number of observations. Quantities on axes are all nondimensional.

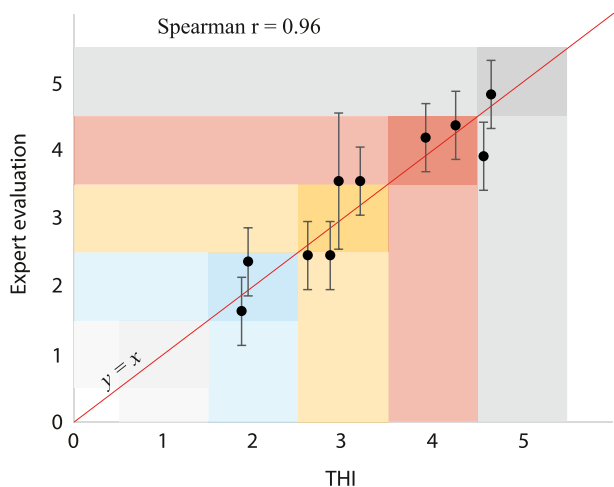


Fig. 4. Comparison of THI values with the averaged expert qualitative evaluation converted into nominal scale from 1 (absent hazard) to 5 (severe hazard). Error bars give an indication of experts’ variability in attributing a class of hazard. The red line represents complete correlation between the two variables, while the observed Spearman correlation coefficient is reported above the graph.

hazard index seems to work in conjunction with the qualitative class of hazard by complementing the indication of the transcriptomics hazard that would result in the simple use of a “label” (from absent to severe). For example, the THI value can give an indication of the degree of risk for borderline situations, where also experts’ evaluations were shown to hardly align. The next step towards a better interpretation of

transcriptomic hazard could provide the proposed methodology with a quantitative approach to estimate and communicate the uncertainty of the hazard output.

3.3. Challenges, benefits, and opportunities of THI in environmental risk assessment

The recent development of new genomic-enabled technologies has brought valuable insights in ecological, evolutionary, and environmental science, making affordable refined studies of the genome and transcriptome of non-model species (Van Aggelen et al., 2010). Among them, NGS (Next Generation Sequencing) technological and computational advances of which RNA sequencing (RNA-seq) is a major step, allowed to apply genome-wide gene expression analysis to investigate the effects of anthropogenic impact (e.g. toxicants) on organisms. Transcriptomics provides information from a different biological level by simultaneously characterizing hundreds to thousands of expressed genes (which will putatively translate to proteins/enzymes). It can be considered a “super-biomarker”. In fact, it conceives a much greater amount of information than that generated by the “classical” biomarkers, which can only focus on one or a handful of proteins/enzymes (van Straalen and Feder, 2012).

Despite its potential, transcriptomics is not exempt from challenges mainly related to data analysis and, consequently, to functional interpretations of genes expression. The global adoption of a single pipeline that covers all aspects of the RNAseq analysis (i.e. from sample’s QC to the generation of the table containing the *p-value* and fold change for each gene in the genome/transcriptome) would add a level of standardisation and comparability between datasets that are then used for the interpretation of the biological outcome.

To gain biological insights, different methods for gene expression

analyses have flourished in the past years, fostering a great variation in the results obtained depending on the choice of methodology, parameters and thresholds applied. Gene set analysis methods differ in their various components, from the underlying assumptions to the way they test for significance. The extensive review compiled by Maleki et al. (2020) well illustrates the strengths and weaknesses of the main gene set analysis methods (i.e., overrepresentation analysis, functional class scoring, and pathway topology-based methods) ascribing the problem of reproducibility, specificity and sensitivity of the results to the absence of a gold standard expression database to test the methods against. The result is a lack of consensus on the best practice to follow when analysing and interpreting transcriptional data. Under these circumstances, the use of the GSEA, as in Subramanian et al. (2005), is here proposed, keeping in mind both GSEA potential over single-gene analysis and its limitations. It is not excluded in the future the possibility of considering new methods denoted by higher specificity and sensitivity, or able to overcome GSEA shortcomings such as the influence of the gene set sizes in the results and the presence of gene sets overlapping due to multi-functional genes.

The latest limitation was here addressed with the adoption of the Hallmark gene sets collection which was further refined by merging redundant Hallmark gene sets. However, the application of Hallmark gene sets in non-model species, often lacking of a high-quality genome and/or annotated transcriptome compared to mammalian models, may provide partial insight into the real dysfunction of specific molecular pathways. The EcoToxModules recently proposed by Ewald et al. (2020) represents an interesting attempt to create a gene set collection shared by scientists and stakeholders and a second one unbiased by functional annotation. Whilst these gene sets were developed in fathead minnow, the same method can be applied to other species with a sequenced transcriptome and can be also considered in future to refine species specific lists of gene sets included in the definition of the THI.

Bearing in mind that a standardized protocol for data processing and gene expression analysis would improve the biological interpretation of generated data, THI has here demonstrated to perform well in summarizing large and complex data into a single index that can be used in ERA. With respect to single-gene analysis results or to simpler gene set analysis metrics, THI couples changes occurring at the whole transcriptome level with their relative biological importance. Compared to other approaches that try to summarize transcriptional data, this methodology does not focus only on few (or a handful of) genes known *a priori* to be associated with specific toxicity pathways, as in the case of Benchmark Dose approaches (BMD). On the contrary, THI considers the entire sets of genes in the transcriptome and their degree of change in comparison with a control condition. Focusing only on a particular group of genes in the attempt to elicit a threshold (Point of Departure, POD) for transcriptional responses might jeopardize the prediction of other adverse effects or key toxicological events, especially when transcriptional responses do not show predictable concentration-dependence. THI tries to overcome this limitation, but for its correct interpretation and use it is crucial to not lose sight of the biological meaning of transcriptomic data and to combine this information with other LoEs providing information about effect at higher biological level.

The possible application to both model and non-model species as long as there is an annotated genome or transcriptome available is another beneficial aspect of the THI. In this study, the proposed method has been applied to two bivalve and one freshwater fish species without the need to adjust and/or modify the bioinformatics and statistic pipelines. The methodology requires to link genes in the studied species with the Ensembl Human gene ids. Several ways are currently available to perform such task, like the Ensembl biomART tools (if the studied species is available on Ensembl), the KEGG's KOALA tools (e.g. BlastKOALA, GhostKOALA, KofamKOALA), the protein BLAST or OrthoFinder programs (by running them against the human proteome set) (Durinck et al., 2009; Kanehisa et al., 2016; Emms and Kelly, 2019). Further, as demonstrated here for *Mg* and *Rp* datasets, the approach is also flexible

in the sense that each user is free to eliminate some of the gene sets if not relevant to the studied organism, or if there is evidence suggesting that the function of genes in those species is different. This is a key aspect for a future extension of the proposed methodology to environmental risk studies of species placed also at different trophic levels. In addition, given the accessibility of transcriptomic data from published studies in public databases (e.g., SRA in NCBI), the method can be easily applied to previously collected data (similarly to what was done here with the brown trout and the western clawed frog studies), without the need to generate new data again. Researchers would then take advantage of a great deal of information that has already been generated over the past years in order to compare risks for different chemicals, concentrations and/or species.

With regard to the main objective of this work, transcriptomics can now be introduced in a QWoE approach applied to ERA contexts. The index can be normalized to the same scale of indexes from other bodies of evidence and weighed before being aggregated with other LoEs (Bates et al., 2018). The contribution of transcriptomics evidence to the final assessment should reflect the relevance and reliability of transcriptomics results. This depends on the collective properties of transcriptional data, such as resolution, study design and execution, or standardization of procedures, which the assessors should evaluate and distill into a weight. The evaluation requires assessors to manage with a certain degree of confidence transcriptomics, and this might be one of the factors that has refrained from quantitatively introduce transcriptional data in WoE approaches. Compared to widely validated LoEs, such as chemistry or ecotoxicological bioassays, transcriptomics lacks of quality standards and of a universally accepted standardized method for data processing. As a consequence, the exploitation of transcriptomics for multi-disciplinary risk assessment has remained narrow. With this study, where a methodology for data processing and results interpretation is provided, the hope is that transcriptomics will be part of QWoE so that hazard can be inferred at different levels of organism functioning.

3.4. Conclusive remarks

In the pursuit of quantitatively summarizing transcriptomic data to support ERA, a new hazard index was developed that reflects the level of genes expression alteration in aquatic organisms exposed to chemicals and the biological importance of such modifications. According to the THI, five classes of hazard (i.e. from absent to severe) were calibrated and validated using previously collected data on two bivalve species and simulated transcriptional profiles. The consistency between THI and the independent expert opinion across the entire spectrum of effects represents an excellent indication of the ability of the methodology to “capture” and synthesize transcriptomic changes. The flexibility of proposed functional database and algorithms facilitates transferability of the methodology to other species, as demonstrated here with the brown trout and clawed frog studies. Additional testing and refinement are surely desirable, especially when it comes to applicability to *in situ* scenarios characterized by multiple stressors and mixture of chemicals. At the moment, the methodology incorporates uncertainty in the processing of data but not in the reached conclusions. Towards a statistical QWoE that estimates the probability of impairment, the next step would provide THI with a method for calculating the hazard probability for the transcriptomics LoE. In the meantime, a transparent methodology that covers all aspects, from data processing to results interpretation, is proposed and validated across different aquatic species. Within the multidisciplinary vision adopted by ecological QWoE approaches, the proposed method constitutes an opening towards the construction of a supportive LoE based on transcriptional evidence, something that has been long sought after. This will strengthen and complement the interpretation of results from other analyses in support to a more accurate assessment of environmental risk in aquatic ecosystems.

Author contributions

Martina Cecchetto: conceptualization, methodology, formal analysis, writing original draft, review & editing; **Luca Peruzza:** conceptualization, methodology, formal analysis, writing original draft, review & editing; **Elisa Giubilato:** conceptualization, methodology, formal analysis, review & editing; **Iliaria Bernardini:** formal analysis, investigation; **Giulia Dalla Rovere:** formal analysis, investigation; **Antonio Marcomini, Tomaso Patarnello and Luca Bargelloni:** validation, writing review & editing, funding acquisition; **Francesco Regoli:** conceptualization, validation, writing review & editing; **Elena Semenzin:** conceptualization, review & editing, supervision, project administration, funding acquisition; **Massimo Milan:** conceptualization, writing original draft, review & editing, supervision, funding acquisition.

Funding

The scientific activity was performed in the Research Programme Venezia2021, coordinated by CORILA, with the contribution of the Provveditorato for the Public Works of Veneto, Trentino Alto Adige and Friuli Venezia Giulia.

Authors acknowledge the funding of the Italian Ministry of Education, University and Research (MIUR) for the project “Centro di Eccellenza per la Salute degli Animali Acquatici - ECCE AQUA”.

This work does not include results of new experiments with animals but makes use of already published experimental data from animal testing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors would like to thank Giuseppe D’Errico for providing useful comments to the development of the methodology. The authors are also thankful to Serena Ferraresso, Maria Elena Martino, Marco Gerdol, Chiara Manfrin, Daniela Bertotto, Giuseppe Radaelli, Marica Mezzelani, Stefania Gorbi, Alessandro Nardi, Maura Benedetti, Arnaud Huvet and Samuele Greco for their help in model validation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envres.2023.115745>.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Ankley, G.T., Daston, G.P., Degitz, S.J., Denslow, N.D., Hoke, R.A., Kennedy, S.W., Miracle, A.L., Perkins, E.J., Snape, J., Tillitt, D.E., Tyler, C.R., Versteeg, D., 2006. TOXICOGENOMICS in regulatory ecotoxicology. *Environ. Sci. Technol.* <https://doi.org/10.1021/es0630184>.
- Bajard, L., Adamovsky, O., Audouze, K., Baken, K., Barouki, R., Beltman, J.B., Beronius, A., Bonefeld-Jørgensen, E.C., Cano-Sancho, G., de Baat, M.L., Di Tillio, F., Fernandez, M.F., FitzGerald, R.E., Gundacker, C., Hernandez, A.F., Hilscherova, K., Karakitsios, S., Kuchovska, E., Long, M., Luijten, M., Majid, S., Marx-Stoelting, P., Mustieles, V., Negi, C.K., Sarigiannis, D., Scholz, S., Sovadinova, I., Stierum, R., Tanabe, S., Tolleson, K.E., van den Brand, A.D., Vogs, C., Wielsøe, M., Wittwehr, C., Blaha, L., 2023. Application of AOPs to Assist Regulatory Assessment of Chemical

- Risks – Case Studies, Needs and Recommendations. *Environmental Research.* <https://doi.org/10.1016/j.envres.2022.114650>.
- Barjhoux, I., Fechner, L.C., Lebrun, J.D., Anzil, A., Ayrault, S., Budzinski, H., Cachot, J., Charron, L., Chaumot, A., Clérandeau, C., Dedouge-Geffard, O., Faburé, J., François, A., Geffard, O., George, I., Labadie, P., Lévi, Y., Munoz, G., Noury, P., Oziol, L., Quéau, H., Servais, P., Uher, E., Urien, N., Geffard, A., 2018. Application of a multidisciplinary and integrative weight-of-evidence approach to a 1-year monitoring survey of the Seine River. *Environ. Sci. Pollut. Res.* <https://doi.org/10.1007/s11356-016-6993-6>.
- Bates, M.E., Massey, O.C., Wood, M.D., 2018. *Weight-of-Evidence Concepts: Introduction and Application to Sediment Management.* US Army Corps of Engineers.
- Bebianno, M.J., Pereira, C.G., Rey, F., Cravo, A., Duarte, D., D’Errico, G., Regoli, F., 2015. Integrated Approach to Assess Ecosystem Health in Harbor Areas. *Science of the Total Environment.* <https://doi.org/10.1016/j.scitotenv.2015.01.050>.
- Benedetti, M., Ciaprin, F., Piva, F., Onorati, F., Fattorini, D., Notti, A., Ausili, A., Regoli, F., 2012. A multidisciplinary weight-of-evidence approach for classifying polluted sediments: integrating sediment chemistry, bioavailability, biomarkers responses and bioassays. *Environ. Int.* <https://doi.org/10.1016/j.envint.2011.08.003>.
- Bernardini, I., Matozzo, V., Valsecchi, S., Peruzza, L., Dalla Rovere, G., Polesello, S., Iori, S., Marin, M.G., Fabrello, J., Ciscato, M., Masiero, L., Bonato, M., Sanvito, G., Boffo, L., Bargelloni, L., Milan, M., Patarnello, T., 2021. The new PFAS C6O4 and its effects on marine invertebrates: first evidence of transcriptional and microbiota changes in the Manila clam *Ruditapes philippinarum*. *Environ. Int.* <https://doi.org/10.1016/j.envint.2021.106484>.
- Bernardini, I., Fabrello, J., Vecchiato, M., Ferraresso, S., Babbucci, M., Peruzza, L., Dalla Rovere, G., Masiero, L., Marin, M.G., Bargelloni, L., Gambaro, A., Patarnello, T., Matozzo, V., Milan, M., 2022. Effects of environmental concentrations of the fragrance amyl salicylate on the Mediterranean mussel *Mytilus galloprovincialis*. *Environ. Pollut.* <https://doi.org/10.1016/j.envpol.2022.119502>.
- Borgwardt, F., Robinson, L., Trauner, D., Teixeira, H., Nogueira, J.A., Lillebø, A.I., Piet, G., Kuemmerlen, M., O’Higgins, T., McDonald, H., Arevalo-Torres, J., Barbosa, A.L., Iglesias-Campos, A., Hein, T., Culhane, F., 2019. Exploring Variability in Environmental Impact Risk from Human Activities across Aquatic Ecosystems. *Science of The Total Environment.* <https://doi.org/10.1016/j.scitotenv.2018.10.339>.
- Chapman, M.P., McDonald, B.G., Lawrence, G.S., 2002. Weight-of-Evidence issues and framework for sediment quality (and other) assessments. *Human. Ecol. Risk Assessment.* <https://doi.org/10.1080/20028091057457>.
- d’Errico, G., Nardi, A., Benedetti, M., Mezzelani, M., Fattorini, D., Di Carlo, M., Pittura, L., Giuliani, M.E., Macchia, S., Vitiello, V., Sartori, D., Scuderi, A., Morroni, L., Chiaretti, G., Gorbi, S., Pellegrini, D., Regoli, F., 2021. Application of a multidisciplinary weight of evidence approach as a tool for monitoring the ecological risk of dredging activities. *Front. Mar. Sci.* <https://doi.org/10.3389/fmars.2021.765256>.
- Dean, J.L., Zhao, Q.J., Lambert, J.C., Hawkins, B.S., Thomas, R.S., Wesselkamper, S.C., 2017. Application of gene set enrichment analysis for identification of chemically induced, biologically relevant transcriptomic networks and potential utilization in human health risk assessment. *Toxicol. Sci.* <https://doi.org/10.1093/toxsci/kfx021>.
- Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* <https://doi.org/10.1038/nprot.2009.97>.
- Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* <https://doi.org/10.1186/s13059-019-1832-y>.
- Ewald, J.D., Soufan, O., Crump, D., Hecker, M., Xia, J., Basu, N., 2020. EcoToxModules: custom gene sets to organize and analyze toxicogenomics data from ecological species. *Environ. Sci. Technol.* <https://doi.org/10.1021/acs.est.9b06607>.
- Fan, J., Semenzin, E., Meng, W., Giubilato, E., Zhang, Y., Critto, A., Zabeo, A., Zhou, Y., Ding, S., Wuan, J., He, M., Lin, C., 2015. Ecological status classification of the Taizi River Basin, China: a comparison of integrated risk assessment approaches. *Environ. Sci. Pollut. Res.* <https://doi.org/10.1007/s11356-015-4629-x>.
- Gottardo, S., Semenzin, E., Giove, S., Zabeo, A., Critto, A., de Zwart, D., Ginebreda, A., Marcomini, A., 2011. Integrated Risk Assessment for WFD Ecological Status Classification Applied to Llobregat River Basin (Spain). Part I—Fuzzy Approach to Aggregate Biological Indicators. *Science of The Total Environment.* <https://doi.org/10.1016/j.scitotenv.2011.07.052>.
- Iori, S., Rovere, G.D., Ezzat, L., Smits, M., Ferraresso, S.S., Babbucci, M., Marin, M.G., Masiero, L., Fabrello, J., Garro, E., Carraro, L., Cardazzo, B., Patarnello, T., Matozzo, V., Bargelloni, L., Milan, M., 2020. The effects of glyphosate and AMPA on the mediterranean mussel *Mytilus galloprovincialis* and its microbiota. *Environ. Res.* <https://doi.org/10.1016/j.envres.2019.108984>.
- Johnson, A.C., Sumpter, J.P., Depledge, M.H., 2021. The Weight-of-Evidence approach and the need for greater international acceptance of its use in tackling questions of chemical harm to the environment. *Environ. Toxicol. Chem.* <https://doi.org/10.1002/etc.5184>.
- Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2015.11.006>.
- Lehtonen, K.K., d’Errico, G., Korpinen, S., Regoli, F., Ahkola, H., Kinnunen, T., Lastumäki, A., 2019. Mussel caging and the weight of evidence approach in the assessment of chemical contamination in coastal waters of Finland (baltic sea). *Front. Mar. Sci.* <https://doi.org/10.3389/fmars.2019.00688>.
- Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The molecular signatures database hallmark gene set collection. *Cell Syst.* <https://doi.org/10.1016/j.cels.2015.12.004>.

- Linkov, I., Loney, D., Cormier, S., Satterstrom, F.K., Bridges, T., 2009. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2009.05.004>.
- Lucia, G., Giuliani, M.E., d'Errico, G., Booms, E., Benedetti, M., Di Carlo, M., Fattorini, D., Gorbi, S., Regoli, F., 2023. Toxicological effects of cigarette butts for marine organisms. *Environ. Int.* <https://doi.org/10.1016/j.envint.2023.107733>.
- Maleki, F., Owens, K., Hogan, D.J., Kusalik, A.J., 2020. Gene set analysis: challenges, opportunities, and future Research. *Front. Genet.* <https://doi.org/10.3389/fgene.2020.00654>.
- Mezzelani, M., Nardi, A., Bernardini, I., Milan, M., Peruzza, L., d'Errico, G., Fattorini, D., Gorbi, S., Patarnello, T., Regoli, F., 2021. Environmental pharmaceuticals and climate change: the case study of carbamazepine in *M. galloprovincialis* under ocean acidification scenario. *Environ. Int.*, 106269 <https://doi.org/10.1016/j.envint.2020.106269>.
- Micheletti, C., Gottardo, S., Critto, A., Chiarato, S., Marcomini, A., 2011. Environmental quality of transitional waters: the lagoon of Venice case study. *Environ. Int.* <https://doi.org/10.1016/j.envint.2010.06.009>.
- Milan, M., Pauletto, M., Boffo, L., Carrer, C., Sorrentino, F., Ferrari, G., Pavan, L., Patarnello, T., Bargelloni, L., 2015. Transcriptomic resources for environmental risk assessment: a case study in the Venice lagoon. *Environ. Pollut.* <https://doi.org/10.1016/j.envpol.2014.12.005>.
- Milan, M., Dalla Rovere, G., Smits, M., Ferrarasso, S., Pastore, P., Marin, M.G., Bogianni, S., Patarnello, T., Bargelloni, L., Matozzo, V., 2018. Ecotoxicological effects of the herbicide glyphosate in non-target aquatic species: transcriptional responses in the mussel *Mytilus galloprovincialis*. *Environ. Pollut.* <https://doi.org/10.1016/j.envpol.2018.02.049>.
- Pagé-Larivière, F., Crump, D., O'Brien, J.M., 2019. Transcriptomic points-of-departure from short-term exposure studies are protective of chronic effects for fish exposed to estrogenic chemicals. *Toxicol. Appl. Pharmacol.* <https://doi.org/10.1016/j.taap.2019>.
- Piva, F., Ciaprin, F., Onorati, F., Benedetti, M., Fattorini, D., Ausili, A., Regoli, F., 2011. Assessing sediment hazard through a weight of evidence approach with bioindicator organisms: a practical model to elaborate data from sediment chemistry, bioavailability, biomarkers and ecotoxicological bioassays. *Chemosphere.* <https://doi.org/10.1016/j.chemosphere.2010.12.064>.
- Regoli, F., d'Errico, G., Nardi, A., Mezzelani, M., Fattorini, D., Benedetti, M., Di Carlo, M., Pellegrini, D., Gorbi, S., 2019. Application of a weight of evidence approach for monitoring complex environmental scenarios: the case-study of off-shore platforms. *Front. Mar. Sci.* <https://doi.org/10.3389/fmars.2019.00377>.
- Regnault, C., Worms, I.A.M., Oger-Desfeux, C., MelodeLima, C., Veyrenc, S., Bayle, M., Combourieu, B., Bonin, A., Renaud, J., Raveton, M., Reynaud, S., 2014. Impaired liver function in *Xenopus tropicalis* exposed to benzo[a]pyrene: transcriptomic and metabolic evidence. *BMC Genom.* <https://doi.org/10.1186/1471-2164-15-666>.
- Risso, D., Ngai, J., Speed, T., Dudoit, S., 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt.2931>.
- Sauer, U.G., Deferme, L., Gribaldo, L., Hackermüller, J., Tralau, T., van Ravenzwaay, B., Yauk, C., Poole, A., Tong, W., Gant, T.W., 2017. The challenge of the application of omics technologies in chemicals risk assessment: background and outlook. *Regul. Toxicol. Pharmacol.* <https://doi.org/10.1016/j.yrtph.2017.09.020>.
- Schmitz, M., Deutschmann, B., Markt, N., Backhaus, T., Brack, W., Brauns, M., Brickmann, M., Seiler, T.B., Fink, P., Tang, S., Beitel, S., Doering, J.A., Hecker, M., Shao, Y., Schulze, T., Weitere, M., Wild, R., Velki, M., Hollert, H., 2022. Demonstration of an Aggregated Biomarker Response Approach to Assess the Impact of Point and Diffuse Contaminant Sources in Feral Fish in a Small River Case Study. *Science of the Total Environment.* <https://doi.org/10.1016/j.scitotenv.2021.150020>.
- Solimini, A.G., Ptacnik, R., Cardoso, A.C., 2009. Towards holistic assessment of the functioning of ecosystems under the Water Framework Directive. *Trends Anal. Chem.* <https://doi.org/10.1016/j.trac.2008.10.015>.
- Soufan, O., Ewald, J., Zhou, J., Hacariz, O., Boulanger, E., Alcaraz, A.J., Hickey, G., Maguire, S., Pain, G., Hogan, N., Hecker, M., Crump, D., Head, J., Basu, N., Xia, J., 2022. EcoToxExplorer: leveraging design thinking to develop standardized web-based transcriptomics analytics platform for diverse users. *Environ. Toxicol. Chem.* <https://doi.org/10.1002/etc.5251>.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA.* <https://doi.org/10.1073/pnas.0506580102>.
- Suter, G.W., Cormier, S., 2011. Why and how to combine evidence in environmental assessments: weighing evidence and building cases. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2010.12.029>.
- Suter, G.W., 2021. Environmental assessors should use state-of-practice weight-of-evidence processes. *Environ. Toxicol. Chem.* <https://doi.org/10.1002/etc.5213>.
- US Environmental Protection Agency, 2016. Weight of Evidence in Ecological Assessment. EPA/100/R-16/0010.
- van Straalen, N.M., Feder, M.E., 2012. Ecological and evolutionary functional genomics-how can it contribute to the risk assessment of chemicals? *Environ. Sci. Technol.* <https://doi.org/10.1021/es2034153>.
- Van Aggelen, G., Ankley, G.T., Baldwin, W.S., Bearden, D.W., Benson, W.H., Chipman, J.K., Collette, T.W., Craft, J.A., Denslow, N.D., Embry, M.R., Falciani, F., George, S.G., Helbing, C.C., Hoekstra, P.F., Iguchi, T., Kagami, Y., Katsiadaki, I., Kille, P., Liu, L., Lord, P.G., McIntyre, T., O'Neill, A., Osachoff, H., Perkins, E.J., Santos, E.M., Skirrow, R.C., Snape, J.R., Tyler, C.R., Versteeg, D., Viant, M.R., Volz, D.C., Williams, T.D., Yu, L., 2010. Integrating omic technologies into aquatic ecological risk assessment and environmental monitoring: hurdles, achievements, and future outlook. *Environ. Health Perspect.* <https://doi.org/10.1289/ehp.0900985>.
- Verheijen, M., Tong, W., Shi, L., Gant, T.W., Seligman, B., Caiment, F., 2020. Towards the development of an omics data analysis framework. *Regul. Toxicol. Pharmacol.* <https://doi.org/10.1016/j.yrtph.2020.104621>.
- Villeneuve, D.L., Garcia-Reyero, N., Escalon, B.L., Jensen, K.M., Cavallin, J.E., Makynen, E.A., Durhan, E.J., Kahl, M.D., Thomas, L.M., Perkins, E.J., Ankley, G.T., 2012. Ecotoxicogenomics to support ecological risk assessment: a case study with bisphenol A in fish. *Environ. Sci. Technol.* <https://doi.org/10.1021/es201150a>.
- Webster, U., Santos, B.M.C., 2015. Global transcriptomic profiling demonstrates induction of oxidative stress and of compensatory cellular stress responses in brown trout exposed to glyphosate and Roundup. *Genomics.* <https://doi.org/10.1186/s12864-015-1254-5>.
- Wernersson, A., Carere, M., Maggi, C., Tusil, P., Soldan, P., James, A., Sanchez, W., Dulio, V., Broeg, K., Reifferscheid, G., Buchinger, S., Maas, H., Van Der Grinten, E., O'Toole, S., Ausili, A., Manfra, L., Marziali, L., Polesello, S., Lacchetti, I., Mancini, L., Lilja, K., Linderoth, M., Lundeberg, T., Fjällborg, B., Porsbring, T., Larsson, D.G.J., Bengtsson-Palme, J., Förllin, L., Kienle, C., Kunz, P., Vermeirssen, E., Werner, I., Robinson, C.D., Lyons, B., Katsiadaki, I., Whalley, C., den Haan, K., Messiaen, M., Clayton, H., Lettieri, T., Negrão Carvalho, R., Gawlik, B.M., Hollert, H., Di Paolo, C., Brack, W., Kammann, U., Kase, K., 2015. The European technical report on aquatic effect-based monitoring tools under the water framework directive. *Environ. Sci. Eur.* <https://doi.org/10.1186/s12302-015-0039-4>.
- WFD 2000/60/EC. Water Framework Directive. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy.
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation.* <https://doi.org/10.1016/j.xinn.2021.100141>.