Leveraging Multi-Omics Technologies for Studying the Effects of Endocrine Disrupting Chemicals on Thyroid In Vitro Models

Leveraging Multi-Omics Technologies for Studying the Effects of Endocrine Disrupting Chemicals on Thyroid *In Vitro* Models

Marta Nazzari

Cover design and layout by Marta Nazzari

Printed by Gildeprint

ISBN: 978-94-6469-870-1

Leveraging Multi-Omics Technologies for Studying the Effects of Endocrine Disrupting Chemicals on Thyroid *In Vitro* Models

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof. Dr. Pamela Habibović
in accordance with the decision of the Board of Deans

to be defended in public
on Wednesday 10th April 2024 at 10:00 hours

by

Marta Nazzari
born on 2nd January 1994 in Venice, Italy

Het Gebruik van Multi-Omics Technologieën voor het Bestuderen van de Effecten van Hormoonverstorende Chemicaliën op *In Vitro* Cel Modellen van de Schildklier

Marta Nazzari

Supervisor

Dr. Florian Caiment

Co-supervisor

Dr. Twan van den Beucken

Assessment Committee

Prof. Dr. Maarten Honing, Professor of Analytics in Systems Imaging, Maastricht University

Prof. Dr. Ir. Juliette Legler, Professor of Toxicology, Utrecht University

Prof. Karine Audouze, full Professor in Bioinformatics, Université Paris Cité

Dr. Susan L. M. Steinbusch-Coort, Department of Bioinformatics, Maastricht University

The research presented in this thesis was conducted at GROW School for Oncology and Reproduction of Maastricht University.

Table of Contents

Summary of the Thesis

This thesis contains work performed in the context of SCREENED, a European Union's Horizon 2020 project part of the EURION cluster, aimed at developing new test and screening methods to identify endocrine disrupting chemicals (EDCs). SCREENED focused on the thyroid, an essential endocrine organ understudied within the field of toxicology. Other partners within the project developed a protocol for differentiating thyroid organoids from human and mouse embryonic stem cells (ESC) which were used as an *in vitro* model to test the response to EDCs. Some of these exposure experiments performed during the project are reported in this thesis.

For the bulk transcriptome analyses in Chapters 2 and 3, we employed a relatively new method able to assess simultaneously the mRNA and miRNA expression in a sample called "Combo-Seq". In **Chapter 2**, we evaluated the robustness of this method and compared it to conventional separate poly(A) and small RNA libraries. Since we observed some limitations and inaccuracies with the pipeline recommended by the kit manufacturer for the analysis of Combo-Seq data, we developed a new one, better adapted to this purpose. In the Chapter, we compared the two pipelines at several steps of data analysis, from data processing to differential expression analysis.

In **Chapter 3**, we exposed mouse ESC-derived thyroid follicles to five incremental, biologically relevant doses of four phthalates for 24 hours and analyzed the changes in the transcriptome via RNA-Sequencing. Gene Set Enrichment Analysis showed a common upregulation of genes involved in fatty acid metabolism by all four phthalates, as well as downregulation of genes involved in signaling by GTPases, tyrosine kinases and TGFB family members and extracellular matrix organization. In all treatments we observed an upregulation of *Ing5*, whose protein product is involved in histone acetylation. We then exposed a thyroid cell line for 5 days to a selection of two phthalates to identify any treatment effect on chromatin accessibility, but we did not observe any changes.

In **Chapter 4,** we analyzed the effects of 16 EDCs belonging to four different classes on mouse thyroid organoids by way of transcriptomics and proteomics. We observed dose-response curves for several genes and a few miRNAs and, for some, doses that could represent potential points of departure. We used transcriptomics data to perform differential expression analysis grouping EDCs by class to identify possible class effects and proteomics data to identify proteins differentially expressed after EDC treatment. We combined the two datasets to identify if changes in genes or miRNA expression could be predictive of the protein levels of target proteins and to derive Random Forest classification models to determine whether a sample belongs to any of the EDC classes analyzed.

In **Chapter 5,** we studied if different sex hormones contexts would affect the thyroid model response to EDCs. To this end, we exposed human ESC-derived thyroid follicles to two mixtures of estrogen, progesterone and dihydrotestosterone that would resemble the serum levels of human females and males of reproductive age for three days. Afterwards, we added BAP and PCB153 for 24 hours, and at the end of the exposure period, we performed 10x Genomics single cell RNA-Sequencing. PCB153 had very limited effects on the transcriptome. We observed synergies between BAP and the two hormonal mixes, especially the "male" cocktail, affecting ribosomal genes as well as genes involved in promoting inflammation, lipid transport and metabolism and oxidative phosphorylation. Additionally, the presence of the "female" mixture increased the expression of aryl hydrocarbon receptor targets.

Samenvatting van het Proefschrift

Dit proefschrift bevat werk dat is uitgevoerd in het kader van SCREENED, een Horizon 2020-project van de Europese Unie dat deel van het EURION-cluster uitmaakt, gericht op de ontwikkeling van nieuwe test- en screeningsmethoden om hormoonverstorende stoffen (in het Engels "endocrine disrupting chemicals" (EDCs) genaamd) te identificeren. SCREENED richtte zich op de schildklier, een essentieel endocrien orgaan dat binnen de toxicologie zeer beperkt bestudeerd is. Andere project partners hebben een protocol ontwikkeld voor de differentiatie van schildklierorganoïden uit menselijke en muizenembryonale stamcellen (ESC). Hiermee kan een *in vitro* cel model gemaakt worden voor het testen van de effecten van EDCs. De resultaten van de blootstellingsexperimenten die tijdens dit project werden uitgevoerd, worden in dit proefschrift gerapporteerd.

Voor de bulk transcriptoom analyses in Hoofdstukken 2 en 3 gebruikten we een relatief nieuwe methode die in staat is om tegelijkertijd de mRNA en miRNA expressie in een monster te bepalen, genaamd "Combo-Seq". In **Hoofdstuk 2** hebben we de robuustheid van deze methode geëvalueerd en vergeleken met conventionele afzonderlijke poly(A) en kleine RNA-bibliotheken. Voor de analyze van deze RNAseq data hebben wij een nieuwe en verbeterde pijplijn gemaakt, omdat de analyse pijplijn van de fabrikant een aantal beperkingen en onnauwkeurigheden vertoonde. In dit hoofdstuk vergeleken we de twee pijplijnen in verschillende stappen van de gegevensanalyse, van gegevensverwerking tot differentiële expressieanalyse.

In **Hoofdstuk 3** stelden we ESC-afgeleide schildklierfollikels van muizen bloot aan vijf oplopende, biologisch relevante doses van vier ftalaten gedurende 24 uur en analyseerden we de veranderingen in het transcriptoom met behulp van RNA-Sequencing. Gene Set Enrichment Analysis toonde een gemeenschappelijke inductie van genen betrokken bij vetzuurmetabolisme door alle vier ftalaten, evenals repressie van genen betrokken bij signalering door GTPases, tyrosinekinases en TGFB familieleden en extracellulaire matrixorganisatie. Bij alle blootstellingen zagen we een toename van *Ing5*, waarvan het eiwitproduct betrokken is bij histonacetylering. Vervolgens stelden we een schildkliercellijn gedurende 5 dagen bloot aan een selectie van twee ftalaten om het effect van de behandeling op de chromatinetoegankelijkheid vast te stellen, maar we namen geen veranderingen waar.

In **Hoofdstuk 4** analyseerden we de effecten van 16 EDCs uit vier verschillende klassen op organoïden van de schildklier bij muizen door middel van transcriptomics en proteomics. We observeerden dosis-respons curves voor verschillende genen en een paar miRNAs, en voor

sommige EDCs konden we een dosis vaststellen die als potentiële point of departure kon dienen. We gebruikten transcriptomics gegevens om een differentiële expressieanalyse uit te voeren waarbij EDCs per klasse werden gegroepeerd om mogelijke klasse-effecten te identificeren en proteomics gegevens om eiwitten te identificeren die differentieel tot expressie kwamen na een blootstelling met EDCs. We combineerden de twee datasets om vast te stellen of veranderingen in genen of miRNA expressie voorspellend zouden kunnen zijn voor de eiwitniveaus van doeleiwitten. Daarnaast hebben deze data gebruikt om Random Forest classificatiemodellen af te leiden die gebruikt kunnen worden om EDCs te classificeren.

In **Hoofdstuk 5** onderzochten we of verschillende omgevingscontexten van geslachtshormonen de respons van het schildkliermodel op EDCs zouden beïnvloeden. Hiertoe stelden we menselijke ESC-afgeleide schildklierfollikels gedurende drie dagen bloot aan twee mengsels van oestrogeen, progesteron en dihydrotestosteron die de serumniveaus van menselijke vrouwen en mannen in de reproductieve leeftijd simuleren. Vervolgens voegden we gedurende 24 uur BAP en PCB153 toe, en aan het eind van de blootstellingsperiode voerden we 10x Genomics single cell RNA-Sequencing uit. PCB153 had zeer beperkte effecten op het transcriptoom. We zagen synergieën tussen BAP en de twee hormoonmengsels, vooral de "mannelijke" cocktail, die invloed hadden op ribosomale genen en genen die betrokken zijn bij de bevordering van ontsteking, lipidentransport en -metabolisme en oxidatieve fosforylering. Bovendien verhoogde de aanwezigheid van het "vrouwelijke" mengsel de expressie van arylkoolwaterstofreceptordoelwitten.

Impact of the Thesis

Endocrine disrupting chemicals (EDCs) are a large group of compounds comprising many manmade substances that pollute the environment and can interfere with the normal functioning of the endocrine system, including the thyroid, and as such they constitute an important environmental concern. They are found in common everyday products (like food packaging, medical plastic material, wire and cable sheathing, paints and coatings, textiles, fuel), and daily exposure of the general population occurs via the diet, air, skin, and water. My PhD, as part of the SCREENED (SCREENing for Endocrine Disruptors) project, focused on the thyroid, an essential endocrine organ understudied within the field of toxicology. SCREENED aimed at answering the following questions: can we develop 3-dimensional *in vitro* models of thyroid for studying the effects of EDCs? And are these compounds having a deleterious effect directly on the thyroid?

To this end, our partners at the Université Libre De Bruxelles in Belgium developed a protocol for differentiating thyroid organoids from human and mouse embryonic stem cells which we used for our experiments together with thyroid cell lines. We focused on four EDCs classes named organophosphate flame retardants (OPFRs), phthalates, polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) and performed several screenings during the project using a relevant dose range that would be relevant to human daily exposure, testing the difference between static and organ-on-a-chip culture conditions, as well as evaluating the response over a short- or long-term exposure. We used several omics techniques, which allow to perform an exhaustive profiling of several biological molecules, to study how EDCs can affect the transcriptome, proteome and epigenetic status of cells.

To provide some examples, we observed that our *in vitro* thyroid models can respond to aryl hydrocarbon receptor (AHR) agonists such as PAHs or some PCBs with the induction of the cytochrome P450 (the main detoxifying enzymes of the human body) genes *CYP1A1* and *CYP1B1*, to our knowledge a phenomenon to date only described in a handful of thyroid cancer cell lines. We also observed that phthalates can induce fatty acid metabolism and downregulate the transduction of important signaling proteins and extracellular matrix organization. Combining gene, miRNA and protein expression data, we built a machine learning classification model that could help us identify if an unknown sample was exposed to one of the EDCs classes we studied.

We showcased how omics approaches can be used in toxicology experiments to elucidate the cell response to harmful chemicals and provided hypotheses to be further tested with targeted experiments. Together with our other SCREENED partners, we started developing a model and tested it, laying the basis for a future use of *in vitro* models for endocrine disruption

testing. We generated a wealth of omics data derived from thyroid models exposed to EDCs that we deposited (or will deposit) in repositories, making it publicly available. Compared to the greater focus that other organs, like liver and intestine, receive, the field of thyroid toxicology is relatively understudied, and we started to fill a gap that we believe can be of use to other researchers.

Our work has led us to publish two of the chapters included in this thesis and, at the date of writing, having one under review. Striving to make our research FAIR (Findable, Accessible, Interoperable and Reusable), we always opted for open access journals, making our findings discoverable to anyone who is interested, and shared the data and scripts used for the analyses.

Chapter 1

Introduction

1. The Endocrine System

In mammals, the endocrine system comprises several hormone-producing organs: the hypothalamus, the pituitary gland, the thyroid gland, the parathyroid glands, the adrenal glands, the beta-cells in the pancreas, the ovaries (in females) and the testes (in males). Hormones regulate body growth, development, metabolism and homeostasis, and reproduction (1). Hormonal regulation in endocrine organs is regulated by the hypothalamus, which releases hormones that target the pituitary gland. In turn, the pituitary releases other tropic hormones that induce hormone production and release in the thyroid, gonads, and adrenal glands. These connections are termed hypothalamic-pituitary-thyroid (HPT)/gonadal/adrenal axes (2).

There are three types of hormones, based on the way they are synthesized: the first type is protein and polypeptide hormones, which are constituted by an amino acid chain. They usually derive from longer polypeptide chains, named prohormones, which are then cleaved to their active form. In turn, the prohormones often derive from a preprohormone, which can be processed in different parts of the body into different types of prohormones. Peptide hormones are stored into intracellular granules, to prevent their degradation by proteases, and include, among others, the thyrotropin release hormone (TRH) and the thyroid stimulating hormone (TSH). The second type of hormones is the steroid hormones, derived from cholesterol. Steroid hormones are generally not stored intracellularly but synthesized and released upon cell stimulation by diffusion. Their release is limited by the rate-limiting of final enzyme in the synthesis pathway. The third type of hormones is amino acid derivatives, which include the thyroid hormones (THs) and catecholamines (derived from tyrosine), and indolamines (derived from tryptophan) .The third type of hormones is amino acid derivatives, which include the thyroid hormones (THs) and catecholamines (derived from tyrosine), and indolamines (derived from tryptophan) (2, 3).

1.1 The Thyroid and the Synthesis of the Thyroid Hormone

The thyroid is an endocrine gland positioned in the lower part of the anterior neck and is responsible for the production of the THs, whose receptors are expressed throughout the body (4). It plays a central role in metabolism regulation, homeostasis and growth (1). TSH is the main regulator of TH production and is secreted by the adenohypophysis. In turn, TSH production is regulated both by circulating TH levels and TRH, synthesized in the hypothalamus. In a feedback loop mechanism, TH negatively regulates the production of TRH. The system that comprises the thyroid, the adenohypophysis and the hypothalamus is collectively termed hypothalamic-pituitary-thyroid (HPT) axis (5).

Normal thyroid functioning relies on a negative-feedback loop that involves signaling between the hypothalamus, adenohypophysis, and thyroid, collectively referred to as hypothalamic-pituitary-thyroid (HPT) axis: the hypothalamus is responsible for the synthesis of the TRH and the hypothalamic axonal projections from the paraventricular nucleus release TRH in the portal capillaries, which is transported via the hypophysial portal system to the adenohypophysis. TRH binds to its receptor (TRHR1) and induces the synthesis and release of the TSH by the adenohypophysis (5). TSH is in turn released in the bloodstream, through which it reaches the thyroid, binding to the TSH receptors (TSHR) on the thyrocytes surface, finally inducing the synthesis of the TH.

Anatomically, the thyroid is composed of two symmetrical lobes connected by an isthmus. The main cell type of the thyroid is constituted by thyrocytes, which organize in small hollow spheres called follicles, composed of 20 to 40 cells. Thyrocytes are responsible for synthesizing the TH, whose precursor is stored in the center of the follicle (the lumen) in a dense matrix termed colloid. The other cell type in the thyroid are the C-cells, which make up for around 0.1% of the thyroid cell population and are dispersed among the thyrocytes (6). They synthesize calcitonin, an hormone necessary for calcium regulation (7). The synthesis of the TH starts with the active transport of iodine inside the thyrocyte via the sodium iodide symporter (NIS in human) from the basal side and moves via passive diffusion to the apical side. In the follicle lumen, it is covalently bound via oxidation to the tyrosyl (Tyr) residues of the protein thyroglobulin (TG) via the action of the membrane-bound enzyme thyroid peroxidase (TPO) (6, 8). Hydrogen peroxide (H_2O_2) is necessary for this oxidative reaction to occur and is produced by the transmembrane glycoproteins dual oxidase 1 (DUOX1) and 2 (DUOX2) (9). Proper maturation and localization of the DUOX proteins requires the interaction with auxiliary proteins called dual oxidase maturation factor 1 and 2 (DUOXA1, DUOXA2), which interact with DUOX1 and DUOX2, respectively (10). One or two iodine atoms can be bound to a Tyr residue, generating 3-iodotyrosine (MIT) and 3,5-diiodotyrosine (DIT) (9). TPO also catalyzes the subsequent phenolic coupling of these two iodotyrosyls, mostly generating the precursors of the TH 3,5,3'-triiodothyronine (T3) and 3,5,3',5'-tetraiodothyronine (T4), still covalently bound to TG. Coupling of one MIT and one DIT generates a T3 molecule, while coupling of two DIT molecules produces T4 (6, 11). Other iodothyronines can be produced, namely 3,5-diiodothyronine (T2) and 3,3',5'-triiodothyronine (reverse T3 or rT3), but they account for a very small fraction of the total TH production (6). Following TSH stimulation, TG is phagocyted from the colloid into the cytoplasm and degraded in the lysosomes, freeing TH, that can be transported outside the thyrocyte mainly by the SLC16A2 monocarboxylate transporter 8 (MCT8) (8).

1.2 Distribution, Metabolism and Excretion of the Thyroid Hormone

1.2.1 Distribution to the Target Organs and Target Genes Expression

The thyroid mostly produces T4 (around 80% T4 and 20% T3 (12)), reflected in the very different total plasma concentration of the two types of TH (1.8 nM for T3, 100 nM for T4). After synthesis, the TH is released in the bloodstream, where most of it is bound by, in order of affinity, thyroxine hormone-binding globulin (TBG), transthyretin (TTG) and serum albumin to prevent degradation. Thus, the serum concentration of free TH is very low (5 pM for T3 and 20 pM for T4) (13). In the target organs, the TH binds to the thyroid hormone receptor (THR), which in its activated for can function as a monomer, a homodimer but preferentially as heterodimer with the retinoid X receptor (RXR). The THR is coded by two different genes: *THRA*, coding for TRa and *THRB*, coding for TRb. Several isoforms of the THR exist, but the functional ones known to date are TRa1, TRb1 and TRb2 (14). The activated THR can (mostly) induce but also repress gene expression (14) and the target genes vary depending on the tissue and developmental stage (15).

1.2.2 Metabolism and Excretion

Once T4 reaches the target organs, it is converted to T3 via removal of one iodine atom by deiodinases (DIO). In human, type 1 deiodinases (DIO1) are located on the plasma membrane and is expressed in the liver, kidney, thyroid and hypophysis, while type 2 deiodinases (DIO2) are located on the endoplasmic reticulum membrane and have a more widespread (but variable in its level) expression in thyroid, heart, brain, spinal cord, skeletal muscle, placenta, skin, retina, cochlea, kidney, brown adipose tissue and pancreas (5, 16-19). DIO1 can deiodinate both the outer and inner ring of T4, converting it to T3 and rT3 (an inactive form of the TH), respectively (20) however, DIO1 preferred substrate is rT3, sulfated T3 (T3S) and T4 (T4S) (21). DIO2 preferred substrate is T4 and can only perform outer ring deiodination, generating T3 (20). An additional type 3 deiodinase (DIO3) exists, which removes an iodine atom from the inner T4 and T3 ring, generating rT3 and T2 (another inactive form), respectively. DIO3 is important during embryonic development to prevent exposure to too high TH levels, and is primarily expressed in the fetus, placenta, and pregnant uterus. In the adult, it is only expressed in the central nervous system (22) and skin (19, 23). In addition to deiodination, the TH can be inactivated by sulfonation, which facilitates deiodination by DIO1, and increases its solubility, facilitating the excretion via the bile and urine. The TH can also be the substrate for glucuronidation in the liver, kidney, and intestine. This process facilitates the excretion via the

bile and feces. Overall, however, the sulfonation and glucuronidation seem to mostly have to role of creating a pool of inactive TH reservoir, as these reactions are reversible and only a small proportion of the TH is actually excreted (20).

2. Endocrine Disrupting Chemicals

The term "endocrine disrupting chemical" (EDC) refers to a broad class of environmental pollutants that can negatively interfere with the endocrine system and as consequence cause adverse effects in an intact organism, its progeny, populations or subpopulations. They are found in many everyday products, and daily exposure for the general population occurs via the diet, air, skin, and water (24).

In Figure 1, we report the 10 key characteristics (KCs) recently published by the US National Institute of Environmental Health Sciences (NIEHS) that can be used when evaluating chemicals as EDCs (25), and that can we therefore refer to for defining which the characteristics of EDCs are.

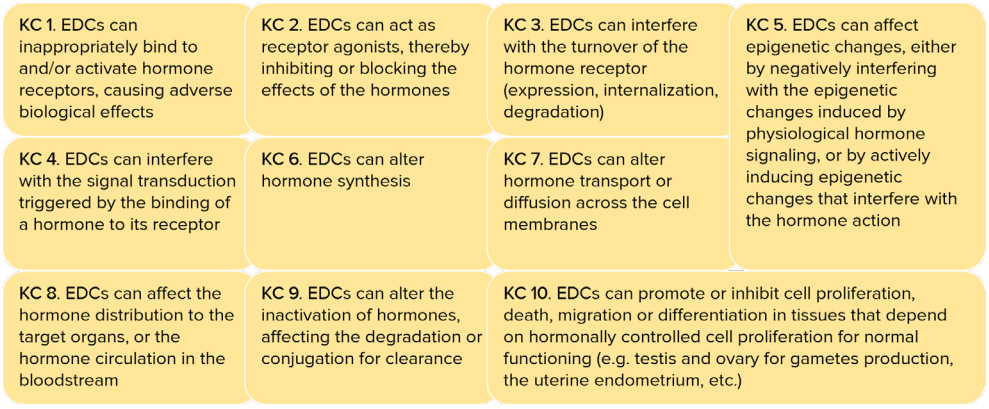


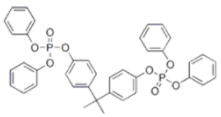
Figure 1. Key characteristics (KC) of endocrine disrupting chemicals as defined by the US National Institute of Environmental Health Sciences (NIEHS).

In the work presented in this thesis, we focused on four EDC classes: organophosphate flame retardants (OPFRs), phthalates, polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) (Table 1). They have been recognized as able to interfere with the hormones systems including the TH system, either directly affecting the thyroid, or indirectly affecting any of the steps involved in TH action after its release in the bloodstream.

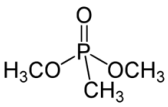
Table 1. Endocrine disrupting chemicals studied in this thesis grouped by class.

Organophosphate flame retardants (OPFRs)

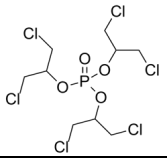
BADP
Bisphenol A bis(diphenyl phosphate)



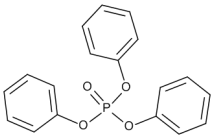
DMMP
Dimethyl methylphosphonate



TDCPP
Tris(1,3-dichloro-2-propyl) phosphate

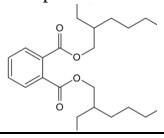


TPP
Triphenyl phosphate

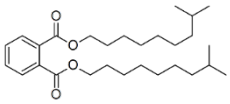


Phthalates

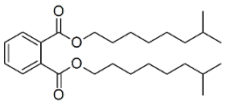
DEHP
Bis(2-ethylhexyl) phthalate



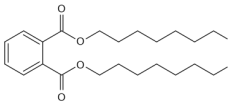
DIDP
Di-*iso*-decyl phthalate



DINP
Di-*iso*-nonyl phthalate

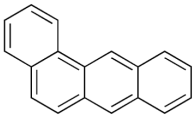


DnOP
Di-*n*-octyl phthalate

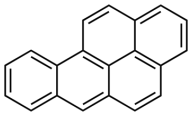


Polycyclic aromatic hydrocarbons (PAHs)

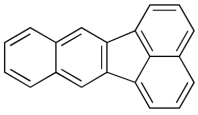
BAA
Benz[a]anthracene



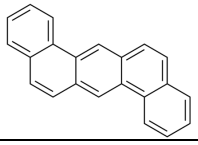
BAP
Benzo[a]pyrene



BKF
Benzo[k]fluoranthene

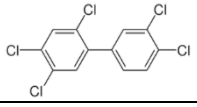


DAHA
Dibenzo[a,h]anthracene

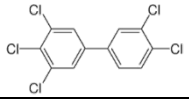


Polychlorinated biphenyls (PCBs)

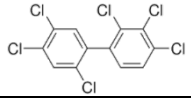
PCB118



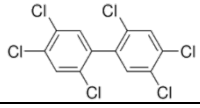
PCB126



PCB138



PCB153



2.1 Organophosphate Flame Retardants (OPFRs)

Organophosphate Flame Retardants (OPFRs) include a wide range of halogenated and non-halogenated compounds containing phosphorus. The most common OPFRs include three classes: phosphate esters, phosphonates and phosphinates. They are found in engineering plastics, coatings, polyurethane foams and textiles (26) and there is evidence of their association with thyroid disruption. Tris(1-chloro-2-propyl) phosphate (TCPP) has been shown to increase *Dio1* expression and tris(1,3-dichloro-2-propyl) phosphate (TDCPP) to negatively affect serum T4

levels in the developing chicken embryo (27). In human, a negative association has been found between TDCPP concentration in house dust and serum T4 levels in adult men (28). *In vitro*, TPP upregulated *tshβ*, *trα*, and *trβ* in the TH-responsive GH3 cell line (a hypophysis cell line) showing how TPP can have opposite effects of T3 treatment. In FRTL-5 cells (a thyroid-derived cell line), TPP treatment increased the expression of genes related to TH synthesis (29). In zebrafish, TDCPP and TPP have been shown to alter TH serum levels and affect key regulatory genes in the hypothalamic-pituitary-thyroid (HPT) axis in a sex-specific manner (30).

2.2 Phthalates

Phthalates are a class of manmade compounds used in the manufacturing industry as solvents or added as plasticizers, mainly to polyvinyl chloride (PVC) or other polymers, to confer flexibility and softness (31, 32). Phthalates are alkyl or dialkyl esters of phthalic acid and their functional groups can be linear, branched, or circular (33). Depending on their size, phthalates are classified into low and high molecular weight (MW) (34, 35). Low MW phthalates include benzyl butyl phthalate (BBP), diethyl phthalate (DEP), di-*iso*-butyl phthalate (DiBP), dimethyl phthalate (DMP) and di-*n*-butyl phthalate (DnBP), while high MW ones comprise bis(2-ethylhexyl) phthalate (DEHP), di(2-propylheptyl) phthalate (DPHP), di-*iso*-decylphthalate (DIDP), di-*iso*-nonylphthalate (DINP), and di-*n*-octylphthalate (DnOP) (36). They are found in common household items, medical devices, construction material and consumer products (36). As they are not covalently bound to the matrix they are added into, they can leave the material by direct release, evaporation leaching or abrasion (36). As a result, human exposure to phthalates can occur via ingestion, inhalation or dermal absorption (36), with ingestion via foodstuff being the most prominent via of exposure in the general population (37). Human biomonitoring studies conducted on the general population in Asia, Europe and North America show a widespread exposure of the general population to phthalates (36, 38-41). Once ingested, they are rapidly metabolized in the digestive tract to their monoester form, which are the species responsible for the phthalates' toxicity. Low MW phthalates metabolites are then excreted through the urine, while high MW metabolites are excreted both via the urine and feces (31). While they do not bioaccumulate, the persistent exposure of the population is cause for concern. Short and medium chain phthalates have been associated with higher toxicity than long chain ones which has led to their banning or restriction in children's toys or teething products (42, 43). Some phthalates have been associated with negatively impacting male fertility and altered development of the male reproductive tract (36), as well as female fertility, impacting folliculogenesis and steroidogenesis (44). They have been shown to interfere with prenatal and postnatal development in animal models (45). In the thyroid, phthalate treatment has been shown to have an effect *in vitro* and *in vivo*, causing histological changes, such as reduced follicle

size and colloid density, hypertrophy of the Golgi apparatus, increase in number and size of lysosomes and alteration of the TH levels (32, 46-49). *In vivo* and *in vitro* studies show how DEHP, one of the most diffused phthalates, can interfere with Tsh/TshR signaling and alter thyrocyte morphology (50, 51), and can compete with T3 for TR binding (49). In human studies, altered serum levels of T3 and T4 have been correlated with plasticizers metabolites in urine (47, 52).

2.3 Polycyclic Aromatic Hydrocarbons (PAHs)

Polycyclic Aromatic Hydrocarbons (PAHs) comprise a large group of organic compounds composed of two or more benzene rings and containing only carbon and hydrogen. They are poorly soluble in water but highly lipophilic and soluble in organic solvents (53). They are formed during the incomplete combustion of organic material, including materials employed in energy production at the industrial or household level, and tobacco smoke. The greatest source of exposure in the general population is thought to be contaminated or burnt food (54). Due to their poor water solubility, PAHs bind to soil particles or occur as a separate phase (55). The US Environmental Protection Agency (EPA) has classified seven PAHs (benz[a]anthracene (BAA), benzo[a]pyrene (BAP), benzo[b]fluoranthene, benzo[k]fluoranthene (BKF), chrysene, dibenz[a,h]anthracene (DAHA), and indeno[1,2,3-cd]pyrene) as Group B2, "probable human carcinogens" (56). In THR-transfected HepG2 cells, 1-naphthol and 2-naphthol showed antagonistic activity against the THR (57). *In vitro*, PAHs derivatives (hydroxides, ketones and quinones) can act as aryl hydrocarbon receptor (AHR) agonists and antagonists, have THR-potentiation activity and TTR-binding activity (58). AHR activation leads to the induction of cytochrome P450 enzymes, which can metabolize PAHs into toxic compounds with carcinogenic activity (59). Treatment with the AhR inducer b-naphthoflavone has been shown to lead to a decrease in plasma levels of T4 in fish (60) and both T4 and T3 in rat (61), and the authors suggest this is probably due to the induction of CYP1A and other hepatic microsomal enzymes, with consequent increased hepatic clearance by glucuronidation of TH. In an *in vitro* human recombinant TPO reporter assay, Song and colleagues showed that BKF and DAHA are able to, respectively, disrupt and induce TPO activity (62).

2.4 Polychlorinated Biphenyls (PCBs)

Polychlorinated Biphenyls (PCBs) are organic synthetic compounds that were produced and used until the 1970s as coolants and lubricants in many types of electrical equipment, both in industrial and consumer contexts. Contamination of the environment can

occur via burning, leaking, spilling, or breaking down of the material they are contained in. Due to their high chemical stability, they break down slowly and remain in the environment for a very long time, and due to their volatility, they can be carried via the atmosphere also very far away from where they were released. In consequence, PCBs are found all over the world. In human, PCB contamination mainly occurs via contaminated food and air. When ingested, they bioaccumulate due to their lipophilicity (63).

The base structure of PCBs is two phenyl rings connected by a carbon-carbon bond. Hundreds of PCBs exists which differ for the number and position of chlorine atom substitutions. Congeners with no chlorine substitutions in the *ortho* position (Figure 2) are termed coplanar PCBs and include number 77, 81, 126, and 169. They are also termed "dioxin like" because of their resemblance in structure to dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin or 2,3,7,8-TCDD), and are potent AHR agonists. As such, they are recognized as the most toxic PCBs. Congeners with one or more *ortho* substitution are not planar due to the steric hindrance of the bulk Cl atoms and have lower toxicity but can nonetheless have deleterious effects for example on the TH system.

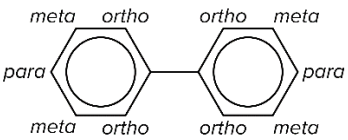


Figure 2. Basic structure of a polychlorinated biphenyl reporting the nomenclature of the functional groups position.

In fact, compounded evidence shows that PCBs can interfere with the production and transformation of TH in the thyroid and peripheral tissues, with the transport of TH in the blood, and increase the TH clearance in peripheral tissues (64). For example, PCBs can affect TH action by increasing T4 hepatic clearance (65, 66) and inhibiting T3 binding to the THR or by causing the dissociation of the THR to the thyroid response element (TRE). Hydroxylated PCBs, which derive from PCB hepatic metabolism, can compete with the TH for TTR binding in the blood (67). Hepatic UDP-glucuronosyltransferase (UDP-GT) was induced in a dose-dependent way in rats fed with the PCB mixture Aroclor 1254, also showing a decrease in serum T4, an increase in rT3 and thyroid hypertrophy and hyperactivity. No alterations in *Tsh* or serum T3 were observed (68). Similar results were obtained by Hood and Klaassen, who showed that Aroclor 1254 is able to decrease serum T4 but not T3 or *Tsh* levels in rat due to an increase in activity in the hepatic UDP-GT toward T4 but not T3 (69).

2.5 Regulatory Testing for the Disruption of the Thyroid System

In 2018 the Organization for Economic Co-operation and Development (OECD) published test guidelines (TG) for screening and testing of chemicals for potential endocrine disrupting activity (70). Currently, only *in vivo* tests are approved for evaluating the effects on the thyroid activity. Three others are reported but have only been approved by the US Environmental Protection Agency (EPA) (Table 2).

Table 2. Tests approved by the Organization for Economic Co-operation and Development (OECD) for endocrine disruption testing that include endpoints to evaluate thyroid disruption.

Test name	Status	Mandatory endpoints for thyroid-related activity	Optional endpoints for thyroid-related activity
OECD TG 231 (Amphibian Metamorphosis Assay)	Approved by OECD	Thyroid development. Histopathologic changes in thyroid gland.	
OECD TG 241 (Larval Amphibian Growth and Development Assay (LAGDA)	Approved by OECD	Changes in: - thyroid histopathology - time to metamorphosis	
OECD TG 407 (Repeated Dose 28-Day Oral Toxicity Study in Rodents)	Approved by OECD	Histopathologic changes in thyroid (follicular cell height increase and colloid area decrease).	Serum T3 and T4 decreased, TSH increased. Thyroid weight.
OECD TG 408 (Repeated Dose 90-Day Oral Toxicity Study in Rodents)	Approved by OECD	Serum T4, T3 decreased, TSH increased. Histopathologic changes in thyroid gland. Thyroid weight.	
OECD TG 409 (Repeated Dose 90-Day Oral Toxicity Study in Non-rodents)	Approved by OECD	Increased thyroid weight. Histopathologic changes in thyroid gland.	
OECD TG 411 (Subchronic Dermal Toxicity: 90-Day Study)	Approved by OECD	Histopathologic changes in thyroid gland.	

Test name	Status	Mandatory endpoints for thyroid-related activity	Optional endpoints for thyroid-related activity
OECD TG 412 (28-Day (Subacute) Inhalation Toxicity Study)	Approved by OECD	Increased thyroid weight. Histopathologic changes in thyroid gland.	
OECD TG 413 (Subchronic Inhalation Toxicity: 90-Day Study)	Approved by OECD	Increased thyroid weight. Histopathologic changes in thyroid gland.	
OECD TG 414 (Prenatal Developmental Toxicity Study)	Approved by OECD	Increased thyroid weight. Histopathologic changes in thyroid gland. Serum T4, decreased, TSH increased.	
OECD TG 416 (Two-Generation Reproduction Toxicity Study)	Approved by OECD	Increased thyroid weight. Histopathologic changes in thyroid gland.	
OECD TG 421 (Reproduction/Developmental Toxicity Screening Test)	Approved by OECD	Increased thyroid weight ("when necessary"). Histopathologic changes in thyroid gland. Serum T4 decreased, TSH increased ("if relevant").	
OECD TG 422 (Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test)	Approved by OECD	Increased thyroid weight ("when necessary"). Histopathologic changes in thyroid gland. Serum T4 decreased, TSH increased ("if relevant").	
OECD TG 441 (Hershberger Bioassay – Adult Male after Castration)	Approved by OECD		Reduction in serum T4 and T3

Test name	Status	Mandatory endpoints for thyroid-related activity	Optional endpoints for thyroid-related activity
OECD TG 443 (Extended One-Generation Reproductive Toxicity Study)	Approved by OECD	Increased thyroid weight. Histopathologic changes in thyroid gland.	
OECD TG 451-3 (Combined Chronic Toxicity/Carcinogenicity Studies)	Approved by OECD	Increased thyroid weight. Histopathologic changes in thyroid gland.	
Xenopus Embryonic Thyroid Signaling Assay (XETA)	Under validation	THbZIP-GFP construct (TH-sensitive promoter). Fluorescence is measured as effect of TH activation	
US EPA OPPTS 890.1450 (Pubertal Development and Thyroid Function Assay in Peripubertal Female Rats)	Approved by EPA	Increased thyroid weight. Histopathologic changes in thyroid (follicular cell weight increase and colloid area decrease). Serum T4 decreased, TSH increased.	
US EPA OPPTS 890.1500 (Pubertal Development and Thyroid Function Assay in Peripubertal Male Rats)	Approved by EPA	Increased thyroid weight. Histopathologic changes in thyroid (follicular cell weight increase and colloid area decrease). Serum T4 decreased, TSH increased.	
US EPA OCSPP 890.2100/740-C-15-003 ATGT (Avian Two-Generation Toxicity Test in the Japanese Quail)	Approved by EPA	T3, T4	

GFP: Green Fluorescent Protein; THR: Thyroid Hormone Receptor

These tests use mammalian and non-mammalian model organisms at various stages of development to study various endpoints of thyroid disruption, such as T3, T4 and TSH serum levels, direct effects on thyroid histology as well as alterations in development and maturation.

In vitro tests have been described but as of 2018 no guidance has been written since they are still at the research stage, and none have yet been validated and standardized at the international level. Moreover, they can test thyroid agonists and antagonists, but thyroid disruption can occur at other points in the thyroid system that these tests cannot evaluate.

Despite the number of testing options, the OECD itself states in the document several limitations presented by these assays. For example, when toxicity is observed, it may not be due to endocrine disruption. As such, possibly confounding effects of systemic toxicity on endocrine endpoints need to be considered. Sometimes, positive (i.e. toxic) effects may not affect development and reproduction further in life.

On the opposite end, absence of acute toxicity or effects does not mean absence of endocrine disrupting activity. There could be confounding factors, for example the test duration was not long enough to observe effects since the compound needs to bioaccumulate to a certain level before displaying toxicity, or it is not an EDC in amphibians, or the life stage used for testing is not relevant. In other cases, some indirect thyroid effects (e.g. CYP450 induction) are difficult to interpret in the context of endocrine disruption.

A retrospective review of 124 reproductive screening studies on mice performed using some of the OECD-approved tests mentioned above (OECD TG 408, 414, 421, 422, 443) concluded that including TH measurements does not provide specific information needed to assess endocrine disruption, as TH alterations are recorded but conclusions on the underlying causes cannot be drawn. Instead of adding additional endpoints to the *in vivo* studies, the authors recommend investing on the development and validation of *in vitro* assays, as they can be more functional in elucidating the mode of action in humans (71).

Since the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) Regulation was established by the EU in 2006 to manage the risk of chemical substances with an import or production of 1 ton/year or more, companies have been required to perform and report the results of toxicity evaluations. As of December 2022, the combined amount of laboratory animals used or planned to be used for a subset of test categories (reproductive toxicity tests, developmental toxicity tests, and repeated-dose toxicity tests for human health) has been estimated to be 4.2 million, not including the pups or the progeny in the case of multi-generational studies. Questions have been raised about the advantages posed by the use of *in vivo* testing beside the fulfilment of legal requirements and whether equivalent information could be gained by new approach methodologies (NAMs) (72, 73).

3. Omics Technologies in Regulatory Risk Assessment

One of the challenges related to the use of omics technologies in a regulatory setting lies in the desired outcome of toxicological *in vivo* testing, that is a positive or negative apical endpoint, namely an observable pathological alteration at the cell or organ level. Often, to define a substance as hazardous, it is not necessary to uncover the molecular mechanism underlying the alteration. In this context, the added value of omics technologies lies in the ability to predict, rather than observe, a pathological alteration, for example in the identification of a Molecular Initiating Event (MIE) and Key Event Relationships (KERs) in the context of Adverse Outcome Pathway (AOP), or identifying the point of departure (POD), which is a dose at which a biological response that can be predictive of a disease state rises above background levels (74). MIE refers to the direct interaction of a chemical with a biological target, which causes a succession of measurable Key Events (KEs) connected by causal KERs. Eventually, the sequence of KEs leads to an Adverse Outcome (AO) at the organism or population level. The whole succession of events, from MIE to AO, is termed AOP (75). Additionally, they could provide evidence for a successful read-across (an approach that uses relevant information from analogous chemicals to predict the properties of ones for which data is lacking (76)), which has proven unsuccessful with current *in vitro* testing methodologies, with 75% of studies being rejected due to of lack of compliance and causing companies being asked to perform standard tests (77).

Currently, omics technologies have a limited role in formal regulatory testing. The OECD approved in June 2022 the first *in vitro* test for skin sensitization (OECD TG 442E) that addresses the Key Event "Activation of dendritic cells" belonging to the AOP 40 "Covalent Protein binding leading to Skin Sensitization" (Figure 3). These guidelines collect four tests, including an assay that evaluates the expression of 196 genes associated with the process of activation of monocytes and dendritic cells following exposure to sensitizers ("Genomic Allergen Rapid Detection (GARD™) for assessment of skin sensitizers (GARD™skin)"). The outcome of the test is a binary hazard identification of skin sensitizers (78).

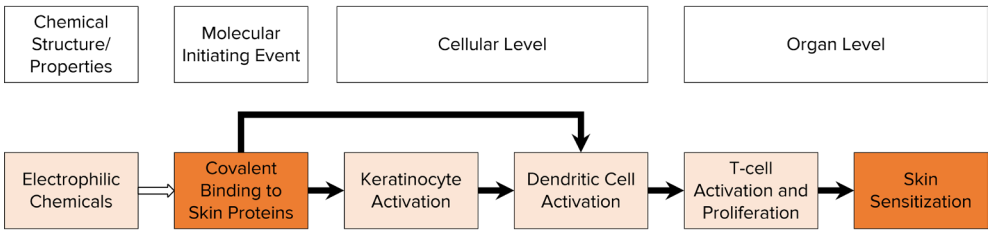


Figure 3. Schematic representation of the Adverse Outcome Pathway 40 "Covalent Protein binding leading to Skin Sensitization". According to the Molecular Initiating Event (MIE) 396 "Covalent Binding, Protein", a chemical binds covalently to the skin proteins. The causal relationships between the MIE, the

three Key Events (KE) "Activation, Keratinocytes" (KE 826), "Activation, Dendritic Cells" (KE 398) and "Activation/Proliferation, T-cells" (KE 272) and the eventual Adverse Outcome (AO) "Sensitization, skin" (AO 827) are depicted with black arrows.

Additionally, initiatives at the European and North American level have brought to the development of the Omics Data Analysis Framework for Regulatory Application (R-ODAF) (79, 80) and the National Toxicology Program's (NTP) Approach to Genomic Dose-Response Modeling (81), which provide guidelines for the analysis of toxicological transcriptomic data.

At present, there is no formal reporting scheme that includes all steps essential for performing analyses and promoting transparency and reproducibility. However, a framework for using omics technologies has been outlined by the OECD called the "OECD reporting framework" (82) to provide guidance to define a way of standardizing the report of omics and non-omics toxicological data and results for the application in regulatory toxicology and is currently undergoing trialing and evaluation. It is composed of both general and specific modules that have been developed for specific technologies (i.e. microarray, targeted and untargeted RNA-Seq and qPCR arrays for transcriptomics, mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy for metabolomics). This framework will be extended with new modules to incorporate novel technologies in the future. It includes reporting the experimental details that led to data collection, the omics assay, the data acquisition and analysis including the statistical analysis used for identifying differentially abundant transcripts or metabolites (Figure 4).

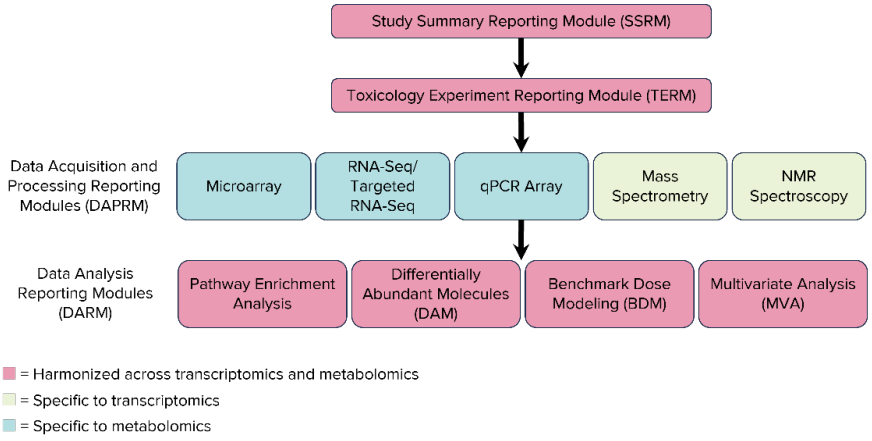


Figure 4. Schematic representation of the OECD reporting framework for transcriptomics and metabolomics. (Reproduced from Harrill et al. 2021).

In this context, this thesis aims at identifying endocrine disruption and elucidating the direct effects of four EDC classes in the thyroid using *in vitro* embryonic stem cell (ESC)-derived models from both human and mouse using omics technologies. For the bulk RNA-Seq analyses, we selected the Combo-Seq kit, allowing us to profile both mRNA and miRNAs from the same sample, and in Chapter 2 we compared it to conventional, separated, poly(A) and small RNA libraries and developed a custom pipeline for the data processing. In Chapter 3, we performed an in-depth study of the effect of phthalates on the mouse model. We exposed the organoids to 5 increasing doses of four compounds for 24 hours and observed a consistent upregulation of *Ing5*, whose product is a member of histone H3 and H4 acetylase complexes. Suspecting a potential effect on chromatin rearrangement, we further investigated the effects of a 5 days of treatment on chromatin accessibility via ATAC-Seq. In Chapter 4, we broadened the focus, using the full panel of EDCs selected for SCREENED, plus two compounds with a known direct (deleterious) effect on TH production. We combined transcriptomic and proteomic dataset to investigate common changes, concomitant miRNAs and target proteins dysregulation, and to construct random forest classification models. In Chapter 5, we used the human model to study the influence of sex hormones on the response to the two compounds BAP and PCB153 using single cell RNA-Seq. In Chapter 6, we discuss the overall results and draw conclusions on the general, recurrent topics touched in this thesis.

4. References

1. Hiller-Sturmhöfel S, Bartke A. The endocrine system: an overview. Alcohol Health Res World. 1998;22(3):153-64.
2. Kleine B, Rossmanith WG. Hormones and the Endocrine System2016. doi:https://doi.org/10.1007/978-3-319-15060-4.
3. Lyons FM, Meeran K. The Physiology of the Endocrine System. International Anesthesiology Clinics. 1997;35(4):1-22.
4. Cheng SY, Leonard JL, Davis PJ. Molecular aspects of thyroid hormone actions. Endocr Rev. 2010;31(2):139-70, doi:https://doi.org/10.1210/er.2009-0007.
5. Luster M, Duntas LH, Wartofsky L. The Thyroid and Its Diseases: Springer; 2019.
6. Benvenga S, Tuccari G, Ieni A, Vita R. Thyroid gland: anatomy and physiology. Encyclopedia of Endocrine Diseases. 2018;4:382-90, doi:https://doi.org/10.1016/B978-0-12-801238-3.96022-7.
7. Srinivasan A, Wong FK, Karponis D. Calcitonin: A useful old friend. J Musculoskelet Neuronal Interact. 2020;20(4):600-9.
8. Carvalho DP, Dupuy C. Thyroid hormone biosynthesis and release. Mol Cell Endocrinol. 2017;458:6-15, doi:https://10.1016/j.mce.2017.01.038.
9. Mondal S, Raja K, Schweizer U, Mugesh G. Chemistry and Biology in the Biosynthesis and Action of Thyroid Hormones. Angew Chem Int Ed Engl. 2016;55(27):7606-30, doi:10.1002/anie.201601116.
10. Grasberger H, Refetoff S. Identification of the maturation factor for dual oxidase. Evolution of an eukaryotic operon equivalent. J Biol Chem. 2006;281(27):18269-72, doi:https://doi.org/10.1074/jbc.C600095200.
11. Coscia F, Taler-Vercic A, Chang VT, Sinn L, O'Reilly FJ, Izore T, et al. The structure of human thyroglobulin. Nature. 2020;578(7796):627-30, doi:10.1038/s41586-020-1995-4.
12. Crantz FR, Silva JE, Larsen PR. An analysis of the sources and quantity of 3,5,3'-triiodothyronine specifically bound to nuclear receptors in rat cerebral cortex and cerebellum. Endocrinology. 1982;110(2):367-75, doi:https://doi.org/10.1210/endo-110-2-367.
13. Schweizer U, Kohrle J. Function of thyroid hormone transporters in the central nervous system. Biochim Biophys Acta. 2013;1830(7):3965-73, doi:https://doi.org/10.1016/j.bbagen.2012.07.015.
14. Wu Y, Koenig RJ. Gene Regulation by Thyroid Hormone. Trends in Endocrinology & Metabolism. 2000;11(6):207-11, doi:https://doi.org/10.1016/s1043-2760(00)00263-0.
15. Zekri Y, Guyot R, Flamant F. An Atlas of Thyroid Hormone Receptors' Target Genes in Mouse Tissues. Int J Mol Sci. 2022;23(19), doi:https://doi.org/10.3390/ijms231911444.
16. Bianco AC, Salvatore D, Gereben Bz, Berry MJ, Larsen PR. Biochemistry, Cellular and Molecular Biology, and Physiological Roles of the Iodothyronine Selenodeiodinases. Endocrine Reviews. 2002;23(1):38-89, doi:https://doi.org/10.1210/edrv.23.1.0455.
17. Dentice M, Marsili A, Zavacki A, Larsen PR, Salvatore D. The deiodinases and the control of intracellular thyroid hormone signaling during cellular differentiation. Biochim Biophys Acta. 2013;1830(7):3937-45, doi:https://doi.org/10.1016/j.bbagen.2012.05.007.

18. Larsen PR, Zavacki AM. The role of the iodothyronine deiodinases in the physiology and pathophysiology of thyroid hormone action. Eur Thyroid J. 2012;1(4):232-42, doi:https://doi.org/10.1159/000343922.
19. Sabatino L, Vassalle C, Del Seppia C, Iervasi G. Deiodinases and the Three Types of Thyroid Hormone Deiodination Reactions. Endocrinol Metab (Seoul). 2021;36(5):952-64, doi:https://doi.org/10.3803/EnM.2021.1198.
20. van der Spek AH, Fliers E, Boelen A. The classic pathways of thyroid hormone metabolism. Mol Cell Endocrinol. 2017;458:29-38, doi:https://doi.org/10.1016/j.mce.2017.01.025.
21. Moreno M, Berry MJ, Horst C, Thoma R, Goglia F, Harney JW, et al. Activation and inactivation of thyroid hormone by type I iodothyronine deiodinase. FEBS Letters. 1994;344(2-3):143-6, doi:https://doi.org/10.1016/0014-5793(94)00365-3.
22. Kaplan MM, Yaskoski KA. Phenolic and tyrosyl ring deiodination of iodothyronines in rat brain homogenates. J Clin Invest. 1980;66(3):551-62, doi:https://doi.org/10.1172/jci109887.
23. Huang TS, Chopra IJ, Beredo A, Solomon DH, Chua Teco GN. Skin is an active site for the inner ring monodeiodination of thyroxine to 3,3',5'-triiodothyronine. Endocrinology. 1985;117(5):2106-13, doi:https://doi.org/10.1210/endo-117-5-2106.
24. National Institute of Environmental Health Sciences. Endocrine Disruptors [updated August 18, 2022. Available from: https://www.niehs.nih.gov/health/topics/agents/endocrine/index.cfm.
25. La Merrill MA, Vandenberg LN, Smith MT, Goodson W, Browne P, Patisaul HB, et al. Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification. Nat Rev Endocrinol. 2020;16(1):45-57, doi:10.1038/s41574-019-0273-8.
26. Phosphorus, Inorganic and Nitrogen Flame Retardants Association (PINFA) [Available from: https://www.pinfa.eu/.
27. Farhat A, Crump D, Chiu S, Williams KL, Letcher RJ, Gauthier LT, Kennedy SW. In Ovo Effects of Two Organophosphate Flame Retardants—TCPP and TDCPP—on Pipping Success, Development, mRNA Expression, and Thyroid Hormone Levels in Chicken Embryos. Toxicological Sciences. 2013;134(1):92-102, doi:https://doi.org/10.1093/toxsci/kft100.
28. Meeker JD, Stapleton HM. House dust concentrations of organophosphate flame retardants in relation to hormone levels and semen quality parameters. Environ Health Perspect. 2010;118(3):318-23, doi:https://doi.org/10.1289/ehp.0901332.
29. Kim S, Jung J, Lee I, Jung D, Youn H, Choi K. Thyroid disruption by triphenyl phosphate, an organophosphate flame retardant, in zebrafish (Danio rerio) embryos/larvae, and in GH3 and FRTL-5 cell lines. Aquat Toxicol. 2015;160:188-96, doi:https://doi.org/10.1016/j.aquatox.2015.01.016.
30. Liu X, Cai Y, Wang Y, Xu S, Ji K, Choi K. Effects of tris(1,3-dichloro-2-propyl) phosphate (TDCPP) and triphenyl phosphate (TPP) on sex-dependent alterations of thyroid hormones in adult zebrafish. Ecotoxicol Environ Saf. 2019;170:25-32, doi:https://doi.org/10.1016/j.ecoenv.2018.11.058.

31. Babich MA. Overview of phthalates toxicity. US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/phthalover.pdf.
32. Carlson KR. Toxicity review of Di(2-ethylhexyl) Phthalate (DEHP). US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/ToxicityReviewOfDEHP.pdf.
33. U.S EPA. Phthalates Action Plan2012. Available from: https://www.epa.gov/sites/default/files/2015-09/documents/phthalates_actionplan_revised_2012-03-14.pdf.
34. Holland M. Socio-economic assessment of phthalates. 2018, doi:https://doi.org/10.1787/a38a0e34-en.
35. Phthalates NRCUCotHRo. 2 - Phthalate Exposure Assessment in Humans. Phthalates and Cumulative Risk Assessment: The Tasks Ahead. Washington (DC): National Academies Press (US); 2008. <https://www.ncbi.nlm.nih.gov/books/NBK215044/>.
36. Wittassek M, Koch HM, Angerer J, Bruning T. Assessing exposure to phthalates - the human biomonitoring approach. Mol Nutr Food Res. 2011;55(1):7-31, doi:https://doi.org/10.1002/mnfr.201000121.
37. Versar Inc. Review of Exposure Data and Assessments for Select Dialkyl Ortho-PhthalatesFebruary 2010. Available from: https://www.cpsc.gov/s3fs-public/pthalexp.pdf.
38. (CDC) CfDCaP. Fourth National Report on Human Exposure to Environmental Chemicals, Updated Tables. January 2019;1, doi:https://doi.org/10.15620/cdc75822.
39. Haines DA, Saravanabhavan G, Werry K, Khoury C. An overview of human biomonitoring of environmental chemicals in the Canadian Health Measures Survey: 2007-2019. Int J Hyg Environ Health. 2017;220(2 Pt A):13-28, doi:https://doi.org/10.1016/j.ijheh.2016.08.002.
40. Wittassek M, Angerer J. Phthalates: metabolism and exposure. Int J Androl. 2008;31(2):131-8, doi:https://doi.org/10.1111/j.1365-2605.2007.00837.x.
41. Koch HM, Angerer J. Chapter 3A. Phthalates: Biomarkers and Human Biomonitoring. Biomarkers and Human Biomonitoring. Issues in Toxicology. Cambridge: Royal Society of Chemistry; 2011. p. 179-233.
42. Consumer Product Safety Improvement Act of 2008, Title I—Children's Product Safety, Sec. 108, Public Law 110-314 (2008). <https://www.cpsc.gov/s3fs-public/pdfs/blk_pdf_cpsia.pdf>.
43. OJ. L 344, 27.12.2005, p. 43.
44. Fletcher EJ, Santacruz-Márquez R, Mourikes VE, Neff AM, Laws MJ, Flaws JA. Effects of Phthalate Mixtures on Ovarian Folliculogenesis and Steroidogenesis. Toxics. 2022;10(5):251, doi:https://doi.org/10.3390/toxics10050251.
45. Phthalates NRCUCotHRo. 3 - Toxicity Assessment. Phthalates and Cumulative Risk Assessment: The Tasks Ahead. Washington (DC): National Academies Press (US); 2008. <https://www.ncbi.nlm.nih.gov/books/NBK215030/>.
46. Babich MA, Osterhout CA. Toxicity review of Diisononyl Phthalate (DINP). US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/ToxicityReviewOfDINP.pdf.

47. Bereketoglu C, Pradhan A. Plasticizers: negative impacts on the thyroid hormone system. Environ Sci Pollut Res Int. 2022;29(26):38912-27, doi:https://doi.org/10.1007/s11356-022-19594-0.
48. Carlson KR. Toxicity review of Di-n-Octyl Phthalate (DnOP). US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. March 2010. Available from: https://www.cpsc.gov/s3fs-public/pdfs/blk_media_toxicityDNOP.pdf.
49. Ghisari M, Bonefeld-Jorgensen EC. Effects of plasticizers and their mixtures on estrogen receptor and thyroid hormone functions. Toxicol Lett. 2009;189(1):67-77, doi:https://doi.org/10.1016/j.toxlet.2009.05.004.
50. Wu H, Zhang W, Zhang Y, Kang Z, Miao X, Na X. Novel insights into di-(2-ethylhexyl)phthalate activation: Implications for the hypothalamus-pituitary-thyroid axis. Mol Med Rep. 2021;23(4), doi:https://doi.org/10.3892/mmr.2021.11930.
51. Kim MJ, Kim HH, Song YS, Kim OH, Choi K, Kim S, et al. DEHP Down-Regulates Tshr Gene Expression in Rat Thyroid Tissues and FRTL-5 Rat Thyrocytes: A Potential Mechanism of Thyroid Disruption. Endocrinol Metab (Seoul). 2021;36(2):447-54, doi:https://doi.org/10.3803/EnM.2020.920.
52. Huang P-C, Kuo P-L, Guo Y-L, Liao P-C, Lee C-C. Associations between urinary phthalate monoesters and thyroid hormones in pregnant women. Human Reproduction. 2007;22(10):2715-22, doi:https://doi.org/10.1093/humrep/dem205.
53. Hyunok Choi, Roy Harrison, Hannu Komulainen, Saborit JMD. Polycyclic aromatic hydrocarbons. 2010. Geneva: World Health Organization. WHO Guidelines for Indoor Air Quality: Selected Pollutants. Available from: https://www.ncbi.nlm.nih.gov/books/NBK138709/.
54. World Health Organization. Regional Office for Europe. Air quality guidelines for Europe, 2nd ed.: WHO Regional Publications, European Series; 91; 2000. Available from: https://apps.who.int/iris/handle/10665/107335.
55. PAHs: An Ecotoxicological Perspective. Douben PET, editor: John Wiley & Sons, Ltd; 2003. doi:https://doi.org/10.1002/0470867132.
56. Washington DC: US Environmental Protection Agency. Polycyclic organic matter (POM)2016. <https://www.epa.gov/sites/default/files/2016-09/documents/polycyclic-organic-matter.pdf>.
57. Sun H, Shen O-X, Xu X-L, Song L, Wang X-R. Carbaryl, 1-naphthol and 2-naphthol inhibit the beta-1 thyroid hormone receptor-mediated transcription in vitro. Toxicology. 2008;249(2):238-42, doi:https://doi.org/10.1016/j.tox.2008.05.008.
58. Bekki K, Takigami H, Suzuki G, Tang N, Hayakawa K. Evaluation of Toxic Activities of Polycyclic Aromatic Hydrocarbon Derivatives Using In Vitro Bioassays. Journal of Health Science. 2009;55(4):601-10, doi:https://doi.org/10.1248/jhs.55.601.
59. Atsdr. Toxicological profile for polycyclic aromatic hydrocarbons. ATSDR Tox Profile. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service; 1995. Report No.: CIS/97/00215. <https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=25>.
60. Teles M, Oliveira M, Pacheco M, Santos MA. Endocrine and metabolic changes in Anguilla anguilla L. following exposure to β-naphthoflavone—a microsomal enzyme

inducer. Environment International. 2005;31(1):99-104, doi:https://doi.org/10.1016/j.envint.2004.07.003.
61. Johnson S, McKillop D, Miller J, Smith IK. The Effects on Rat Thyroid Function of an Hepatic Microsomal Enzyme Inducer. Human & Experimental Toxicology. 1993;12(2):153-8, doi:https://doi.org/10.1177/096032719301200021.
62. Song M, Kim Y-J, Park Y-K, Ryu J-C. Changes in thyroid peroxidase activity in response to various chemicals. Journal of Environmental Monitoring. 2012;14(8):2121-6, doi:10.1039/C2EM30106G.
63. Agency for Toxic Substances and Disease Registry (ATSDR) (US). Toxicological Profile for Polychlorinated Biphenyls (PCBs). Department of Health and Human Services, Public Health Service; 2000. Report No.: NTIS/02928747_a.
64. Agency for Toxic Substances and Disease Registry (ATSDR) (US). 3.6 ENDOCRINE DISRUPTION. 2000. In: Toxicological Profile for Polychlorinated Biphenyls (PCBs) [Internet]. Atlanta (GA). Available from: https://www.ncbi.nlm.nih.gov/books/NBK587429/.
65. Van Birgelen AP, Smit EA, Kampen IM, Groeneveld CN, Fase KM, Van der Kolk J, et al. Subchronic effects of 2,3,7,8-TCDD or PCBs on thyroid hormone metabolism: use in risk assessment. Eur J Pharmacol. 1995;293(1):77-85, doi:https://doi.org/10.1016/0926-6917(95)90021-7.
66. Craft ES, DeVito MJ, Crofton KM. Comparative responsiveness of hypothyroxinemia and hepatic enzyme induction in Long-Evans rats versus C57BL/6J mice exposed to TCDD-like and phenobarbital-like polychlorinated biphenyl congeners. Toxicol Sci. 2002;68(2):372-80, doi:https://doi.org/10.1093/toxsci/68.2.372.
67. Boas M, Feldt-Rasmussen U, Main KM. Thyroid effects of endocrine disrupting chemicals. Molecular and Cellular Endocrinology. 2012;355(2):240-8, doi:https://doi.org/10.1016/j.mce.2011.09.005.
68. Liu J, Liu Y, Barter RA, Klaassen CD. Alteration of thyroid homeostasis by UDP-glucuronosyltransferase inducers in rats: a dose-response study. J Pharmacol Exp Ther. 1995;273(2):977-85.
69. Hood A, Klaassen CD. Differential Effects of Microsomal Enzyme Inducers on in Vitro Thyroxine (T4) and Triiodothyronine (T3) Glucuronidation. Toxicological Sciences. 2000;55(1):78-84, doi:https://doi.org/10.1093/toxsci/55.1.78.
70. OECD. Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Paris: OECD Publishing; 2018.
71. Beekhuijzen M, Rijk JCW, Meijer M, de Raaf MA, Pelgrom S. A critical evaluation of thyroid hormone measurements in OECD test guideline studies: Is there any added value? Reprod Toxicol. 2019;88:56-66, doi:https://doi.org/10.1016/j.reprotox.2019.07.014.
72. Rovida C, Busquet F, Leist M, Hartung T. REACH out-numbered! The future of REACH and animal numbers. ALTEX. 2023;40(3):367-88, doi:https://doi.org/10.14573/altex.2307121.
73. Knight J, Hartung T, Rovida C. 4.2 million and counting… The animal toll for REACH systemic toxicity studies. Altex. 2023;40(3):389-407, doi:https://doi.org/10.14573/altex.2303201.

74. Gant TW, Auerbach SS, Von Bergen M, Bouhifd M, Botham PA, Caiment F, et al. Applying genomics in regulatory toxicology: a report of the ECETOC workshop on omics threshold on non-adversity. Arch Toxicol. 2023;97(8):2291-302, doi:https://doi.org/10.1007/s00204-023-03522-3.
75. OECD. Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). OECD Series on Testing and Assessment, No. 260, Paris: OECD Publishing; 2017. <https://www.oecd-ilibrary.org/content/publication/44bb06c1-en> doi:https://doi.org/10.1787/44bb06c1-en.
76. (ECHA) ECA. Grouping of substances and read-across [Available from: https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/grouping-of-substances-and-read-across.
77. ECHA. Report on the operation of REACH and CLP 2021: Publications Office; 2021. doi:https://data.europa.eu/doi/10.2823/041059.
78. OECD. Test No. 442E: In Vitro Skin Sensitisation: In Vitro Skin Sensitisation assays addressing the Key Event on activation of dendritic cells on the Adverse Outcome Pathway for Skin Sensitisation. OECD Guidelines for the Testing of Chemicals, Section 4, Paris: OECD Publishing; 2023. <https://www.oecd-ilibrary.org/content/publication/9789264264359-en> doi:https://doi.org/10.1787/9789264264359-en.
79. CEFIC C4 team. Omics Data Analysis Framework for Regulatory application (R-ODAF) 2021 [Available from: https://github.com/R-ODAF/Main.
80. Verheijen MC, Meier MJ, Asensio JO, Gant TW, Tong W, Yauk CL, Caiment F. R-ODAF: Omics data analysis framework for regulatory application. Regul Toxicol Pharmacol. 2022;131:105143, doi:https://10.1016/j.yrtph.2022.105143.
81. NTP U. NTP research report on National Toxicology Program Approach to genomic dose–response modeling. Research Report 5; 2018.
82. Harrill JA, Viant MR, Yauk CL, Sachana M, Gant TW, Auerbach SS, et al. Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology. Regul Toxicol Pharmacol. 2021;125:105020, doi:https://doi.org/10.1016/j.yrtph.2021.105020.

Chapter 2

CODA: a Combo-Seq Data Analysis Workflow

Marta Nazzari, Duncan Hauser, Marcel van Herwijnen, Mírian Romitti, Daniel J Carvalho, Anna M Kip, Florian Caiment. **CODA: a combo-Seq data analysis workflow.** *Briefings in Bioinformatics.* 2023;24(1), bbac582.

Abstract

The analysis of the combined mRNA and miRNA content of a biological sample can be of interest for answering several research questions, like biomarkers discovery, or mRNA-miRNA interactions. However, the process is costly and time-consuming, separate libraries need to be prepared and sequenced on different flowcells. Combo-Seq is a library prep kit that allows to prepare combined mRNA-miRNA libraries starting from very low total RNA. To date, no dedicated bioinformatics method exists for the processing of Combo-Seq data. In this paper, we describe CODA (Combo-seq Data Analysis), a workflow specifically developed for the processing of Combo-Seq data that employs existing free-to-use tools. We compare CODA to exceRpt, the pipeline suggested by the kit manufacturer for this purpose. We also evaluate how Combo-Seq libraries analyzed with CODA perform compared to conventional poly(A) and small RNA libraries prepared from the same samples. We show that using CODA more successfully trimmed reads are recovered compared to exceRpt, and the difference is more dramatic with short sequencing reads. We demonstrate how Combo-Seq identifies as many genes and fewer miRNAs compared to the standard libraries, and how miRNA validation favours conventional small RNA libraries over Combo-Seq. The CODA code is available at https://github.com/marta-nazzari/CODA.

1. Introduction

The analysis of the RNA content of a biological sample, referred to as transcriptomics, has become a routine practice for many fields of biology. The transcriptome comprises several types of RNA, called biotypes (1), whose composition varies depending on the type of sample or cell model (2-5). Some of the most frequently studied biotypes are messenger RNAs (mRNA) and micro RNAs (miRNAs) due to their link with protein expression levels or for their biomarker potential (6-11). mRNAs generally possess a 3' poly(A) tail and are usually 1 kilobase or longer (12-14), while miRNAs are short (20-22 nucleotides) long non-coding RNAs (15).

Currently, simultaneous mRNA and miRNA analysis from the same sample is often performed by preparing separate sequencing libraries for the two RNA species. These libraries follow two very different protocols for selecting the desired RNA: mRNA libraries protocols either perform positive poly(A) selection, capturing all RNA species that possess a 3' poly(A) tail (so called "poly(A) libraries"), or perform a negative rRNA selection, by using baits targeting the ribosomal RNA (rRNA) to deplete these species from the total RNA (termed "ribodepleted libraries"). To sequence the miRNA content of a sample, a small RNA library needs to be prepared, which selects small RNAs by performing a size-selection. The separate preparation of two libraries can pose a problem for samples that have very low starting material or RNA content. Moreover, mRNA and miRNA sequencing libraries need to be sequenced on separate flowcells, due to the different number of cycles required (since the insert sizes differ greatly), and by the type of reads generated. In fact, mRNA reads are sequenced paired-end, while short RNA libraries single-end. If longer and shorter fragments were mixed in the same flowcell, the short fragments would tend to outcompete the longer fragments, which would result in the former being overrepresented and latter being undersequenced. When sequencing continues through the full fragment, there can be a sharp decline in base quality and the sequencing run could be potentially aborted (16). Lastly, when different library prep kits are used, barcodes and barcode collision must be considered to confirm compatibility for multiplexing.

One commercially available library prep kit that aims to overcome these limitations is the NEXTFLEX® Combo-Seq™ library prep kit that allows to prepare combined mRNA/miRNA libraries starting from very little input total RNA (between 5 ng and 100 ng) (17). In this method, poly(A) RNAs are first selectively retrotranscribed; RNA-DNA hybrids are then digested by RNase H into small fragments; the sample then contains mRNA-derived fragments and short RNAs of comparable length, that are further processed in the same way. As such, miRNAs but also other similarly short RNA species, like small nucleolar RNAs (snoRNAs), can be captured. The final library contains sequences of homogeneous size, that can be then sequenced in a single flowcell.

Recently, Illumina commercialized the v1.5 S4 35 cycle kit (18) for the NovaSeq6000 sequencer, which generates short reads. As the average insert length of a Combo-Seq library is

21-22 nt, it possesses a very suitable length for being sequenced on a 35-cycles flowcell, further reducing the price per sample to study the full transcriptome at very high throughput.

To date, no dedicated pipeline exists to process Combo-Seq generated datasets, and the manufacturer of Combo-Seq recommends using exceRpt (19). This toolkit was developed for the analysis of extracellular RNAs but can be adapted to the analysis of WGS/exome and long RNA-seq data according to the authors (20). The published works we could find using Combo-Seq libraries employ exceRpt for their bioinformatics analysis (17, 21, 22). As it was not originally developed for Combo-Seq, it presents some problems when used on this type of data: the references use custom gene annotations that group together gene counts based on the biotype, making it a hybrid between a gene and transcript level analysis. They are custom-made by the developers and available only for human (hg19 and hg38) and mouse (mm10). The user is thus bound to using them and they cannot be changed or updated, as it is not possible to prepare a reference using a genome version downloaded from common repositories such as Ensembl and Gencode. In consequence, the applicability of Combo-Seq libraries is reduced to only two species with an outdated reference.

To our knowledge, no independent evaluation of Combo-Seq has been performed, and users could wonder if it provides results comparable to using a combination of poly(A) and miRNA libraries. For example, mRNA is fragmented in a different way (enzymatically in the Combo-Seq protocol or chemically in most poly(A) libraries), and mRNA fragments undergo a different size selection, since Combo-Seq retains short poly(A)-derived fragments, while standard poly(A) libraries usually include 300 nt-long inserts and longer (and thus exclude the shorter mRNA fragments from the pool).

In this work, we developed a custom-tailored workflow for the processing of Combo-Seq data which uses existing tools commonly used in RNA-Seq data analysis and compared it to exceRpt. We generated Combo-Seq libraries from two different *in vitro* cell models and sequenced them in 100- or 35-cycles flowcells. We processed them with CODA or exceRpt and noticed how exceRpt discards part of the reads during the trimming step. We show that this is more dramatic as the average read length decreases and it is more biased toward some RNA species. We also provide an evaluation of Combo-Seq performance compared to conventional poly(A) and small RNA libraries prepared from the same RNA samples. We performed differential expression analysis to compare the dysregulated genes and miRNA after benzo[a]pyrene treatment that can be identified with the different libraries. We show that the differentially expressed genes partially overlap between the two types of libraries, while there is no overlap of differentially expressed miRNAs. In addition, we performed miRNA RT-qPCR validation to solve discrepancies between conventional small RNA libraries and Combo-Seq quantification.

2. Materials and Methods

2.1 Thyroid Follicles Culture and Enrichment

2.1.1 Embryonic Stem Cells Culture and Differentiation

Mouse embryonic stem cell-derived thyroid follicles were differentiated and enriched as described previously (23, 24). The A2Lox.Cre_TRE-Nkx2-1/Pax8_Tg-EGFP mouse ESC cells were initially cultured on gamma-irradiated mouse embryonic fibroblasts (MEFs) feeder using mouse stem cell medium [23, 24] and incubated at 37 °C, 5% CO2 and > 95% humidity. For differentiation into thyroid, embryoid bodies (EBs) were generated by hanging drops culture of ESCs (1,000 cells per drop) for 4 days, then collected and embedded in growth factor-restricted Matrigel (354230, Corning); 50 ml MTG drops containing around 20 EBs were plated into 12-well plates. Cells were differentiated using differentiation medium (composed of DMEM (31966021, Gibco) supplemented with 15% FBS, vitamin C (50 µg/mL) (A4544, Sigma), nonessential amino acids (0.1 mM) (11140035, Gibco), sodium pyruvate (1 mM) (11360039, Gibco), penicillin and streptomycin (50 U/mL) (15140122, Gibco), 2-mercaptoethanol (0.1 mM) (31350010, Gibco)) supplemented with 1 µg/mL of doxycycline (D9891-1G, Sigma) for 3 days for *Nkx2-1* and *Pax8* induction, followed by two weeks of maturation by using differentiation medium containing 0.3 nM of 8-Br-cAMP (B 007-500, BioLog).

2.1.2 Follicles Enrichment Protocol

After complete maturation (day 21), Matrigel drops containing the thyroid follicles were washed twice with Hanks's balanced salt solution (HBSS, containing calcium and magnesium) (14025050, Gibco) and incubated in a HBSS solution (1 mL/well) containing 10 U/mL of dispase® II (4942078001, Roche) and 125 U/mL of collagenase type IV (Sigma) for 30-45 min at 37 °C. The enzymes were inactivated with 10% FBS and organoids were centrifuged at 300×g for 3 min. They were resuspended in differentiation medium and enriched for thyroid follicles by filtration using a 100 mm cell strainer (43-50100, pluriSelect Life Science GmbH) and 30 mm reverse strainer (43-50030, pluriSelect Life Science GmbH).

2.2 Datasets

The datasets used in this paper were obtained from two sources: (1) the human epithelial follicular cell line Nthy-ori 3-1 was seeded at a density of 40,000 cells/cm^2 in a 6-well plate and exposed for 24 hours in triplicate to 1 µM or 10 µM benzo[a]pyrene (BAP) (B-1760, Sigma) dissolved in DMSO (1029521000, Merck) (final concentration of DMSO 0.5%). Six DMSO controls were included. Cells were cultured in RPMI 1640 Medium, GlutaMAX™

Supplement (61870036, Gibco) with 10% FBS and 100 U/mL Penicillin-Streptomycin (15140122, Gibco). (2) Enriched thyroid follicles were exposed to DMSO 0.5% for 24 hours (5 biological replicates). For culture, the differentiation medium was supplemented with 8-Br-cAMP (0.3 nM) and TGF-βRI inhibitor SB431542 (10 µM) (1614, Tocris). At the end of the exposure time, cells were lysed in QIAzol Lysis Reagent (79306, Qiagen). Total RNA was extracted using the Direct-zol RNA miniprep (R2051, Zymo Research) for Nthy-ori 3-1 cells and with the miRNAeasy Micro Kit (217084, Qiagen) for the follicles.

2.3 Libraries Preparation and Sequencing

50 ng (Nthy-ori 3-1) and 20 ng (thyroid follicles) of total RNA were used as input for the NEXTFLEX® Combo-Seq™ mRNA/miRNA Kit (NOVA-5139-53, PerkinElmer). All RNA integrity number (RIN) values, as calculated by the Agilent software (25), were 8 or higher. tRNA fragments and Y RNA fragments were depleted with NEXTFLEX® tRNA/YRNA blocker. Thirteen (13) and 16 PCR cycles were performed for Nthy-ori 3-1 and follicles, respectively. Nthy-ori 3-1 samples were sequenced on an S2 Illumina flowcell 100 cycles (v1.5) (Illumina) in single-end mode; follicles samples were sequenced on an S4 Illumina flowcell 35 cycles (v1.5) (Illumina) in single-end mode. Throughout this paper, we will sometimes refer to the RNA-Seq data derived from the Nthy-ori 3-1 and the follicles as "1x100 dataset" and "1x35 dataset", respectively ("1x" means that both libraries were sequenced in single-end mode).

Poly(A) libraries were prepared on an automated system (Zephyr G3® NGS) with the NEXTFLEX® Rapid Directional RNA-Seq Kit 2.0 (NOVA-5198-02, PerkinElmer), NEXTFLEX® Poly(A) Beads 2.0 (NOVA-512992, PerkinElmer) and NEXTFLEX® Unique Dual Index Barcodes (NOVA-512923, PerkinElmer) using 1 µg of total RNA extracted from Nthy-ori 3-1 samples and performing 10 PCR cycles. The libraries were sequenced on an S1 Illumina flowcell 200 cycles (v1.5) (Illumina) in paired-end mode.

miRNA libraries were prepared manually with the NEXTFLEX® Small RNA-Seq Kit v3 (NOVA-5132, PerkinElmer) from 100 ng of total Nthy-ori 3-1 RNA and performing 18 PCR cycles. They were sequenced on an S4 Illumina flowcell 35 cycles (v1.5) (Illumina) in single-end mode.

All prepared libraries were quantified on a Qubit 2.0 Fluorometer (ThermoFisher), and quality control performed on the 2200 TapeStation System (Agilent) or BioAnalyzer 2100 expert (Agilent). The sequencing was done with the NovaSeq 6000 Sequencing System (Illumina).

2.4 Data analysis

2.4.1 RNA-Seq Data Processing

Data from Combo-Seq libraries was processed with exceRpt (v4.6.3, 2018-03-18) or CODA. When using exceRpt, we followed the parameters suggested by PerkinElmer (Table 1) (19).

Table 1. Comparison between default and modified arguments in the exceRpt pipeline to analyze Combo-Seq data as suggested by PerkinElmer.

Argument	Default	Modified
ADAPTER_SEQ	'guessKnown'	AAAAAAAA
MIN_READ_LENGTH	18	15
STAR_outFilterMatchNmin	18	15
RANDOM_BARCODE_LENGTH	0	4
RANDOM_BARCODE_LOCATION	'-5p -3p'	'-5p'

CODA is composed of three steps. Reads are trimmed using Cutadapt (v3.4) (26) and used as input in two different steps: miRNAs are quantified with miRge3.0 (v3.0) (27), while genes using RSEM (v1.3.3) with the --STAR option (v2.7.9a). The STAR parameters that are hard-coded in RSEM follow the ENCODE3's STAR-RSEM pipeline (28), and we opted for these options over the default STAR because they allow read alignment to more loci and permit fewer base mismatches than the default (Table 2) (29, 30).

Table 2. Comparison between default STAR arguments and RSEM-modified STAR parameters.

Argument	STAR (Default)	STAR (ENCODE3's STAR-RSEM pipeline)
--outFilterMultimapNmax	10	20
--outFilterMismatchNmax	10	999
--outFilterMismatchNoverLmax	0.3	0.04
--alignIntronMin	21	20
--alignIntronMax	0	1,000,000
--alignMatesGapMax	0	1,000,000
--alignSJoverhangMin	5	8
--alignSJDBoverhangMin	3	1

The primary assemblies of the human (GRCh38) and mouse (GRCm39) genomes were downloaded from Gencode (https://www.gencodegenes.org/) (31). For miRNA detection, the human (v22) and mouse (v22) annotations were obtained from miRBase (https://www.mirbase.org/) (32). BBMap (v38.94) (33), FastQC (v0.11.5) (34) and multiQC

(v1.11) (35) are used for quality control. The CODA code is available at https://github.com/marta-nazzari/CODA.

Demultiplexed data from small RNA libraries was processed as suggested by PerkinElmer (36) and used as input for miRge3.0. Demultiplexed data from poly(A) libraries was processed using a modified version of the Omics Data Analysis Framework for regulatory application (R-ODAF) (37, 38): reads were trimmed with fastp (39), and aligned and mapped using RSEM with the --STAR option.

2.4.2 Pipelines Comparison

exceRpt performs adapter trimming and reads size selection in several steps. First the 3' adapter sequence is removed, followed by a combined 5' 4N adapter trimming and exclusion of the 5' trimmed inserts shorter than 15 nt. As only reads that are 3' adapter trimmed are reported in the summary statistics by exceRpt, we retrieved the number of reads passing both trimming and size-selection filters from the output .log files. The count of reads trimmed by CODA is the "Reads passing filters" statistic output by Cutadapt.

The length distribution of trimmed reads was retrieved from the summary file "exceRpt_ReadLengths.txt" for the exceRpt pipeline, or after running FastQC for CODA.

To compare CODA and exceRpt, we analyzed the genes and miRNA counts separately. Since exceRpt uses custom-made annotations, a one-to-one feature comparison with the Gencode and miRge3.0 ones is not possible. For this reason, we summed all biotypes of the same gene in a single count for exceRpt. For CODA, we used the gene counts output by RSEM in the file 'SAMPLE_NAME.genes.results'. In addition, as exceRpt filters out reads mapping to all primary endogenous rRNA genes, we mapped all genes to the corresponding biotype using the biomaRt R package (40) (41) and removed all rRNA counts from both datasets. If a gene was identified in only one of the two samples, its count in the other sample was set to 0. To compare miRNA counts, we kept only the annotations that overlapped between the miRge3.0 and exceRpt outputs, as trying to manipulate the annotations to match the discordant annotations could introduce a bias. To evaluate the percentage of mapped miRNA reads, we calculated the total read counts per sample output by miRge3.0 and by exceRpt as reported on the file "exceRpt_miRNA_ReadCounts.txt". If a miRNA was identified in only one of the two pipelines for one sample, its count in the other was set to 0. The differences between the two workflows are reported in detail in Table 3.

Table 3. Main differences between the CODA and exceRpt workflows.

	Parameter	CODA	exceRpt
miRNA workflow	*miRNA reference*	miRBase v22	miRBase v21
	Software for mapping	Bowtie	STAR
	Software for quantification	miRge3.0	Authors' own code for quantification
Genes workflow	*Genes reference*	Gencode	Based on Gencode but modified by the authors
	Reference annotation	GRCh38, version 38 (Ensembl 104)	GRCh38, version 24 (Ensembl 83)
	Software for mapping	STAR v2.7.9a	STAR v2.5.4b
	STAR parameters that differ from the default in either workflow		
	--outFilterMatchNmin	0 (default)	15
	--outFilterMultimapNmax	20	10 (default)
	--outFilterMismatchNmax	999	10 (default)
	--outFilterMismatchNoverLmax	0.04	0.3 (default)
	--alignIntronMin	20	21 (default)
	--alignIntronMax	1,000,000	0 (default)
	--alignMatesGapMax	1,000,000	0 (default)
	--alignSJoverhangMin	8	9 (default)
	--alignSJDBoverhangMin	1	3 (default)
	--outFilterMatchNminOverLread	0.66 (default)	0.9
	--outFilterMismatchNmax	10 (default)	1
	Software for quantification	RSEM	Authors' own code for quantification

2.4.3 Trimmers Comparison

To show how the trimming steps of CODA and exceRpt compare to other trimmers, we processed the raw reads of dataset 1x100 with Cutadapt as set in CODA, exceRpt, AdapterRemoval (v2.3.3) [59] or BBDuk (v38.94) from the BBMap suite [33]. In Table 4 we report the parameters used for trimming with AdapterRemoval and BBDuk, and the ones used in Cutadapt (CODA) and exceRpt, for easy comparison.

Table 4. Arguments used for trimming the 1x100 dataset with Cutadapt as set up in CODA, exceRpt, AdapterRemoval and BBDuk. The variable [SAMPLE] is here a placeholder for the sample ID, and [OUTPUT_FOLDER] for the folder path where the trimmed files are saved.

Cutadapt (CODA)	exceRpt	AdapterRemoval	BBDuk

`cutadapt \`
`-u 4 \`
`-a AAAAAAAA \`
`-j 5 \`
	`fastx_clipper \`
	`-c \`
	`-Q33 \`
	`-a AAAAAAAA \`
		`AdapterRemoval \`
		`--file1 [SAMPLE] \`
		`--threads 10 \`
		`--adapter1 AAAAAAAA \`
			`bbduk.sh \`
			`in=[SAMPLE] \`
			`out=[OUTPUT_FOLDER][SAMPLE]".fastq.gz" \`

Cutadapt (CODA)	exceRpt	AdapterRemoval	BBDuk

`--minimum-length 15 \`
`--output [OUTPUT_FOLDER][SAMPLE].fastq.gz \`
`[SAMPLE]`
	`-l 15 \`
	`-v \`
	`-n \`
	`-M 7 \`
	`-i [SAMPLE] \`
	`-z \`
	`-o [OUTPUT_FOLDER][SAMPLE].fastq.clipped.fastq.tmp.gz`
		`--trim5p 4 \`
		`--gzip \`
		`--minlength 15 \`
		`--output1 [OUTPUT_FOLDER][SAMPLE]".fastq.gz" \`
		`--basename [OUTPUT_FOLDER][SAMPLE]`
			`minlen=15 \`
			`literal=AAAAAAAA \`
			`ftl=4 \`
			`ktrim=r \`
			`k=8 \`
			`mink=3`

2.4.3 Principal Component Analysis and Correlation Analysis

Principal component analyses were performed on variance-stabilized expression levels of normalized gene and miRNA read counts using the R package PCAtools (v2.4.0) (42). Pearson correlation was used to calculate the correlation between genes read count. To calculate the miRNA expression correlation among samples prepared with Combo-Seq and small RNA library prep kit, we ranked the miRNAs based on level of expression (miRNA with the highest read count = highest rank). When multiple miRNAs had the same read count, they were assigned the same rank with the highest value. miRNAs for which the read count was 0 in all samples were removed. To evaluate the correlation between miRNAs we used the non-parametric Spearman correlation.

2.4.4 Differential Expression and Gene Ontology Analysis

Differential expression analysis was performed with R using the DESeq2 (43) and edgeR (44). To select relevant differentially expressed genes and miRNA, stringent filtering was applied using a modified version of the R-ODAF. Briefly, a gene was considered expressed if its count per million (CPM) value is ≥1 in at least 75% of the replicates of either group (i.e., BAP or DMSO). In addition, differentially expressed genes and miRNAs identified by DESeq2 were filtered to remove spurious spikes (for details, see the paper by Verheijen, Meier et al. 2022) (37). To increase statistical power, all BAP samples were grouped together and compared to the DMSO controls. Gene ontology (GO) (2021) (45, 46) and Reactome (2022) (47) enrichment analyses were performed using the web-based tool Enrichr (48) and the FDR was set to 0.01.

2.5 miRNAs Reverse Transcription-qPCR

Total RNA from the six Nthy-ori 3-1 DMSO control samples was used for cDNA synthesis with the TaqMan® Advanced miRNA cDNA Synthesis Kit (A28007, Applied Biosystems) according to manufacturer's protocol. The synthesized cDNA was used for qPCR using the TaqMan™ Fast Advanced Master Mix (4444556, Applied Biosystems) and TaqMan™

Advanced miRNA Assay (A25576, Applied Biosystems) following the manufacturer's protocol for hsa-miR-122-5p (477855_mir), hsa-miR-361-3p (478055_mir), hsa-miR-622 (479106_mir). The program used for the qPCR reaction was 20 seconds at 95 °C (1 cycle), 3 seconds at 95 °C – 30 seconds at 60 °C (40 cycles) on a CFX Connect™ Real-Time System (Bio-Rad). Each sample was analyzed in 4 technical replicates. For each sample a technical replicate was retained if its Ct value difference from its closest other replicates was lower than 0.6.

3. Results

3.1 Description of CODA

For the processing of Combo-Seq data, we developed a pipeline named "CODA" (Combo-Seq Data Analysis), composed of three different steps (Figure 1). The first step uses Cutadapt to trim the 5' and 3' adapters and discard reads that are shorter than 15 nt. Since Combo-Seq generates libraries from both polyadenylated RNA species and miRNA, the workflow splits in two and the trimmed files are used as input for gene and miRNA mapping and quantification. To analyze the reads that derive from poly(A)-tailed species, the trimmed reads are aligned to the reference genome with STAR and quantified using RSEM. To identify miRNAs, the trimmed reads are used as input for miRge3.0. The genes and miRNAs count files output by both tools for each sample are then merged into a single table for genes and for miRNAs. The pipeline outputs a report with useful QC metrics that can be inspected by the user.

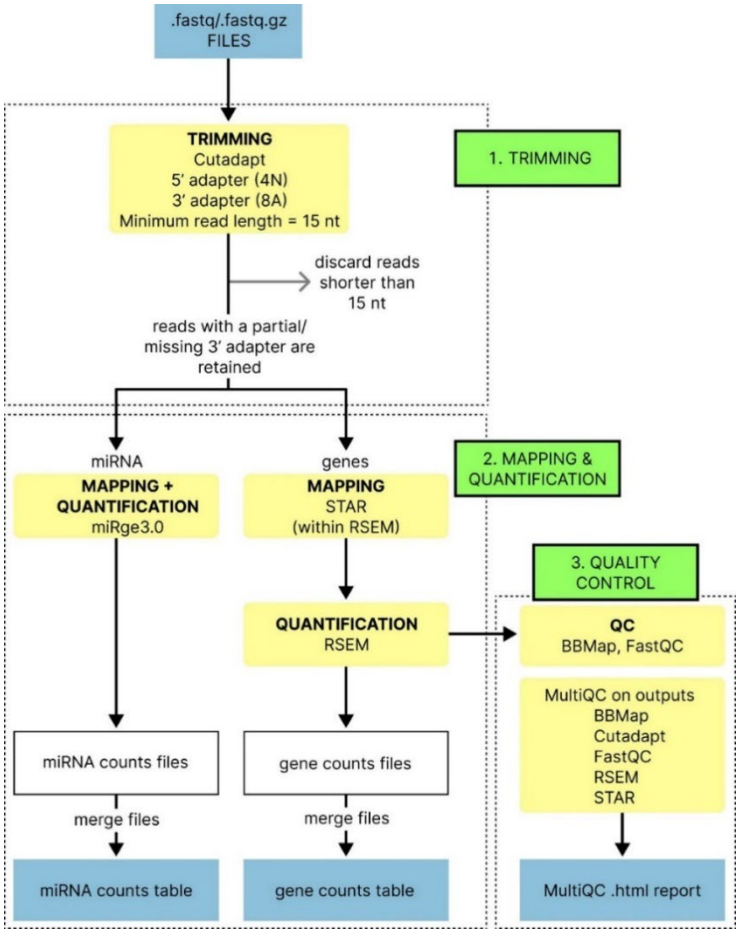


Figure 1. Schematic representation of CODA: fastq or fastq.gz sequencing files are used as input and 5' and 3' sequencing adapters are removed using Cutadapt. Reads shorter than 15 nt are also discarded. This trimming step retains also reads with a partial or missing 3' adapter (a point further discussed in the section "Trimming and read length distribution" below). Mapping and quantification of miRNA is then performed using miRge3.0. Genes mapping and quantification is then performed with RSEM using STAR as aligner (which follows the criteria of the ENCODE3's STAR-RSEM pipeline). As each tool outputs a single file per sample, the count files are then merged into a single table for genes or miRNAs. The last step uses the BBMap suite and FastQC to gather some summary statistics on the trimmed/mapped reads and MultiQC is used to compile all information into a .html report.

3.2 Comparison of CODA with exceRpt

3.2.1 Trimming and Read Length Distribution

Since the first step of both pipelines is adapter trimming, we evaluated the number of retained trimmed reads and their length distribution. Figure 2A shows the read length

distribution of the 1x100 samples trimmed with CODA or exceRpt. The maximum read length of samples processed with CODA (97 nt) is longer than the maximum length identified by exceRpt (90 nt). The median number of reads that pass adapter trimming is 89.44% and 90.90% of total input reads when using exceRpt or CODA, respectively (Figure 2B). On average, an additional 1.46% of total raw reads are retained by CODA and their length is between 91 and 97 nt. This increase is even more evident with datasets that have shorter reads, like the one generated from thyroid follicles sequenced on a 1x35-cycles flowcell (Figure 2C).

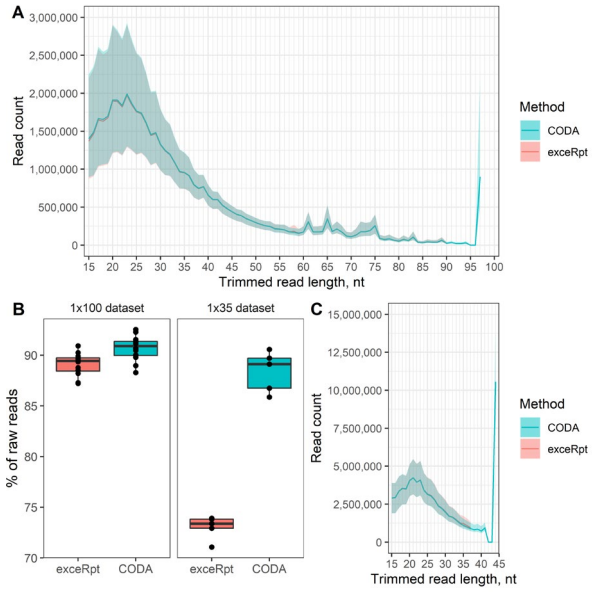


Figure 2. (A) Read length distribution of 1x100 dataset processed with exceRpt (red) or trimmed with CODA (blue). (B) Distribution of trimmed reads expressed as percentage of total raw read count. (C) Read length distribution of 1x35 dataset processed with exceRpt (red) or trimmed with CODA (blue). For plots A and C, the line represents the average count, while the edges of the shaded area correspond to the highest and lowest count among the replicates.

The maximum read length when using CODA is 44 nt, while it is 37 nt when using exceRpt. The median percentage of reads that successfully pass 5' and 3' adapter trimming when using exceRpt is 73.39% of total sequenced reads. This number increases to 89.12% when the same samples are trimmed with CODA, retaining an additional 15.73% of the raw reads. The difference in maximum read length identified by the two methods likely lies in the choice of trimmer. When sequencing single-end libraries (like Combo-Seq), if a fragment is longer than the total number of cycles of the chosen flowcell, only part of the fragment will be sequenced, thus partly or completely excluding the 3' adapter (since single-end libraries are always sequenced from the 5' end). Cutadapt has the option to retain such reads, while the one used by exceRpt

with the options set by the exceRpt developers (fastx -M 7) allows only 1 nt mismatch (Figure 3). If the reads are not clipped, they are discarded (determined by the -c flag in the fastx command) (49).

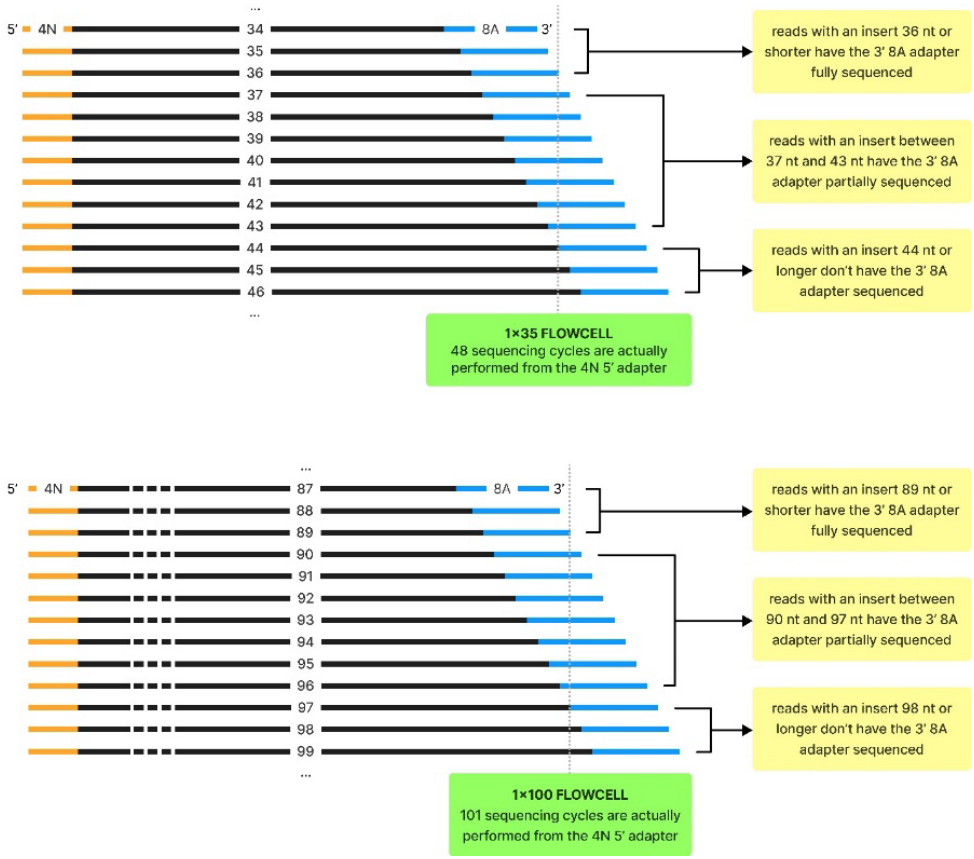


Figure 3. Schematic representation of how Combo-Seq reads are sequenced in a 1x35 or 1x100 Illumina flowcell. Each line represents a read, the black portion denotes the insert and the number in the middle reports the insert length. The 5' 4N and 3' 8A adapters are colored in yellow and blue, respectively, and the 5' and 3' ends are reported. The dots above and below the reads mean that there can be sequencing reads with longer or shorter inserts, as reads in a library are characterized by is a distribution of possible insert lengths. The green boxes explain that in 1x35 or 1x100 Illumina flowcells, the actual number of sequencing cycles is 48 and 101, respectively. For this reason, starting from the 5' 4N adapter, 48 nt and 101 nt can be sequenced in total in a 1x35 or 1x100 setting, respectively. In Combo-Seq libraries, when the length of adapters+insert is equal to or lower than the total sequencing cycles, they will be sequenced in their entirety. For 1x35 and 1x100 flowcells, this means that the insert needs to be 36 and 89 nt long, respectively. However, if the insert length is longer, only part of the 3' 8A adapter, or none of it, will be sequenced, leaving the sequenced read with an incomplete, or missing, adapter. Cutadapt provides the option to handle partial adapters, and in CODA we allow a minimum overlap of a partial adapter of 3 nt

(which is the default value and should result in just 0.07 bases being lost per read [60]). This means that an adapter that is 1A or 2A long will not be considered a partial adapter, but a part of the insert itself. This explains why in the 1x100 dataset the read count for inserts that are 95 or 96 nt long is 0 (as they are incorporated in the 97 nt long reads). Similarly, in the 1x35 dataset the reads that are 42 or 43 nt long (meaning they have a 3' adapter that is 2A or 1A, respectively) are incorporated in the 44 nt-long group. On the other hand, exceRpt uses fastx_clipper from the fastx-toolkit package [49] for 3' adapter trimming with the minimum overlap flag (-M) set to 7: this means that there has to be a minimum overlap of 7 bases between the provided 3' adapter (8A in case of Combo-Seq) and the adapter found in the read. If the match is less than 7, the read is discarded.

Effectively, exceRpt discards reads where the 3' adapter is either 6 A or shorter, or missing altogether. This factor is the most impactful on the trimmed read length distribution, as 1x100 dataset reads trimmed with two other trimmers retaining non-clipped reads show a similar distribution as Cutadapt, with a peak at the maximum read length, 97 nt (Figure 4).

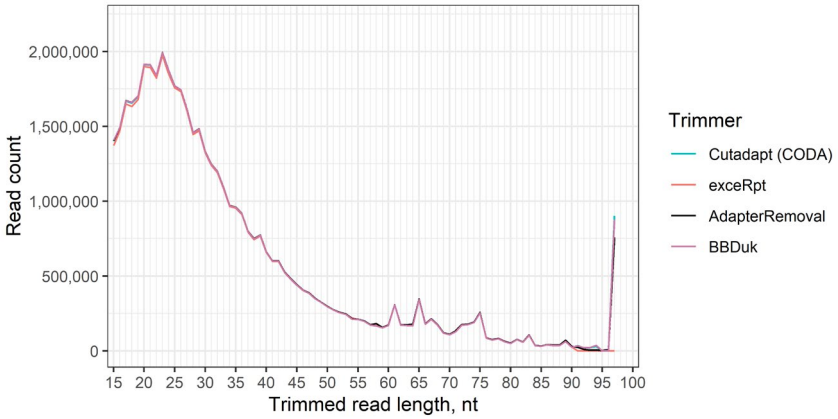


Figure 4. Reads distribution of the 1x100 dataset trimmed with several trimming software: Cutadapt as set up in CODA (cyan), exceRpt (red), AdapterRemoval (black), BBDuk (pink).

3.2.2 Comparison of Mapping and Quantification

Since Combo-Seq libraries capture both poly(A) species and miRNAs, our pipeline performs separately gene and miRNA mapping and quantification. We compared the reads mapping to genes or miRNA using CODA and exceRpt.

The mapping efficiency is comparable, evidenced by similar percentages of trimmed reads mapping to genes or miRNA in both datasets (Table 5). However, since the number of reads successfully passing trimming is higher when using CODA, samples processed with it have more total mapping reads compared to exceRpt, especially in the 1x35 dataset, going from a median 52.01 M (exceRpt) to 60.69 M (CODA) reads. The median count of miRNA-mapping reads is instead comparable between the two methods, supporting the observation that, for shorter inserts like miRNAs, the 3' adapters are fully sequenced, and the reads properly trimmed.

Table 5. Median counts and proportions of reads mapped to the reference transcriptome (or to exceRpt Gencode annotations) and miRNA and quantified with CODA or with exceRpt in the 1x100 and 1x35 datasets.

	1x100 dataset				1x35 dataset			
	Read count		Percentage of trimmed reads		Read count		Percentage of trimmed reads	
	CODA	exceRpt	CODA	exceRpt	CODA	exceRpt	CODA	exceRpt
Total trimmed reads	46.32 M	45.50 M	100%	100%	73.37 M	61.69 M	100%	100%
Genes mapping reads	40.07 M	39.36 M	86.50%	86.42%	60.69 M	52.01 M	84.14%	85.60%
miRNA mapping reads	0.13 M	0.12 M	0.27%	0.25%	1.40 M	1.39 M	1.80%	5.95%

3.2.3 Comparison of Genes and miRNAs

To analyze how many and which genes are identified by either pipeline, we first analyzed the overlap of the genes which have a raw read count greater than 0 in the 1x100 and 1x35 datasets processed either with CODA or exceRpt. While the total read count is higher when the samples are processed with CODA, exceRpt maps the reads to more genes (Figure 5A-C, G). This is not the case for miRNAs, where a comparable number of mapped features is detected by both pipelines (Figure 5D-F, H).

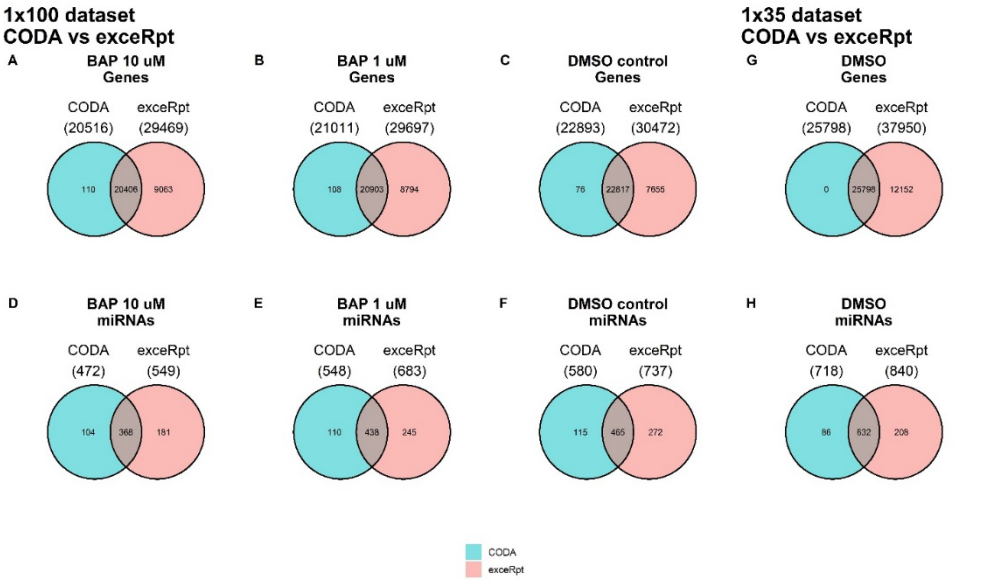


Figure 5. Venn diagrams of genes and miRNAs with raw read count > 0 identified in the 1x100 dataset (genes: A-C, miRNAs: D-F) or 1x35 dataset (genes: G, miRNAs: H) processed either with CODA (cyan) or exceRpt (red) and their overlap. For the 1x100 dataset, the samples are grouped by treatment (i.e., BAP 10 uM, BAP 1 uM, DMSO control). The total number of identified species is indicated in brackets below the method.

We then analyzed the raw read count distribution per RNA biotype in the 1x100 dataset (Figure 6A) and observed that exceRpt assigns reads to low-expressed genes belonging to several biotypes, whose count is instead 0 in CODA. The read count distribution for protein coding genes appears to be bimodal in both pipelines, with two peaks identifiable for "low" and "high" expressed genes (Figure 6B).

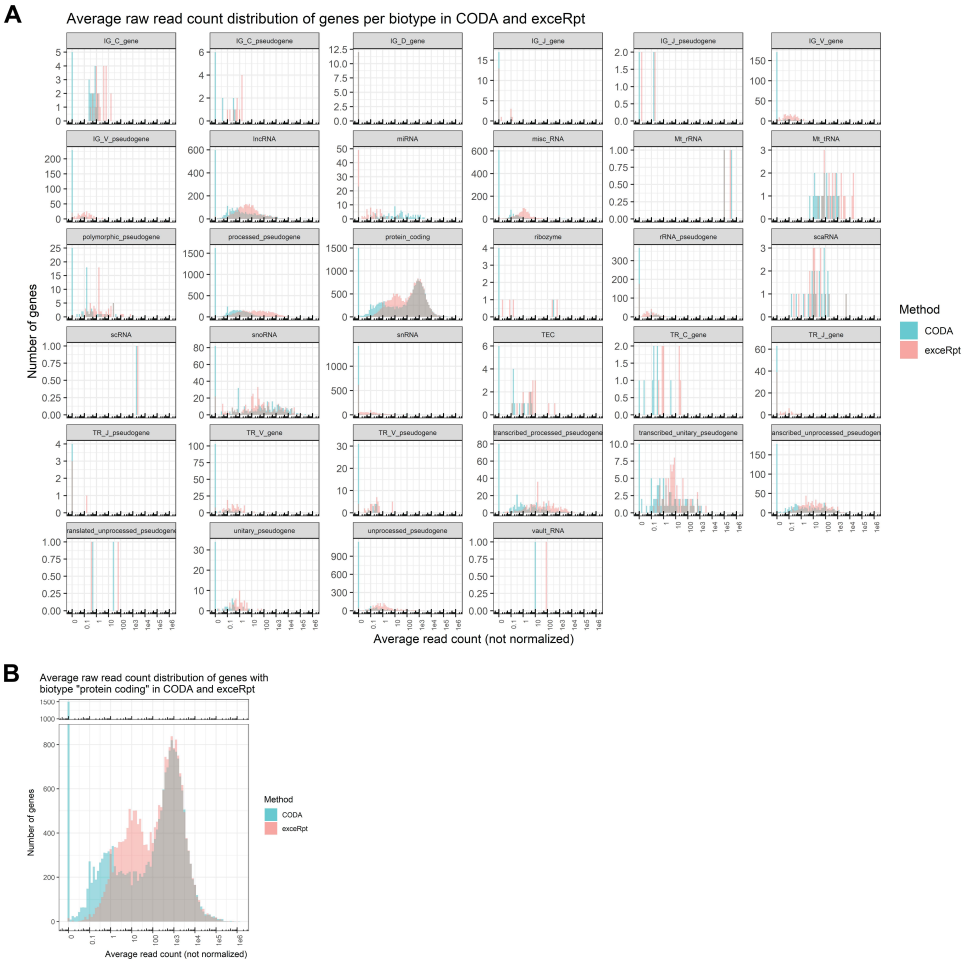


Figure 6. (A) Average raw read gene count in the 1x100 dataset processed with either CODA (cyan) or exceRpt (red) divided by RNA biotype. (B) The "protein coding" panel from figure A has been expanded. In both figures the x axis is log10-scaled for easier reading.

Looking at the PCA plots, samples cluster along PC1 according to the pipelined use both for genes (Figure 7A-B) and miRNAs (Figure 7C-D), for both the 1x100 and 1x35 datasets, showing how the processing method is the biggest source of variation. Additionally, clustering along PC2 reflects the BAP-treated versus control condition for the Nthy-ori 3-1 dataset.

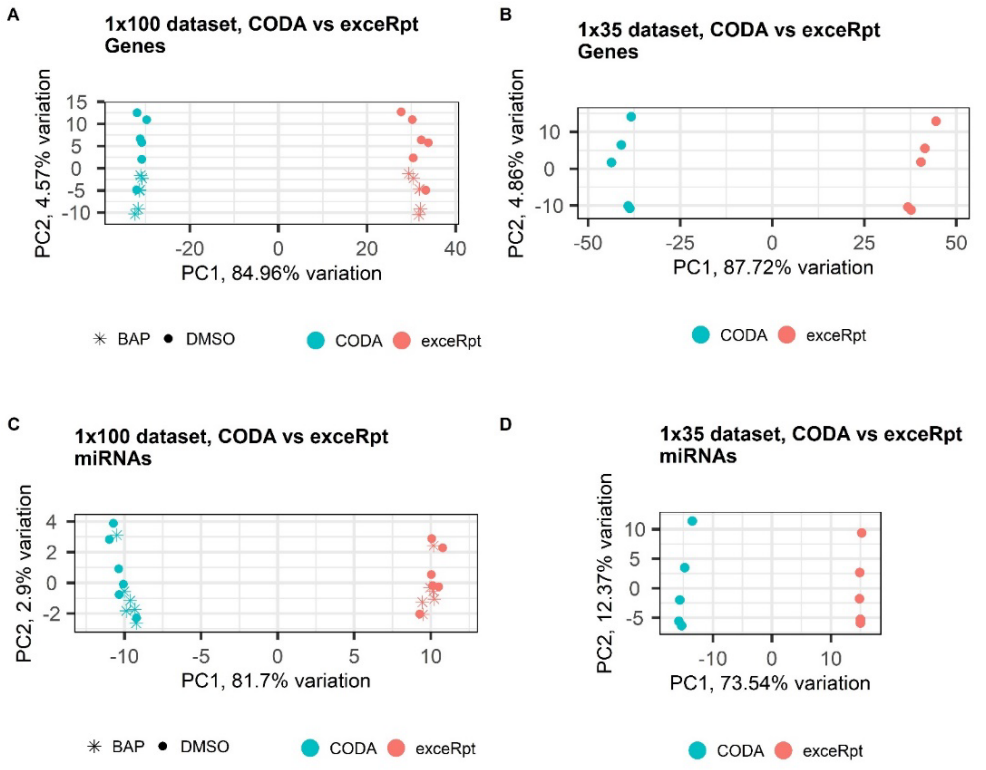


Figure 7. Principal component analysis (PCA) plots showing PC1 and PC2 of PCA analysis carried out on variance-stabilized normalized gene counts for (A) 1x100 and (B) 1x35 samples processed with either pipeline. Plots showing PC1 and PC2 of PCA analysis carried out on variance-stabilized transformed miRNA counts for (C) 1x100 and (D) 1x35 samples processed with either pipeline (cyan = CODA, red = exceRpt).

Gene counts correlation is stronger among samples analyzed with the same pipeline, which remains relatively high across the two methods for the 1x100 dataset (Figure 8A). In the 1x35 dataset, the correlation among biological replicates analyzed with the same pipeline is also high but shows a lower value across methods (Figure 8B). The relatively higher correlation between CODA and exceRpt in the 1x100 dataset compared to the 1x35 one could be explained by the fact that the median total gene read count is very similar (from 39.36 M with exceRpt to 40.07 M reads with CODA, with an increase of around 0.71 M reads). For the 1x35 dataset, instead, there is a gain of ~8.68 M reads per sample (around +17%, from a median 52.01 M with exceRpt to 60.69 M with CODA).

Correlation of normalized miRNA counts for both datasets is higher among samples analyzed with the same pipeline and between the same sample analyzed with CODA or exceRpt (Figure 8C-D).

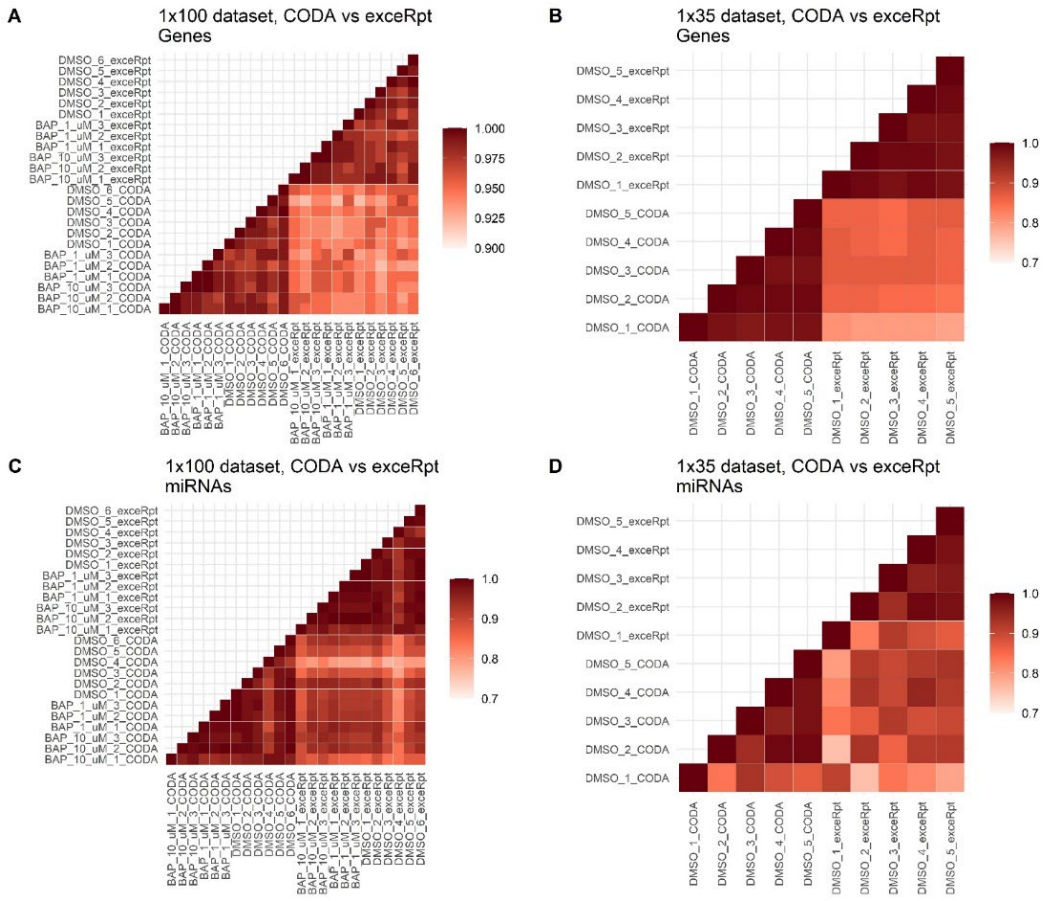


Figure 8. Pearson correlation of normalized gene counts for (A) 1x100 and (B) 1x35 samples. Pearson correlation of normalized miRNA counts for (C) 1x100 and (D) 1x35 samples.

To investigate the correlation difference among samples in the 1x100 dataset compared to the one in the 1x35 dataset, we analyzed the RNA biotype composition of the Nthy-ori 3-1 and follicles DMSO control samples. To reduce the background noise and highlight the most consistent differences, we focused on biotypes representing at least 1% of total mapping reads on average (Figure 9A-B, Table 6). Reads mapping to lncRNA and snoRNA are mainly between 60 nt and 100 nt long and are thus identified in the 1x100 dataset samples by both pipelines. Mitochondrial rRNA (Mt rRNA) reads show two peaks, at 89 nt and 91 nt, but the 91 nt peak

is not identified by exceRpt (Figure 9C). snoRNA are almost completely missed by exceRpt. Possibly, these reads and other biotypes on average longer than 44 nt do not have a complete 3' adapter and are thus discarded by exceRpt. By recovering longer reads with an incomplete adapter, CODA also recovers protein coding reads, which constitute between 40% and 80% of the reads with an incomplete or partial adapter (Figure 9D).

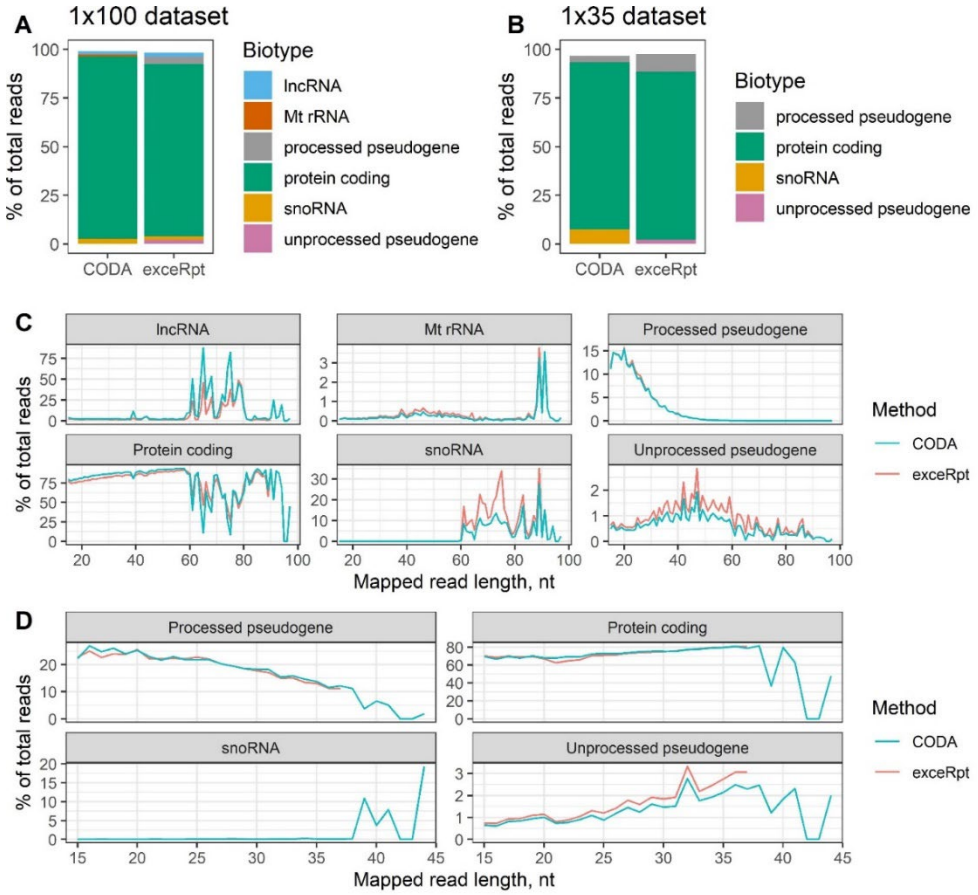


Figure 9. Average biotype composition of (A) 1x100 (6 replicates) and (B) 1x35 (5 replicates) DMSO control samples processed with either CODA or exceRpt. The values are expressed as percentage of total gene read counts. Read length distribution for (C) 1x100 and (D) 1x35 datasets expressed as percentage of total mapped reads grouped per biotype. Only the biotypes representing at least 1% of total reads on average are reported.

Table 6. Average percentage of RNA biotypes relative to the total gene read count for DMSO replicates of 100- or 1x35 datasets processed with CODA or exceRpt. We selected only the biotypes that make up for at least 1% of the total read count in at least one replicate.

	1x100 dataset		1x35 dataset	
	CODA	exceRpt	CODA	exceRpt
lncRNA	1.70	2.02	0.69	0.76
Mt rRNA	1.10	0.96	0.49	0.47
Processed pseudogene	0.47	3.80	3.39	9.18
Protein coding	93.67	88.58	85.89	86.32
snoRNA	2.66	1.84	7.40	0.12
Unprocessed pseudogene	0.12	2.04	0.87	2.12

lncRNA = long non-coding RNA; Mt = mitochondrial; snoRNA = small nucleolar RNA

Differential expression analysis was performed to evaluate the number of differentially expressed (DE) genes and miRNA after BAP exposure of Nthy-ori 3-1 cells (1x100 dataset). MA-plots of the dataset processed with either workflow showed comparable distributions for the genes (Figure 10A-B), while highly expressed miRNAs are characterized by a higher log2 fold change when analyzed by CODA (Figure 10C-D). In addition, the two pipelines identify a comparable number of DE genes and miRNA (CODA: 1,201 DE genes, 1 DE miRNA; exceRpt: 1,251 DE genes, 0 DE miRNA) (FDR = 0.01) (Figure 10E-F) but only a partial overlap between genes (Figure 10G).

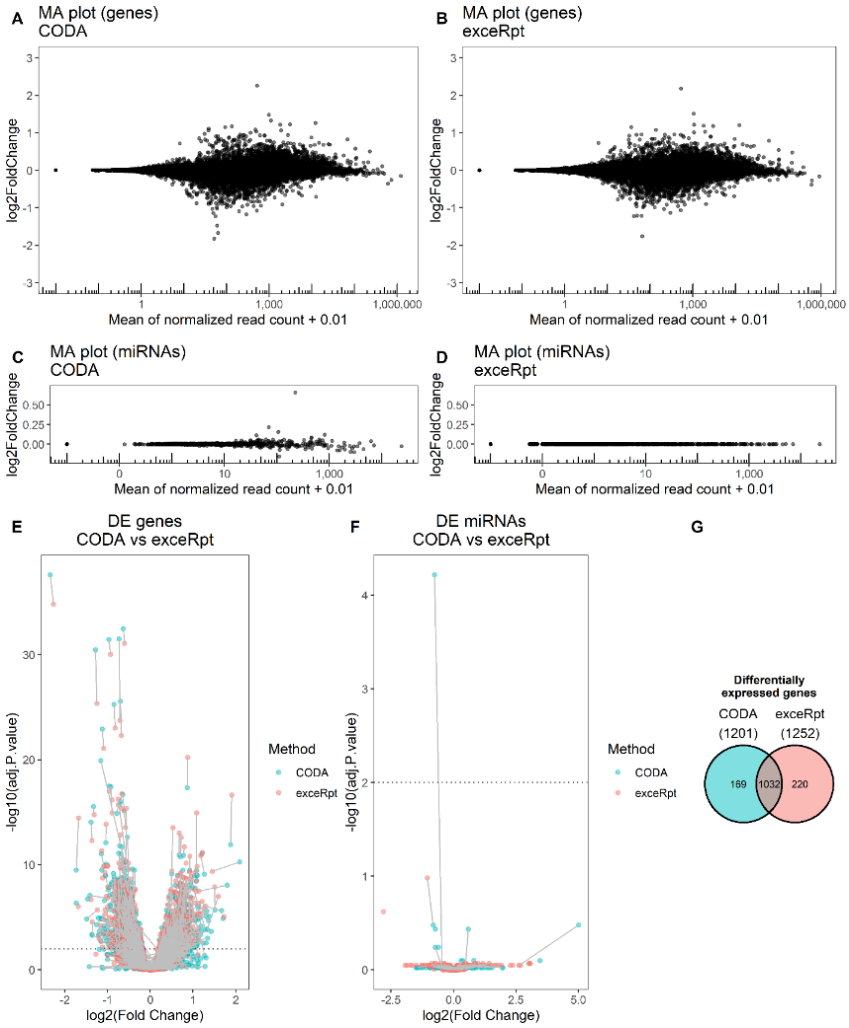


Figure 10. (A, C) MA-plot of the 1x100 dataset samples processed with CODA (A: genes; C: miRNAs). (B, D) MA-plot of 1x100 dataset samples processed with exceRpt (B: genes; D: miRNAs). For plotting purposes, 0.01 has been the mean read count to account for genes/miRNAs whose expression is 0. The ashr method has been used as shrinkage estimator [61]. Volcano plot of (E) genes and (F) miRNAs that pass the CPM filter in the 1x100 dataset processed either with CODA (cyan) or exceRpt (red) and that have been included in the differential expression analysis carried out with DESeq2 comparing samples exposed to BAP to the DMSO control. Every dot represents a gene/miRNA, and the same gene/miRNA identified in both pipelines is connected by a grey line. The dotted line represents the FDR = 0.01. (G) Venn diagram of the overlap between the differentially expressed (DE) genes identified in the 1x100 dataset processed either with CODA (cyan) or exceRpt (red). The total number of DE genes identified by either pipeline is reported in brackets below the pipeline name.

Gene ontology (GO) and Reactome analysis shows how the P-adjusted values for the same pathways are mostly lower in CODA (Figure 11A, C) and showing enrichment for more

terms (Figure 11B, D). As BAP is known to be a strong inducer of the cytochrome P450 enzymes CYP1A1 and CYP1B1 via the activation of the aryl hydrocarbon receptor (AHR) (50), we analyzed the expression of these two genes: both result DE at a comparable level (*CYP1A1*: p-adj = 5.2e-11 in CODA, p-adj = 4.2e-10 in exceRpt; *CYP1B1*: p-adj = 1.5e-05 in CODA, p-adj = 2.3e-10 in exceRpt) (Figure 11E-F). Taken together, these results show how differential expression results are comparable between pipelines, with CODA showing higher sensitivity in gene functional enrichment analysis.

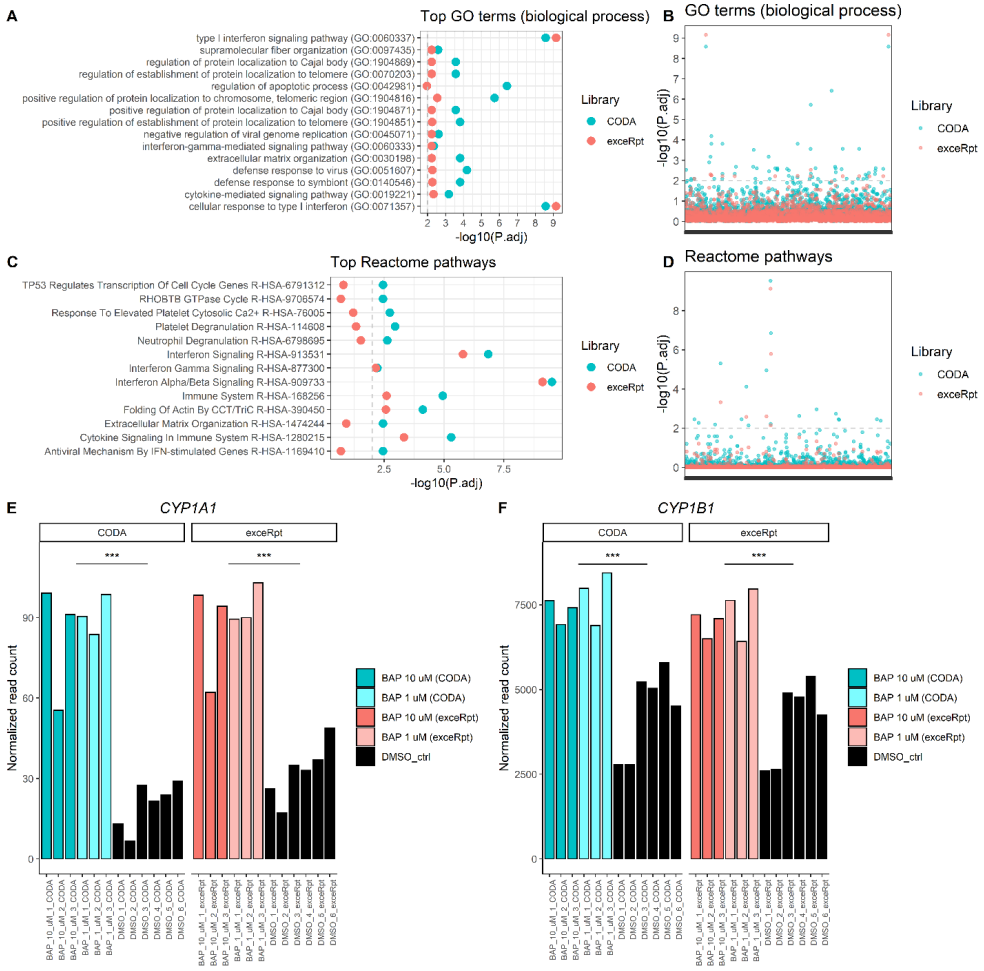


Figure 11. (A-D) Results of gene ontology (GO) (biological process) and Reactome analysis performed on the differentially expressed genes in BAP vs DMSO samples from the 1x100 dataset processed with either CODA (cyan) or exceRpt (red) workflows. In figures (A) and (C) the top 10 terms with lowest P-adjusted value in each group were selected and then plotted together. If two or more terms had the same P-adjusted value, all terms were reported. The dotted grey line corresponds to the set FDR value of 0.01. In figures

(B) and (D) all the terms which resulted from the analysis were reported. Each dot corresponds to one term. The dotted grey line corresponds to the set FDR value of 0.01. (E, F) Normalized expression values of (E) CYP1A1 and (F) CYP1B1 genes as measured in BAP (10 uM dark shade, 1 uM light shade) vs DMSO (black) samples from the 1x100 dataset processed with either CODA (cyan) or exceRpt (red) workflows. The *** symbol denotes a P-adjusted value < 0.001 as resulted from the differential expression analysis.

3.3 Evaluation of Combo-Seq Compared to Poly(A) and Small RNA Libraries

To evaluate the genes and miRNA identified by Combo-Seq libraries, we compared them to conventional poly(A) libraries (for genes) and small RNA libraries (for miRNAs) prepared with the same Nthy-ori 3-1 input RNA. We evaluated the number of expressed genes and miRNA using both libraries and performed differential expression analysis to identify genes and miRNA dysregulated upon BAP treatment.

The median total gene read count was 42.4 M and 35.0 M reads for Combo-Seq and poly(A) libraries, respectively (Figure 12A), while for miRNAs it was 0.13 M (Combo-Seq) and 2.78 M (small RNA libraries) reads (Figure 12B).

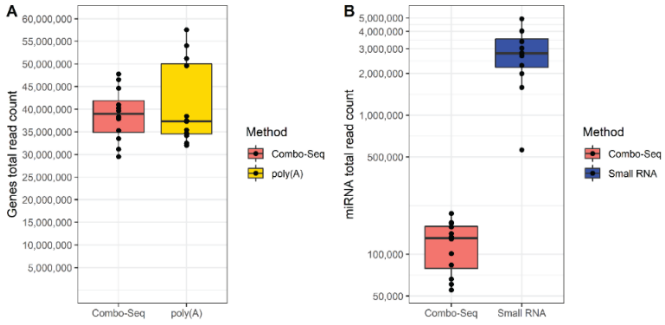


Figure 12. Boxplots of total (non-normalized) count of reads mapped to (A) genes or (B) miRNAs of Nthy-ori 3-1 samples prepared with either Combo-Seq (red), poly(A) (yellow) or small RNA (blue) libraries.

Samples cluster along PC1 based on library preparation method and along PC2 based on treatment both for genes (Figure 13A) and miRNAs (Figure 13B), showing how the type of library is the greatest source of variation. In addition, the greater spread along PC2 for Combo-Seq samples could be attributed to the libraries preparation over different batches, as opposed to the poly(A) and small RNA libraries, which were prepared in single batches, or to the low number of miRNA-mapping reads in Combo-Seq samples, as background noise tends to increase with small sample sizes. Correlation of genes normalized counts is not very high between the two different libraries compared to the correlation within the same methods (Figure 13C). Due to the very different sequencing depths of the two datasets, normalization for library size would

tend to overestimate miRNA count in the samples prepared with Combo-Seq (Figure 14). We then ranked the miRNAs in each sample based on their level of expression (highest read count = highest rank) and calculated Spearman correlation of the ranks across samples. The correlation plot confirms the PCA results, showing a high correlation among samples prepared with the same type of library (Figure 13D).

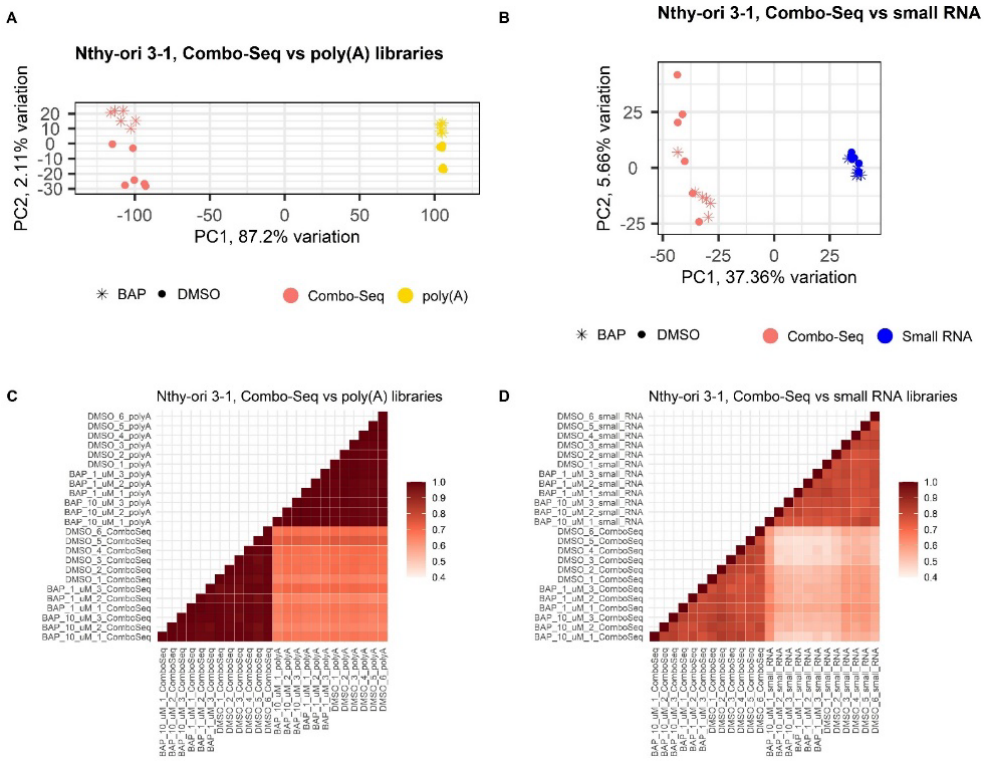


Figure 13. (A) PCA plot of variance-stabilized transformed gene counts of Nthy-ori 3-1 samples prepared using Combo-Seq or poly(A) libraries. (B) PCA plot of variance-stabilized transformed miRNA counts of Nthy-ori 3-1 samples prepared using Combo-Seq or Small RNA libraries. (C) Person correlation of normalized gene counts for Nthy-ori 3-1 samples prepared with Combo-Seq or poly(A) libraries. (D) Spearman correlation of ranked miRNA counts of Nthy-ori 3-1 samples prepared using the Combo-Seq kit a Small RNA kit. Combo-Seq libraries were sequenced on a 1x100 single-end flowcell, poly(A) libraries on a 2x200 paired-end flowcell and small RNA libraries on a 1x35 single-end flowcell.

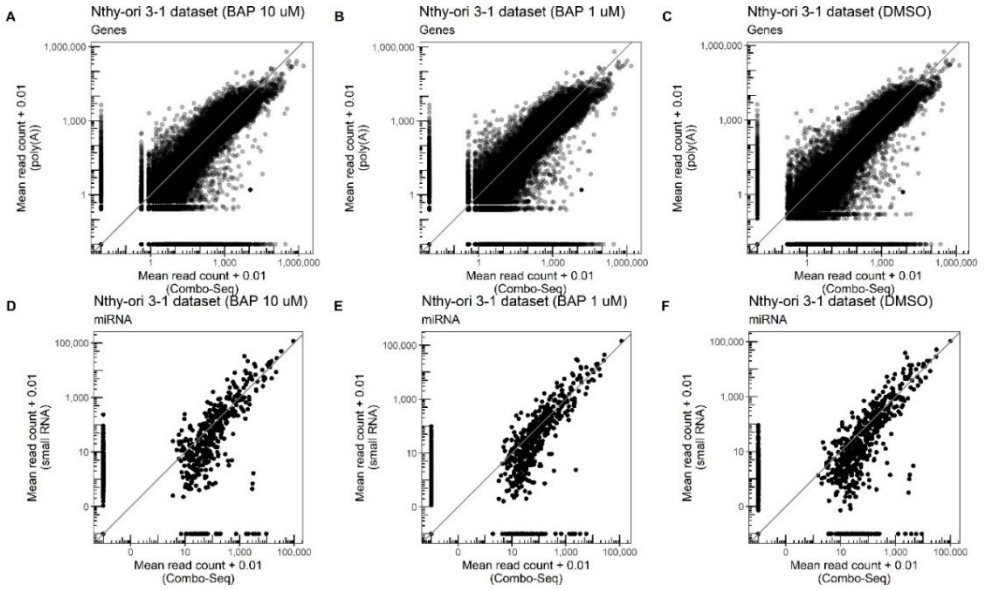


Figure 14. Normalized genes (A to C) or miRNA (D to F) read counts of Nthy-ori 3-1 samples prepared using either Combo-Seq libraries (x-axis) or poly(A) (for genes)/small RNA (for miRNAs) libraries (y-axis). The axes are log10 transformed and 0.01 was added to every value to account for genes or miRNA whose read count was 0. The mean read count was calculated by averaging the normalized count of each group replicates (BAP 10 µM (A and D), BAP 1 µM (B and E), DMSO control (C and F)). Every dot represents a gene or miRNA, and the identity line is reported in grey.

To evaluate how the type of library affects genes and miRNA detection, we compared the transcripts that are thus considered expressed. On average, as many genes are identified in samples prepared with poly(A) as in Combo-Seq libraries (3% more on average) (Figure 15A-C). On the other hand, small RNA libraries identify 1.8 times more miRNAs than Combo-Seq on average, and similarly to genes, most of the ones detected by Combo-Seq overlap with the other library (Figure 15D-F).

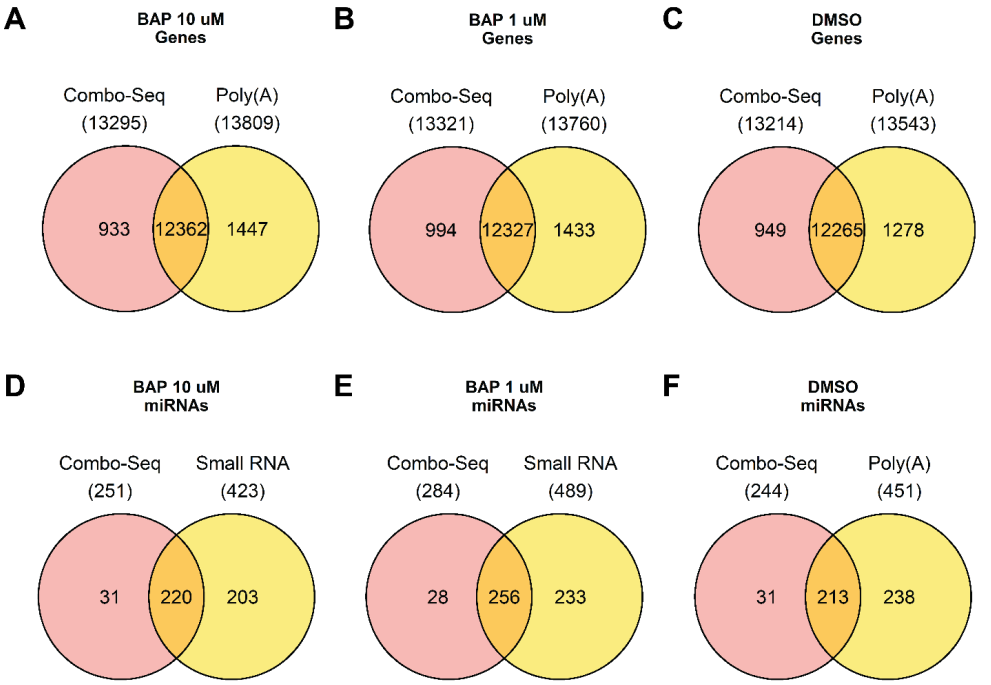


Figure 15. Venn diagram of expressed genes (A-C) and miRNAs (D-F) in Nthy-ori 3-1 samples grouped by treatment (i.e., BAP 10 uM, BAP 1 uM, DMSO control) and prepared with Combo-Seq, poly(A) or small RNA libraries and their overlap. The total number of expressed species is indicated in brackets below the type of library.

Regression analysis of the average normalized read count of expressed genes and miRNAs for each group is not very strong (average R^2 = 0.60 for genes and R^2 = 0.69 for miRNAs) (Figure 16).

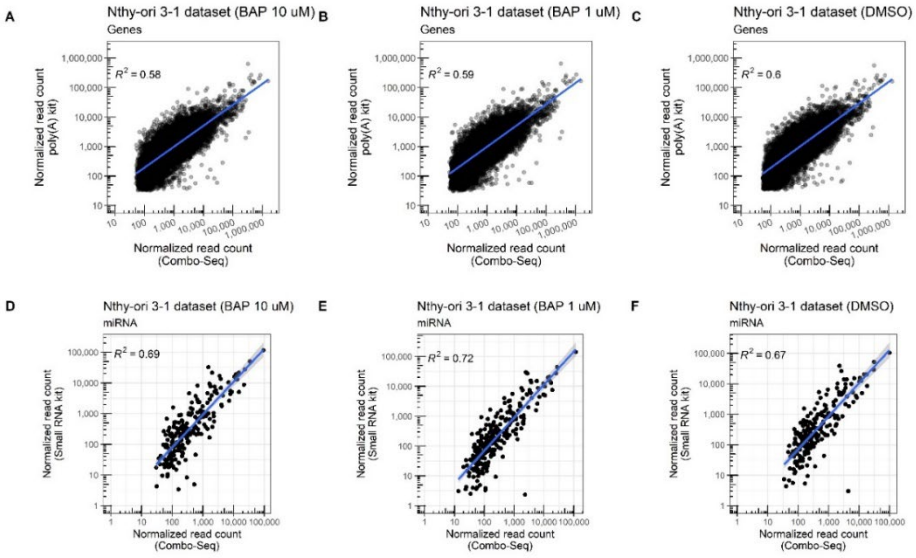


Figure 16. Scatterplots of normalized genes (A to c) or miRNA (D to F) counts passing the relevance filter in Nthy-ori 3-1 samples prepared with Combo-Seq (x-axis) or poly(A)/small RNA (y-axis) library prep kits ((A, D) BAP 10 uM, (B, E) BAP 1 uM, (C, F) DMSO control). Each dot represents a gene or miRNA, and only species that are considered expressed (i.e. passing the relevance filter) by both methods in each condition are reported. A linear regression model was computed for each plot and the R2 value is reported in the top left corner of the plot. The x and y axes are log10 transformed.

Next, to identify differentially expressed genes and miRNA in the BAP-treated samples compared to the DMSO control we performed differential expression analysis. When the samples are prepared with poly(A) libraries or Combo-Seq libraries, 4,462 or 1,186 genes result DE, respectively, 967 of which overlap between the two methods (Figure 17A). GO analysis shows that most top 10 terms are shared by Combo-Seq and poly(A) and are related to the processes of protein localization to telomeres (GO:0070203, GO:1904851, GO:1904816), regulation of apoptosis (GO:0042981, GO:0043065, GO:0043069), extracellular matrix organization (GO:0030198, GO:0097435, GO:0030334, GO:0030335) and regulation of protein localization to Cajal body (GO:1904871, GO:1904869) (Figure 17B). Reactome enrichment analysis shows similar results, where on average the top hits tend to have a lower P-adjusted value in the poly(A) dataset compared to Combo-Seq (Figure 17C).

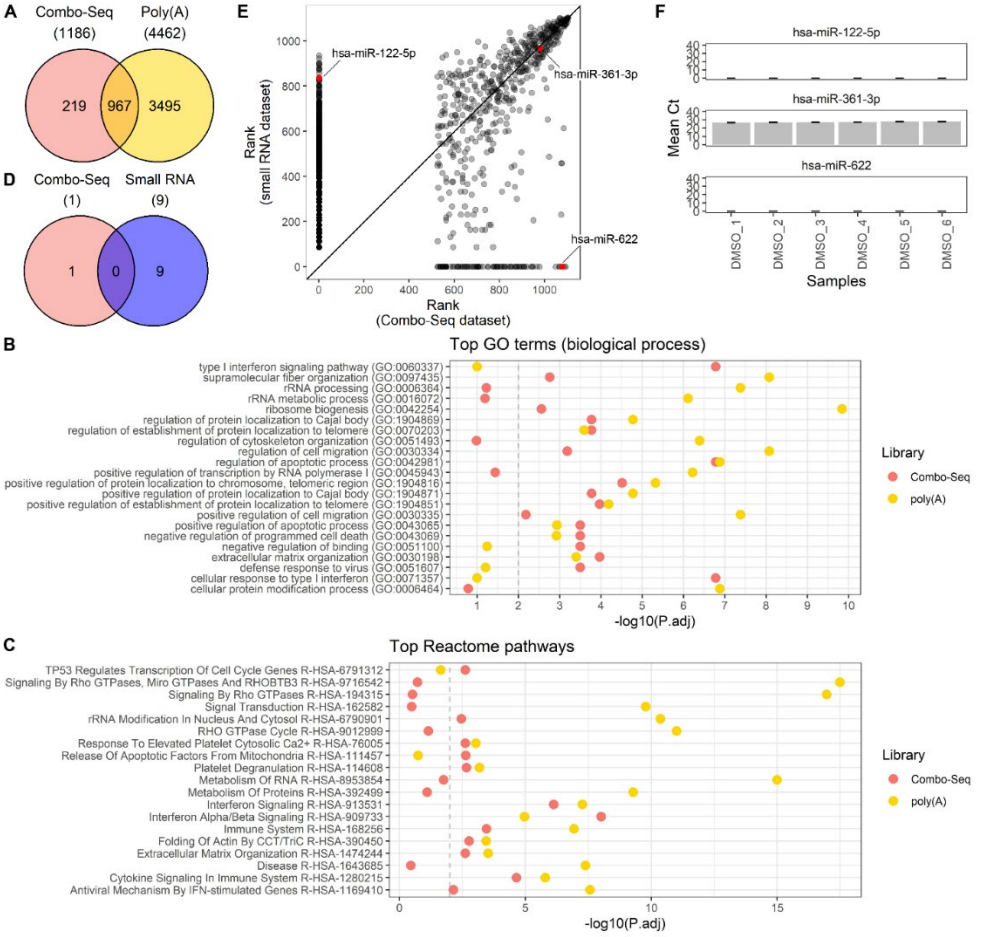


Figure 17. (A) Overlap of differentially expressed genes after BAP treatment compared to the DMSO control in datasets prepared with Combo-Seq libraries (red) or poly(A) (yellow) libraries. (B) Results of gene ontology (GO) (biological process) and (C) Reactome pathway analyses performed on the differentially expressed genes in BAP vs DMSO samples prepared with either Combo-Seq (red) or poly(A) (yellow) libraries. The top 10 GO terms with lowest P-adjusted value in each group were selected and then plotted together. If two or more terms had the same P-adjusted value, all terms were reported. The dotted grey line corresponds to the set FDR value of 0.01. (D) Overlap of differentially expressed miRNA after BAP treatment compared to the DMSO control in datasets prepared with Combo-Seq libraries (red) or small RNA (blue) libraries. (E) Scatterplot representing the rankings of miRNA in mean read count of Nthy-ori 3-1 DMSO control samples. The mean read count was calculated as the average of the replicate samples prepared with either a Combo-Seq or small RNA library prep kit. miRNAs were then ranked based on their level of expression in each condition (most highly expressed miRNA = highest rank). Each dot in the plot represents a miRNA, and miRNAs for which the mean count was 0 in both conditions were removed. A total of 1104 miRNAs were ranked and miRNAs with the same level of expression were assigned the same rank. The miRNA selected for qPCR validation are highlighted in red. (F) RT-qPCR analysis of the selected miRNAs. The bar represents the average Ct value for each sample and the error bars represent

the mean ± sd. Each sample was measured in 4 technical replicates. (Modified from the manuscript this chapter is based on.)

Globally, a greater enrichment in both GO (biological pathway) (Figure 18A) and Reactome terms results from poly(A) samples (Figure 18B).

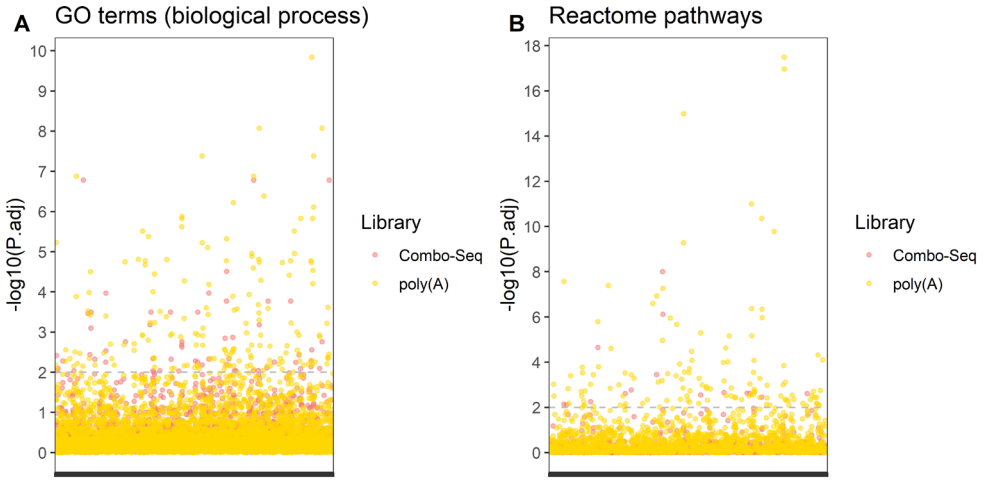


Figure 18. Results of (A) gene ontology (GO) (biological process) and (B) Reactome analyses performed on the differentially expressed genes in BAP vs DMSO samples from the Nthy-ori 3-1 dataset prepared either with the Combo-Seq library prep kit (red) or poly(A) library prep kit (gold). All the terms which resulted from the analyses were reported. Each dot corresponds to one term. The dotted grey line corresponds to the set FDR value of 0.01.

BAP is a genotoxic compound able to induce apoptosis. For this reason, we analyzed the expression of genes known to be induced by TP53, by BAP treatment, or labelled as pro-apoptotic (Table 7). It must be pointed out that neither of the BAP concentrations tested resulted cytotoxic on Nthy-ori 3-1 cells after treatment for up to 72 hours (data not shown). A partial overlap of the dysregulated genes can be observed (up: *BMF*, *CDKN1A*, *CYP1A1*, *CYP1B1*, *FAS*, *MDM2*; down: *BNIP3*, *BOK*, *GADD45A*), while some genes result dysregulated in either dataset (*BAK1*, *BCL2L11*, *BID*, *DDB2*, *RRMB2*).

Table 7. List of genes that are induced by BAP, involved in apoptosis, or induced by TP53. The data refers to the differential expression analysis carried out in the Nthy-ori 3-1 samples prepared either with Combo-Seq or poly(A) library kit. If the gene results upregulated, it is reported in red, if downregulated, in blue. The name of the library is reported, as well as the Ensembl gene ID and a brief description of the protein coded by the gene.

Combo-Seq	Poly(A)	Protein function
BAD	*BAD*	Proapoptotic member of the BCL-2 family

Combo-Seq	Poly(A)	Protein function
BAK1	*BAK1*	Proapoptotic member of the BCL-2 family
BAX	*BAX*	Proapoptotic member of the BCL-2 family
BBC3	*BBC3*	Proapoptotic member of the BCL-2 family
BCL2L11	*BCL2L11*	Proapoptotic member of the BCL-2 family
BID	*BID*	Proapoptotic member of the BCL-2 family
BIK	*BIK*	Proapoptotic member of the BCL-2 family
BMF	*BMF*	Proapoptotic member of the BCL-2 family
BNIP3	*BNIP3*	Proapoptotic member of the BCL-2 family
BOK	*BOK*	Proapoptotic member of the BCL-2 family
CDKN1A	*CDKN1A*	Inhibitor of cyclin-dependent kinase 2 and 4 complexes. Regulated by p53.
CYP1A1	*CYP1A1*	Member of cytochrome P450 family. Induced by the AHR after binding by BAP.
CYP1B1	*CYP1B1*	Member of cytochrome P450 family. Induced by the AHR after binding by BAP.
DDB2	*DDB2*	Involved in DNA repair. Regulated by p53.
FAS	*FAS*	Member of TNF-receptor superfamily. Necessary for the formation of the death-inducing signaling complex (DISC), involved in apoptosis. Regulated by p53.
GADD45A	*GADD45A*	Involved in DNA repair mechanism. Regulated by p53.
MDM2	*MDM2*	Oncogene. Codes for a nuclear-localized E3 ubiquitin ligase. It targets tumor suppressor proteins (like p53) for proteasomal degradation. Regulated by p53.
RRMB2	*RRMB2*	Necessary for DNA synthesis. Regulated by p53.

Using Combo-Seq samples processed with CODA, we identify only 1 DE miRNA, while we identify 9 DE miRNA from small RNA libraries samples. Interestingly, there is no overlap between the DE miRNA in the Combo-Seq and small RNA groups (Figure 17C). In addition, the only DE miRNA in the Combo-Seq group (hsa-miR-3654) is not expressed in the small RNA one. Vice versa, only 6 out of 9 DE miRNA (hsa-miR-1268a/1268b, hsa-miR-186-5p, hsa-miR-222-3p, hsa-miR-30a-3p, hsa-miR-30c-2-3p, hsa-miR-92a-1-5p) are expressed in the Combo-Seq group. To discern which of the two libraries most truthfully detects the miRNAs in our samples, we validated the expression of three miRNAs using RT-qPCR in the DMSO control samples. We selected miRNAs for which the two datasets disagree in either direction (i.e. the rank is high in one dataset a low in the other) or are in concordance (Figure 17D): we selected hsa-miR-622, hsa-miR-122-5p and hsa-miR-361-3p (Figure 19). RT-qPCR shows how hsa-miR-361-3p is the only miRNA detected in the DMSO control samples, at consistent levels among biological replicates. On the other hand, hsa-miR-122-5p and hsa-miR-622 are not detected (Figure 17E).

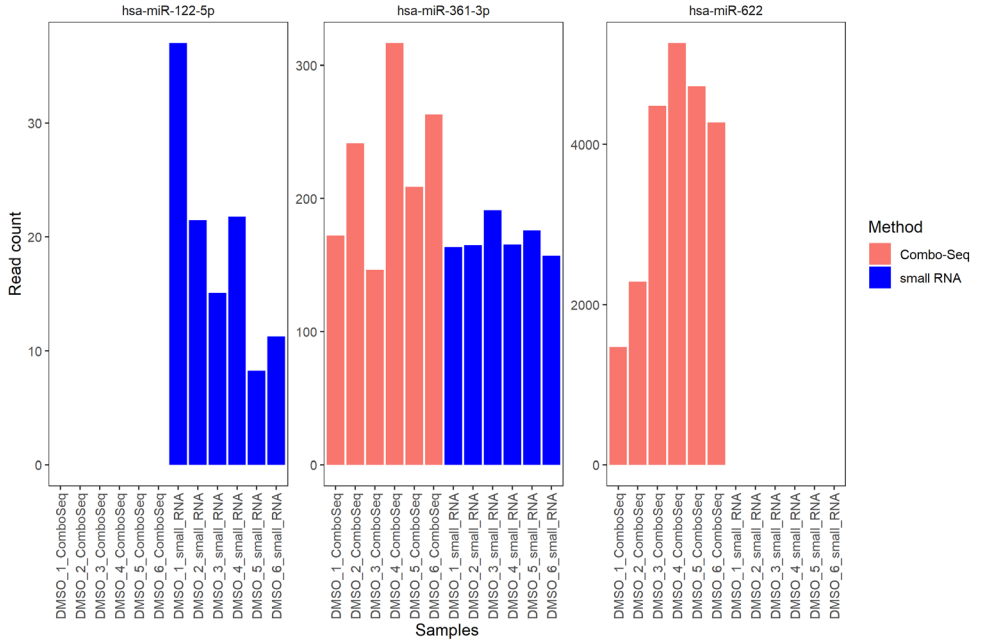


Figure 19. Normalized read count of the miRNAs selected for qPCR evaluation in Nthy-ori 3-1 DMSO samples prepared with Combo-Seq (red) or small RNA (blue) library prep kit.

Intriguingly, hsa-miR-622 is coded within the keratin 18 pseudogene 27 (KRT18P27) (Figure 20A). It is possible that a fragment of KRT18P27 mRNA generated early in the protocol of Combo-Seq library preparation, when poly(A) species are retrotranscribed and the RNA-cDNA hybrid is fragmented by RnaseH, is then erroneously recognized as a miRNA. Analysis of the isomiRs of hsa-miR622 reveals how there is a wide distribution of isomiRs across the six DMSO replicates detected at a low level (Figure 20B). Interestingly, hsa-miR-622 read count calculated by exceRpt is comparable to the one calculated by CODA, supporting the hypothesis that this is not a consequence of a mature miRNA processing, but that at least part of the fragments derives from KRT18P27 fragmentation (Figure 20C).

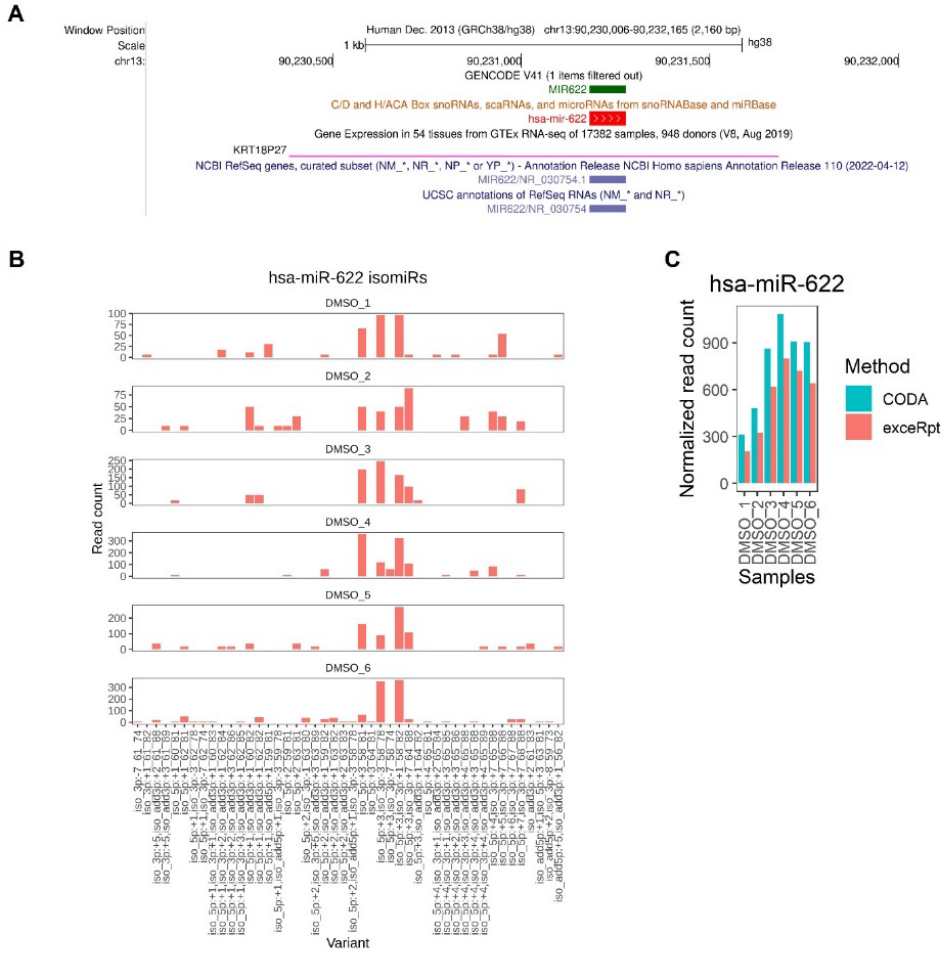


Figure 20. (A) Screenshot of the UCSC Genome Browser showing the location of hsa-miR-622 and KRT18P27 along the genomic coordinates (database accessed on 31/08/2022). (B) Raw isomiRs read count of hsa-miR-622 in Nthy-ori 3-1 DMSO samples prepared with Combo-Seq. On the x axis is reported the type of isomiR and its coordinates. (C) Normalized read count of hsa-miR-622 in Nthy-ori 3-1 DMSO control samples prepared with Combo-Seq as identified by CODA (cyan) and exceRpt (red).

4. Discussion

Combo-Seq is a library prep kit for RNA-Seq that allows to prepare combined mRNA-miRNA libraries starting from the same sample with very little minimum input (down to 5 ng of total RNA). For these reasons, it represents a convenient solution for simultaneously analyzing both RNA species from a single sample, even for samples that contain little RNA, such as biopsies, extracellular fluids, or organoids. In addition, it provides useful information about the relative mRNA and miRNA content of a cell, which to our knowledge cannot be provided by any library preparation kit currently available in the market. However, no specific bioinformatic pipeline was developed for the processing of this data. To this purpose, the manufacturer recommends using the exceRpt pipeline with some modifications (19). Nonetheless, it presents some limitations when adopted for the processing of Combo-Seq data.

In this paper, we illustrated CODA, a pipeline we developed for the processing of Combo-Seq data. It is modular and implements free-to-use tools often employed in RNA-Seq processing analysis. The first step is adapter trimming with Cutadapt, which we chose because its manual clearly states that it can handle partial adapters, a key point especially critical for shorter sequencing reads. In addition, Cutadapt is regularly supported and updated, and offers a clear and extensive documentation. After trimming, our pipeline performs separate alignment and quantification of miRNAs and genes: miRNA detection is carried out with miRge3.0. Gene mapping and quantification is done with RSEM based on the ENCODE3's STAR-RSEM pipeline. It is important to note that, although we selected certain tools, the strength point of CODA is the control that the user has over each step that is carried out and it is possible to change each tool according to the user's preferences. CODA can then be considered a guideline on how to analyze sequencing data deriving from Combo-Seq libraries, and the end user is free to use it or set up their own.

To compare CODA to exceRpt, we generated Combo-Seq libraries from two different cell models and compared the processing of the two pipelines. We showed that, because of the chosen trimmer, the maximum read length of trimmed reads when using CODA is higher than the one with exceRpt, and it results in more reads successfully passing. This is more dramatic the shorter the sequenced reads are. This tends to affect gene-mapping reads, rather than miRNA mapping ones: in fact, when the same samples are processed with CODA, the absolute number of reads mapping to genes increases, especially for shorter sequencing reads, where the proportion of reads with an incomplete/missing adapter increases. On the other hand, the number of reads mapped to miRNAs is almost the same. The two pipelines are comparable at the mapping stages, showing similar percentages of reads passing trimming that are aligned to genes or miRNA, possibly because they both use STAR as mapper for genes, while miRNA mapping, being more stringent and less ambiguous, can be performed by different aligners with similar results. In addition, more genes are assigned reads when samples are processed with exceRpt rather than

CODA. This difference may be due to the quantification step: while exceRpt adopts its own quantification algorithm (20), CODA uses RSEM, which employs an Expectation-Maximization (EM) algorithm in its statistical model, whereby assignment of multimapping reads is determined by estimating the level of expression of deriving from unambiguously mapping reads (29). We hypothesize that while exceRpt attributes every read to its highest scoring location, RSEM instead allocates reads mapping to very low expressed genes to higher expressed paralogues.

We also observed that the read length distribution in Combo-Seq libraries is not homogeneous for all RNAs: some biotypes tend to generate fragments longer than the Combo-Seq average (which is 21-22 nt (51)). As such, most reads coming from these species will have an incomplete/partial adapter when sequenced on a low number of cycles (e.g., using a 1x35 flowcell like we did). If the pipeline used to process the data cannot retain these reads, they will be lost. This can lead to an incorrect estimation of the RNA biotype composition of a sample, and loss of potentially interesting data (52, 53). Although two different RNA extraction kits were used during the Nthy-ori 3-1 and thyroid follicles processing, the generated RNA-Seq data is comparable (54, 55). We believe then the results obtained by the analysis of the 1x100 and 1x35 datasets have been minimally influenced by this factor.

We also compared how Combo-Seq libraries perform in comparison to standard poly(A) and small RNA libraries for the analysis of mRNA or miRNA, respectively. We showed how the type of library is the main source of variation when comparing the two datasets. We noticed that conventional poly(A) libraries identify around 4% more genes, while small RNA libraries identify almost twice as many miRNAs. The difference in genes identified is most likely due to the chemistry underlying the library preparation kit, as very different RNA inputs from the same sample prepared with the same kit show very similar percentages of exonic-, intronic- and intergenic-mapping reads (56). In addition, for this work, the poly(A) libraries were prepared in a single batch in an automated system. The Combo-Seq libraries were prepared manually over different batches and the protocol includes several steps. It is then possible that the variation introduced during the Combo-Seq libraries preparation reflects in a variability in gene expression among biological replicates, which would affect the DE analysis. Considering the miRNAs, most of the variability in detection probably arises from the difference in total miRNA read counts. Indeed, the number of recovered miRNAs from Combo-seq is dependent on the miRNA content of the cells, which seems to be low in the selected Nthy-ori 3-1 cell line. The sequencing depth would therefore need to be much higher to reach a number of reads comparable to the small RNA libraries. Admittedly, the input RNA used for library preparation is a possible confounder, as we did not test how the number of miRNA mapping reads changes with different amounts of input, and using a lower amount, which increases the miRNA/mRNA ratio in Combo-Seq libraries (57), could result in a greater number of miRNA mapping reads and thus in more

miRNAs identified. Small RNA libraries prepared with a high RNA input perform similarly when detecting highly expressed miRNAs compared to using a low input, and detect more low-expressed ones (58).

Differential expression analysis shows how four times more DE genes are identified in poly(A) libraries, but GO top hits are almost the same. 9 versus 1 DE miRNAs result also from the analysis in small RNA and Combo-Seq libraries, respectively. While we observed an overlap in the DE genes between the two methods, we did not get a similar result for the DE miRNAs. In addition, miRNA validation by RT-qPCR is concordant with small RNA libraries: hsa-miR-622 was detected by Combo-Seq only at high levels, but its expression was absent in RT-qPCR. We hypothesize that at least part of the reads assigned to hsa-miR-622 in Combo-Seq samples may instead derive from the fragmentation of the KRT18P27 transcript. hsa-miR-122-5p was detected at low levels and only by small RNA libraries, which as already discussed have a greater read coverage.

In conclusion, Combo-Seq is a convenient solution to capture both poly(A)-tailed and small RNAs starting from very little material and from a single RNA aliquot. In addition, it requires less time and money per sample than the combination of conventional separated poly(A) and small RNA libraries. However, it presents some inconsistencies when compared to standard poly(A) and small RNA libraries, that researchers should be aware of and evaluate when choosing how to prepare their samples.

5. Data availability

The data underlying this article is available in BioStudies with accession E-MTAB-12078.

6. Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825745.

7. References

1. Boivin V, Faucher-Giguere L, Scott M, Abou-Elela S. The cellular landscape of mid-size noncoding RNA. Wiley Interdiscip Rev RNA. 2019;10(4):e1530, doi:10.1002/wrna.1530.
2. Godoy PM, Bhakta NR, Barczak AJ, Cakmak H, Fisher S, MacKenzie TC, et al. Large Differences in Small RNA Composition Between Human Biofluids. Cell Rep. 2018;25(5):1346-58, doi:10.1016/j.celrep.2018.10.014.
3. Potemkin N, Cawood SMF, Treece J, Guevremont D, Rand CJ, McLean C, et al. A method for simultaneous detection of small and long RNA biotypes by ribodepleted RNA-Seq. Sci Rep. 2022;12(1):621, doi:https://doi.org/10.1038/s41598-021-04209-4.
4. Nolte-'t Hoen EN, Buermans HP, Waasdorp M, Stoorvogel W, Wauben MH, t Hoen PA. Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. Nucleic Acids Res. 2012;40(18):9272-85, doi:10.1093/nar/gks658.
5. Boivin V, Deschamps-Francoeur G, Couture S, Nottingham RM, Bouchard-Bourelle P, Lambowitz AM, et al. Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. RNA. 2018;24(7):950-65, doi:10.1261/rna.064493.117.
6. Stoicea N, Du A, Lakis DC, Tipton C, Arias-Morales CE, Bergese SD. The MiRNA Journey from Theory to Practice as a CNS Biomarker. Front Genet. 2016;7:11, doi:10.3389/fgene.2016.00011.
7. Wittmann J, Jack HM. Serum microRNAs as powerful cancer biomarkers. Biochim Biophys Acta. 2010;1806(2):200-7, doi:10.1016/j.bbcan.2010.07.002.
8. Backes C, Meese E, Keller A. Specific miRNA Disease Biomarkers in Blood, Serum and Plasma: Challenges and Prospects. Mol Diagn Ther. 2016;20(6):509-18, doi:10.1007/s40291-016-0221-4.
9. Scholer N, Langer C, Dohner H, Buske C, Kuchenbauer F. Serum microRNAs as a novel class of biomarkers: a comprehensive review of the literature. Exp Hematol. 2010;38(12):1126-30, doi:10.1016/j.exphem.2010.10.004.
10. da Silva JL, Cardoso Nunes NC, Izetti P, de Mesquita GG, de Melo AC. Triple negative breast cancer: A thorough review of biomarkers. Crit Rev Oncol Hematol. 2020;145:102855, doi:10.1016/j.critrevonc.2019.102855.
11. Arantes L, De Carvalho AC, Melendez ME, Lopes Carvalho A. Serum, plasma and saliva biomarkers for head and neck cancer. Expert Rev Mol Diagn. 2018;18(1):85-112, doi:10.1080/14737159.2017.1404906.
12. Anvar SY, Allard G, Tseng E, Sheynkman GM, de Klerk E, Vermaat M, et al. Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. Genome Biol. 2018;19(1):46, doi:10.1186/s13059-018-1418-0.
13. Ramberg S, Hoyheim B, Ostbye TK, Andreassen R. A de novo Full-Length mRNA Transcriptome Generated From Hybrid-Corrected PacBio Long-Reads Improves the Transcript Annotation and Identifies Thousands of Novel Splice Variants in Atlantic Salmon. Front Genet. 2021;12:656334, doi:10.3389/fgene.2021.656334.

14. Liu D, Graber JH. Quantitative comparison of EST libraries requires compensation for systematic biases in cDNA generation. BMC Bioinformatics. 2006;7:77, doi:10.1186/1471-2105-7-77.
15. Eddy SR. Non-coding RNA genes and the modern RNA world. Nat Rev Genet. 2001;2(12):919-29, doi:10.1038/35103511.
16. Illumina. How short inserts affect sequencing performance 2020 [Available from: https://support.illumina.com/bulletins/2020/12/how-short-inserts-affect-sequencing-performance.html.
17. Verheijen MCT, Krauskopf J, Caiment F, Nazaruk M, Wen QF, van Herwijnen MHM, et al. iPSC-derived cortical neurons to study sporadic Alzheimer disease: A transcriptome comparison with post-mortem brain samples. Toxicol Lett. 2022;356:89-99, doi:10.1016/j.toxlet.2021.12.009.
18. Illumina. Introducing the NovaSeq™ 6000 v1.5 reagents 2020 [Available from: https://support.illumina.com/bulletins/2020/11/introducing-the-novaseq--6000-v1-5-reagents.html.
19. PerkinElmer Inc. NEXTFLEX® Combo-Seq Analysis Guidelines 2020 [Available from: https://perkinelmer-appliedgenomics.com/wp-content/uploads/2020/06/NOVA-5139-AG_v01_NEXTFLEX-Combo-seq-Analysis-Guideline.pdf.
20. Rozowsky J, Kitchen RR, Park JJ, Galeev TR, Diao J, Warrell J, et al. exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling. Cell Syst. 2019;8(4):352-7 e3, doi:10.1016/j.cels.2019.03.004.
21. Abdelhamid RF, Ogawa K, Beck G, Ikenaka K, Takeuchi E, Yasumizu Y, et al. piRNA/PIWI Protein Complex as a Potential Biomarker in Sporadic Amyotrophic Lateral Sclerosis. Mol Neurobiol. 2022;59(3):1693-705, doi:10.1007/s12035-021-02686-2.
22. Zheng T, Ellinghaus D, Juzenas S, Cossais F, Burmeister G, Mayr G, et al. Genome-wide analysis of 944 133 individuals provides insights into the etiology of haemorrhoidal disease. Gut. 2021, doi:10.1136/gutjnl-2020-323868.
23. Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao XH, Dumitrescu AM, et al. Generation of functional thyroid from embryonic stem cells. Nature. 2012;491(7422):66-71, doi:https://doi.org/10.1038/nature11525.
24. Romitti M, Eski SE, Fonseca BF, Gillotay P, Singh SP, Costagliola S. Single-Cell Trajectory Inference Guided Enhancement of Thyroid Maturation In Vitro Using TGF-Beta Inhibition. Frontiers in Endocrinology. 2021;12, doi:https://doi.org/10.3389/fendo.2021.657195.
25. Mueller O., Lightfoot S., Schroeder A. RNA Integrity Number (RIN) – Standardization of RNA Quality Control January 21, 2016 [Available from: https://www.agilent.com/cs/library/applications/5989-1165EN.pdf.
26. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1): 10-2, doi:https://doi.org/10.14806/ej.17.1.200.
27. Patil AH, Halushka MK. miRge3.0: a comprehensive microRNA and tRF sequencing analysis pipeline. NAR Genom Bioinform. 2021;3(3):lqab068, doi:https://doi.org/10.1093/nargab/lqab068.

28. Li B. rsem-prepare-reference documentation page [Available from: https://deweylab.github.io/RSEM/rsem-prepare-reference.html.
29. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323, doi:https://doi.org/10.1186/1471-2105-12-323.
30. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21, doi:10.1093/bioinformatics/bts635.
31. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47(D1):D766-D73, doi:https://doi.org/10.1093/nar/gky955.
32. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008;36(Database issue):D154-8, doi:https://doi.org/10.1093/nar/gkm952.
33. Bushnell B. BBMap [Available from: sourceforge.net/projects/bbmap/.
34. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010.
35. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8, doi:10.1093/bioinformatics/btw354.
36. PerkinElmer Inc. NEXTflex™ Small RNA Trimming Instructions [Available from: https://perkinelmer-appliedgenomics.com/wp-content/uploads/marketing/NEXTFLEX/miRNA/NEXTflex_Small_RNA_v3_Trimming_Instructions.pdf.
37. Verheijen MC, Meier MJ, Asensio JO, Gant TW, Tong W, Yauk CL, Caiment F. R-ODAF: Omics data analysis framework for regulatory application. Regul Toxicol Pharmacol. 2022;131:105143, doi:https://10.1016/j.yrtph.2022.105143.
38. CEFIC C4 team. Omics Data Analysis Framework for Regulatory application (R-ODAF) 2021 [Available from: https://github.com/R-ODAF/Main.
39. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90, doi:10.1093/bioinformatics/bty560.
40. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184-91, doi:https://doi.org/10.1038/nprot.2009.97
41. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
42. Blighe K, Lun A. PCAtools: PCAtools: Everything Principal Components Analysis. . R package version 2.6.0 ed2021.
43. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550, doi:https://doi.org/10.1186/s13059-014-0550-8.

44. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40, doi:https://doi.org/10.1093/bioinformatics/btp616.
45. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9, doi:https://doi.org/10.1038/75556.
46. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021;49(D1):D325-d34, doi:https://doi.org/10.1093/nar/gkaa1113.
47. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50(D1):D687-D92, doi:https://doi.org/10.1093/nar/gkab1028.
48. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128, doi:10.1186/1471-2105-14-128.
49. Hannon GJ. FASTX-Toolkit 2010 [Available from: http://hannonlab.cshl.edu/fastx_toolkit.
50. Atsdr. Toxicological profile for polycyclic aromatic hydrocarbons. ATSDR Tox Profile. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service; 1995. Report No.: CIS/97/00215. <https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=25>.
51. Allen K, Morris A, Piehl S, Kushiro K, Kothandaraman A. Combined mRNA & microRNA NGS Library Prep Enables a more Complete Characterization of Cell-free RNA. 2018.
52. Liang J, Wen J, Huang Z, Chen XP, Zhang BX, Chu L. Small Nucleolar RNAs: Insight Into Their Function in Cancer. Front Oncol. 2019;9:587, doi:10.3389/fonc.2019.00587.
53. Calvo Sánchez J, Köhn M. Small but Mighty—The Emerging Role of snoRNAs in Hematological Malignancies. Noncoding RNA. 2021;7(4):68, doi:https://dx.doi.org/10.3390%2Fncrna7040068.
54. Marczyk M, Fu C, Lau R, Du L, Trevarton AJ, Sinn BV, et al. The impact of RNA extraction method on accurate RNA sequencing from formalin-fixed paraffin-embedded tissues. BMC Cancer. 2019;19(1):1189, doi:https://doi.org/10.1186/s12885-019-6363-0.
55. Scholes AN, Lewis JA. Comparison of RNA isolation methods on RNA-Seq: implications for differential expression and meta-analyses. BMC Genomics. 2020;21(1):249, doi:https://doi.org/10.1186/s12864-020-6673-2
56. Wang L, Felts SJ, Van Keulen VP, Pease LR, Zhang Y. Exploring the effect of library preparation on RNA sequencing experiments. Genomics. 2019;111(6):1752-9, doi:https://doi.org/10.1016/j.ygeno.2018.11.030.
57. PerkinElmer AG-. NEXTFLEX® Combo-Seq™ mRNA/miRNA Kit (v20.04) [
58. Yeri A, Courtright A, Danielson K, Hutchins E, Alsop E, Carlson E, et al. Evaluation of commercially available small RNASeq library preparation kits using low input RNA. BMC Genomics. 2018;19(1):331, doi:https://doi.org/10.1186/s12864-018-4726-6.

Chapter 3

Investigation of the Effects of Phthalates on *In Vitro* Thyroid Models with RNA-Seq and ATAC-Seq

Marta Nazzari, Mírian Romitti, Duncan Hauser, Daniel J Carvalho, Stefan Giselbrecht, Lorenzo Moroni, Sabine Costagliola, Florian Caiment. **Investigation of the effects of phthalates on *in vitro* thyroid models with RNA-Seq and ATAC-Seq.** *Frontiers in Endocrinology*. 2023; 14, 1200211

Abstract

Phthalates are a class of endocrine disrupting chemicals (EDCs) which have been shown to negatively correlate with thyroid hormone (TH) serum levels in human and to cause a state of hyperactivity in the thyroid. However. their mechanism of action is not well described at the molecular level. We analyzed the response of mouse thyroid organoids to the exposure to a biologically relevant dose range of the phthalates DEHP, DIDP, DINP and DnOP for 24 hours and simultaneously analyzed mRNA and miRNA expression via RNA-Sequencing. Dose-series analysis showed how the expression of several genes increased or decreased at the highest dose tested. As expected with the low dosing scheme, the compounds induced a modest response on the transcriptome, as we identified changes in only mmu-miR-143-3p after DINP treatment and very few differentially expressed genes. No effect was observed on thyroid markers. *Ing5* was consistently upregulated in three out of four conditions compared to control, and we observed a partial overlap among the genes differentially expressed by the treatments. Gene set enrichment analysis (GSEA) showed an enrichment in the treatment samples of the fatty acid metabolism pathway and in the control of pathways related to the receptor signaling and extracellular matrix (ECM) organization. As ING5 is a component of histones H3 and H4 acetylation complexes, we exposed the human thyroid follicular epithelial cell line Nthy-ori 3-1 to DEHP or DINP 1 µM for 5 days and analyzed changes in chromatin accessibility via ATAC-Seq. While we observed a general increase of accessibility compared to the control, we did not identify significant changes in accessibility in the identified regions. Lastly, we showed that despite having only a few differently expressed genes, different analysis methods could be applied to retrieve relevant information on phthalates, showing the potential of *in vitro* thyroid relevant systems analysis EDCs.

1. Introduction

Phthalates are a class of manmade compounds used in the manufacturing industry as solvents or added as plasticizers, mainly to polyvinyl chloride (PVC) or other polymers, to confer flexibility and softness (1, 2). Phthalates are alkyl or dialkyl esters of phthalic acid and their functional groups can be linear, branched, or circular (3). Depending on their size, phthalates are classified into low and high molecular weight (MW) (4, 5). Low MW phthalates include benzyl butyl phthalate (BBP), diethyl phthalate (DEP), di-*iso*-butyl phthalate (DiBP), dimethyl phthalate (DMP) and di-*n*-butyl phthalate (DnBP), while high MW ones comprise bis(2-ethylhexyl) phthalate (DEHP), di(2-propylheptyl) phthalate (DPHP), di-*iso*-decylphthalate (DIDP), di-*iso*-nonylphthalate (DINP), and di-*n*-octylphthalate (DnOP) (6). They are found in common household items, medical devices, construction material and consumer products (6). Since they are not covalently bound to the plastic matrix they are contained in, phthalates can leach or gas out and contaminate either the environment or be ingested via contaminated food (5). Indeed, food constitutes one of the biggest sources of human exposure to phthalates (7). Human biomonitoring studies conducted on the general population in Asia, Europe and North America show a widespread exposure of the general population to phthalates (6, 8-11). Despite this broad use and pervasive environmental presence, they have been recognized as toxic substances both in humans and other organisms (12). Once ingested, they are rapidly metabolized in the digestive tract to their monoester form, which are the species responsible for the phthalates' toxicity. Low MW phthalates metabolites are then excreted through the urine, while high MW metabolites are excreted both via the urine and feces (1). While they do not bioaccumulate, the persistent exposure of the population is cause for concern. Short and medium chain phthalates have been associated with higher toxicity than long chain ones which has led to their banning or restriction in children's toys or teething products (13, 14). Some of the examined compounds have been reported to cause chronic or subchronic toxicity in several organs and systems, namely liver, kidney, immune system, testes, uterus, ovary, central nervous system, and thyroid *in vivo* (2, 15-17). Phthalates can also negatively interfere with the endocrine system and are thus considered endocrine disrupting chemicals (EDCs). They have been shown to interfere with prenatal and postnatal development in animal models (18), with the female and male reproductive systems (19-21) as well as being possibly linked to obesity and type 2 diabetes (22-24).

The thyroid is an endocrine gland positioned in the lower part of the anterior neck and is responsible for the production of the thyroid hormone (TH), whose receptors are expressed throughout the body (25). The TH is essential for normal growth and development and metabolism regulation (26). Its production is mainly regulated by the thyroid-stimulating hormone (TSH), which is secreted by the adenohypophysis. In turn, TSH production is regulated both by circulating TH levels and the thyrotropin-releasing hormone (TRH),

synthesized in the hypothalamus. The main cell type of the thyroid is constituted by thyrocytes, which organize in small hollow spheres called follicles and are responsible for synthesizing the TH, stored in the center of the follicle (the lumen), in a dense matrix termed colloid. The synthesis of the TH starts with active transport of iodine inside the thyrocyte via the sodium iodide symporter (NIS in human). In the follicle lumen, it is covalently bound via oxidation to the tyrosyl (Tyr) residues of the protein thyroglobulin (TG) via the action of the membrane-bound enzyme thyroid peroxidase (TPO) (27, 28). Following TSH stimulation, TG is degraded in the lysosomes, freeing TH, which can be transported outside the thyrocyte.

In the thyroid, phthalate treatment has been shown to have an effect *in vitro* and *in vivo*, causing histological changes, such as reduced follicle size and colloid density, hypertrophy of the Golgi apparatus, increase in number and size of lysosomes and alteration of the TH levels (2, 15, 16, 29, 30). DEHP has been shown to downregulate *Tshr* (Tsh receptor) expression and interfere with the Tsh/Tshr signaling pathway *in vivo* (31, 32). In human, the presence of phthalate metabolites in urine has been observed in association with alterations of TH and TSH serum levels (22, 29). In addition, there is evidence for phthalates altering the methylation status in sperm cells (33) and adrenal glands (34) of the offspring of exposed rats, as well as the expression or activity of histone deacetylases and histone methyltransferases (35, 36).

Over the years, great effort has been made to develop thyroid organoids using both embryonic and induced pluripotent stem cells (37) which can be used for developing thyroid disease models (38) and performing cancer (39), toxicological and drug screening (40). In the context of toxicology, *in vitro* models can offer high throughput capability, mechanistic insight into endocrine disruption and reduce the use of animal testing, in line with the 3Rs principles for animal welfare (Replacement, Reduction and Refinement (41).

In this work, we analyzed our ability to identify alterations induced by phthalate treatment by using two *in vitro* models of thyroid. To this end, we exposed mouse embryonic stem cell (mESC)-derived thyroid follicles (42) to the high MW phthalates DEHP, DIDP, DINP, DnOP for 24 hours and analyzed the transcriptome via RNA-Sequencing (RNA-Seq) using the Combo-Seq library prep kit for simultaneous analysis of mRNA and miRNA expression. Data analysis revealed the upregulation of the gene *Ing5* in three out of four tested compounds (DEHP, DINP, DnOP) compared to the control. ING5 is a component of the histones H3 and H4 acetyltransferase complexes HBO1-JADE, HBO1-BRPF1 and MOZ/MORF (43, 44). To investigate a potential effect of phthalate treatment on the chromatin status on thyroid models *in vitro*, we exposed the human thyroid follicular epithelial cell line Nthy-ori 3-1 to 1 uM of DEHP or DINP for 5 days and analyzed the genome accessibility with ATAC-Seq (***A**ssay for **T**ransposase-**A**ccessible **C**hromatin*). We used maSigPro to analyze gene expression across the dose series and performed gene set enrichment analysis (GSEA) to identify enriched pathways.

2. Materials and methods

2.1 Chemicals information

The following phthalates were used for the experiments described in this paper: bis(2-ethylhexyl) phthalate (DEHP) (CAS 117-81-7, purity 99.8% ± 0.4%) (67261, Sigma-Aldrich, St. Louis, MO, USA), di-iso-nonylphthalate (DINP) (CAS 28553-12-0, ester content >= 99% mixture of C9 isomers) (376663, Sigma-Aldrich), di-iso-decylphthalate (DIDP) (CAS 26761-40-0, purity >= 99.0%) (80135, Supelco, St. Louis, MO, USA) (DINP) and di-n-octylphthalate (DnOP) (CAS 117-84-0, purity >= 99.5%) (D201154, Sigma-Aldrich).

2.2 Thyroid organoids Differentiation and Enrichment

2.2.1 Embryonic stem cells culture and Thyroid Organoids Differentiation

Thyroid follicles were differentiated from the A2Lox.Cre_TRE-Nkx2-1/Pax8_Tg-EGFP mouse ESC as previously described [42,47]. Briefly, cells were initially cultured on gamma-irradiated mouse embryonic fibroblasts (MEFs) feeder using mouse stem cell medium [42,47] and incubated at 37 °C, 5% CO2 and > 95% humidity. For differentiation into thyroid, embryoid bodies (EBs) were generated by hanging drops culture of ESCs (1000 cells per drop) for 4 days. They were then collected and embedded in growth factor restricted Matrigel (354230, Corning). 50 mL Matrigel drops containing around 20 EBs were plated into 12-well plates. Cells were differentiated using a differentiation medium composed of DMEM (31966021, Gibco) supplemented with 15% FBS, vitamin C (50 µg/mL) (A4544, Sigma), nonessential amino acids (0.1 mM) (11140035, Gibco), sodium pyruvate (1 mM) (11360039, Gibco), penicillin and streptomycin (50 U/mL) (15140122, Gibco), 2-mercaptoethanol (0.1 mM) (31350010, Gibco). The differentiation medium was supplemented with 1 µg/mL of doxycycline (D9891-1G, Sigma) for 3 days for *Nkx2-1* and *Pax8* induction, followed by two weeks of maturation by using the differentiation medium containing 8-Br-cAMP (0.3 nM) (B 007-500, BioLog).

2.2.2 Thyroid Follicles Enrichment Protocol

At day 21, after complete thyroid maturation, Matrigel drops containing the thyroid follicles were washed twice with Hanks's balanced salt solution (HBSS, containing calcium and magnesium) (14025050, Gibco) and incubated in a HBSS solution (1 mL per well) containing 10 U/mL of dispase® II (4942078001, Roche) and 125 U/mL of collagenase type IV (Sigma) for 30-45 min at 37 °C. The enzymes were then inactivated by adding 10% FBS. Organoids were

centrifuged at 1200 rpm for 3 minutes. After rinsing twice with HBSS, the follicles were enriched by filtration using 30 mm (to remove single cells) (43-50030, pluriSelect Life Science GmbH) to and 100 mm reverse strainer (to remove big follicles aggregates) (43-50100, pluriSelect Life Science GmbH).

Resuspended follicles were cultured in subsequent experiments in the differentiation medium described above and supplemented with 8-Br-cAMP (10 µM) and TGF-βRI inhibitor SB431542 (10 µM) (1614, Tocris), hereafter termed "supplemented differentiation medium".

2.3 Exposure to Phthalates and RNA-Seq Library Preparation

2.3.1 Exposure to Phthalates

One thousand (1,000) follicles per well were seeded in triplicate in low-adhesion 48-well cell culture plates in supplemented differentiation medium (Supplementary Methods) and 1-10-100 nM-1-10 µM of DEHP, DINP or DIDP or 2-20-200 nM-2-20 µM of DnOP dissolved in DMSO (1029521000, Merck Millipore, Burlington, MA, USA) (final DMSO concentration 0.5%). Of note, the slightly different dose range for DnOP was caused by an unwanted dilution error. We decided to still consider DnOP not differently than the other 3 phthalates in our following data analysis, considering the dose range still maintains the same scaling between each dose and its order of magnitude is comparable to the others. As control, 1000 follicles per well were seeded in supplemented differentiation medium and 0.5% DMSO (n = 5) or medium alone (n = 3). The plated follicles were incubated at 37 °C, 5% CO2 and > 95% humidity for 24 hours.

2.3.2 RNA Isolation

After 24 hours, the follicles were collected, washed once with PBS, and lysed in QIAzol Lysis Reagent (79306, Qiagen). Total RNA was extracted using the miRNAeasy Micro Kit (217084, Qiagen, Venlo, The Netherlands). All samples had a RIN (RNA Integrity Number) of 8 or higher.

2.3.3 RNA-Seq Libraries Preparation

Twenty (20) ng of total RNA were used to prepare RNA-Seq libraries with the NEXTFLEX® Combo-Seq™ mRNA/miRNA Kit (NOVA-5139-53, PerkinElmer, Waltham, MA, USA). To deplete tRNA fragments and Y RNA fragments, the NEXTFLEX® tRNA/YRNA Blocker was used during the library preparation following the manufacturer's instructions.

Fourteen (14) cycles of PCR were performed during the protocol. For some samples, the final library concentration was below the pooling concentration used for sequencing (1.6 nM). In these cases, the library was prepared again performing 16 cycles. For 3 samples (DEHP 10 µM replicate 3; DIDP 1 nM replicate 3; Untreated control replicate 1) there was not enough RNA to repeat the library preparation and could thus not be sequenced. The prepared libraries were sequenced on an S4 Illumina flowcell 35 cycles (v1.5) (Illumina) in single-end mode.

2.4 Exposures to DEHP or DINP and ATAC-Seq library preparation

2.4.1 Exposure to DEHP or DINP

The human thyroid follicular epithelial cell line Nthy-ori 3-1 was plated at a density of 10,000 cells/cm2 on 6-well plates and cultured in RPMI 1640 Medium with GlutaMAX™ Supplement (61870036, Gibco, Waltham, MA, USA), 10% FBS and Penicillin-Streptomycin (15140122, Gibco) and incubated at 37 ºC, 5% CO2 and > 95% humidity. Cells were left one day to adhere, and the following day the media was changed to culture media with DEHP (n = 6) or DINP (n = 6) at 1 µM in 0.5% DMSO. As solvent control, the culture media was additioned with just 0.5% DMSO (n = 6). Cells were incubated for 5 days refreshing the media with the compound or DMSO only at day 3. At the end of the incubation period, cells were collected and counted manually.

2.4.2 ATAC-Seq libraries preparation

Fifty thousand (50,000) cells per sample were used to prepare ATAC-Seq libraries. Libraries were prepared following the Omni-Atac protocol of Corces et al. (45) with the replacement of NP40 from the original protocol with IGEPAL (I8896-50ml, Merck Millipore). The tagmentase kit used was Illumina Tagment DNA Enzyme and Buffer Small Kit (20034197, Illumina, San Diego, CA, USA) and the indexes IDT® for Illumina® DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 samples) (20027213, Illumina). Seven PCR cycles were used for all samples. The prepared libraries were sequenced on an SP Illumina flowcell 100 cycles (v1.5) (Illumina) in paired-end mode.

All RNA samples and sequencing libraries concentrations were measured with the Qubit 2.0 Fluorometer (ThermoFisher), and quality control performed on a BioAnalyzer 2100 expert (Agilent, Santa Clara, CA, USA) or a 2200 TapeStation System (Agilent).

2.4 Data analysis

All the scripts used for RNA-Seq and ATAC-Seq data analysis have been collected in a markdown file available at https://github.com/marta-nazzari/phthalates_rnaseq_atacseq.

2.4.1 RNA-Seq data processing

The fastq files were processed according to our previously published CODA pipeline (46). Briefly, reads were trimmed from the 5' 4N and 3' 8A adapters using Cutadapt (v3.7) (47) as recommended by the manufacturer (48). To obtain genes read counts, trimmed reads were aligned to the mouse transcriptome (GRCm39 v27) and quantified using RSEM (v1.3.3) with the "--STAR" parameter (v2.7.10a), following the ENCODE3's STAR-RSEM pipeline (49, 50). To analyze miRNAs, the trimmed files were used as input for miRge3.0 (v0.0.9) (51) using miRBase mouse annotations (v22).

2.4.2 ATAC-Seq data processing

The fastq files were preprocessed using the PEPATAC pipeline (v0.10.4) (52) using bowtie2 (v2.4.2) (53) as mapper, samtools (v1.4) (54) as deduplicator and the included Python tool "pyadapt" as trimmer. The human genome GRCh38 v38 build was used for alignment.

2.4.3 RNA-Seq samples biotype mapping and outliers identification

Quantified RNA species were mapped to their respective biotypes using the R (55) package biomaRt (56). We calculated the percentage of mapped reads per biotype and retained only those constituting at least 1% in at least one sample. Outliers for each biotype were identified per treatment group (DEHP, DIDP, DINP, DnOP, DMSO, and Untreated) and calculated as being 1.5 times the interquartile range (IQR) below the 25th percentile or above the 75th percentile:

$$biotype\ x\ in\ sample\ y\ <\ 25th\ percentile\ (biotype\ x\ in\ group\ z) - 1.5 * IQR\ (biotype\ x\ in\ group\ z)$$

or

$$biotype\ x\ in\ sample\ y\ >\ 75th\ percentile\ (biotype\ x\ in\ group\ z) + 1.5 * IQR(biotype\ x\ in\ group\ z)$$

2.4.4 MaSigPro analysis

Normalized gene counts were used for maSigPro (57) analysis according to the maSigPro User's Guide for Next-Generation Sequencing data (58) for a single series course experiment. We set the "tetha" (θ) value to 10 (default), the FDR to 0.05 (default) and the "degree" parameter to 3 (this corresponds to a cubic polynomial regression model). The variable

"Time" with values 0, 1, 2, 3, 4, 5 was used in the model to represent the "Dose" values of 0 (DMSO control), 1 nM, 10 nM, 100 nM, 1 µM and 10 µM (or 2 nM, 20 nM, 200 nM, 2 µM and 20 µM for DnOP).

2.4.5 Differential expression analysis

Differential gene and miRNA expression analysis was performed comparing the phthalate-treated samples to the DMSO solvent control using the R package DESeq2 (59) following a slightly modified version of the guidelines of the Omics Data Analysis Framework for regulatory application (R-ODAF) pipeline developed by our group (60, 61). Briefly, a first filtering step ("relevance threshold") was applied to select the expressed genes/miRNAs by retaining only those whose normalized expression is ≥1 count per million (CPM) in at least 75% of the samples in either group (i.e. treatment versus control). To increase statistical power, all doses of a single compound were grouped together and compared to DMSO control. The RUVg function from the RUVSeq package (62) (k = 2) was used on the genes/miRNAs passing the relevance threshold filter to remove unwanted variation. Then, differential expression analysis on the expressed genes/miRNAs was performed setting the FDR to 0.01. The resulting differentially expressed (DE) genes/miRNAs/snoRNAs were subjected to an additional filtering step ("spurious spikes") to identify those cases in which a very high expression value in only one replicate in a group is responsible for a certain gene/miRNA/snoRNA to result differentially expressed. To this end, the following formula was applied to every DE gene/miRNA/snoRNA for both treatment and control groups $\frac{read\ count\ gene/miRNA_i}{total\ read\ count\ gene/miRNA_i\ in\ group_j} < 1.4 \times \left(number\ of\ replicates\ in\ group_j\right)^{-0.66}$, where i refers to any gene/miRNA/snoRNA, and j to either the treatment or control group. The expression of such genes/miRNAs/snoRNAs was manually checked in all replicates to determine whether a gene that failed this "spurious spike" filter was indeed a technical artefact or could instead be biologically relevant.

2.4.6 Gene set enrichment analysis (GSEA)

GSEA was performed using the R package ReactomePA (v1.40.0) (63) and Reactome as database (64) using DESeq2 "stat" value for gene ranking. For significance, we set a q-value threshold of 0.05.

2.4.7 Differential accessibility analysis

The alignment files (.bam) output by the PEPATAC pipeline were shifted with the deepTools (v3.5.1) (65) utility alignmentSieve to account for the Tn5 transposase duplication

at the cut site. To identify differentially accessible (DA) regions, we used a sliding window approach with the R package csaw (v1.32.0) (66) and a modified version of the script made available by Sheikh and Blais (2022) on bioRxiv (67). For quantification, we used the 5-prime reads, a sliding window of 50 bp without overlap, and a minimum number of counts of 50 for a window to be retained. To calculate the background, we binned the genome in 10 kb bins. To distinguish the signal from the background, we compared each region against the global background and set a fold change compared to the background to 3. The differential accessibility analysis was performed with the R package edgeR (v3.4.0) (68) and we performed batch correction using the RUVs function (k = 5) from the RUV-Seq package (v1.32.0) (62). As multiple testing correction should be performed on regions and not windows (69), we merged the regions identified as "signal" that are at most 500 bp apart, reaching up to a maximum merged region width of 5 kbp, and performed multiple-testing correction using the Benjamini-Hochberg method (FDR = 0.01).

Identified differentially accessible regions were annotated with HOMER (v3.13) (70) (genome version hg38). Regulatory regions annotations were retrieved from the ENCODE Candidate Cis-Regulatory Elements (cCREs) registry (71). Coverage tracks were normalized using BeCorrect (v1.1.0) (72) and visuals extracted from the Integrative Genomics Viewer (IGV) (v2.13.2) (73).

3. Results

3.1 RNA-Seq results

To investigate the effect of phthalates on the transcriptome, we generated RNA-Seq data from mESC-derived thyroid organoids exposed to four phthalates using *in vivo* relevant concentrations (1-10-100 nM-1-10 µM DEHP, DIDP, DINP; 2-20-200 nM-2-20 µM DnOP) for 24 hours. A schematic representation of the exposure regimen is shown in Figure 1. In the following paragraphs, we provide some dataset quality control (QC) metrics followed by the results of gene expression analysis.

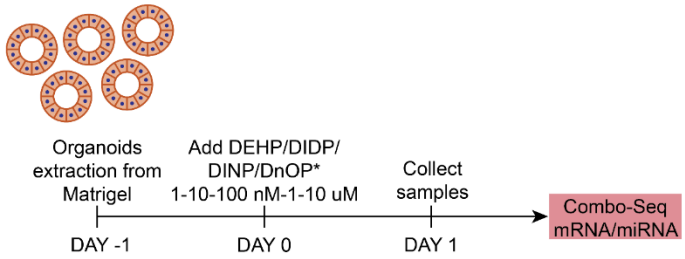


Figure 1. Graphical representations of the exposure regimens of thyroid follicles to the phthalates (* = for DnOP the concentrations used were 2-20-200 nM-2-20 µM).

3.1.1 RNA-Seq data QC and outliers identification

Combo-Seq libraries had a median 51.8 million (M) reads per sample (min = 17.7, max = 92.4 M) (Figure 2A), with a median of 97% of sequenced reads with quality score of 30 or more (min = 96.4%, max = 97.3%) (Figure 2B). The median number of reads mapped to mRNAs was 41.8 M (min = 14.3, max = 78.8 M) and to miRNAs 1.30 M (min = 0.16, max = 2.2 M) (Figure 2C).

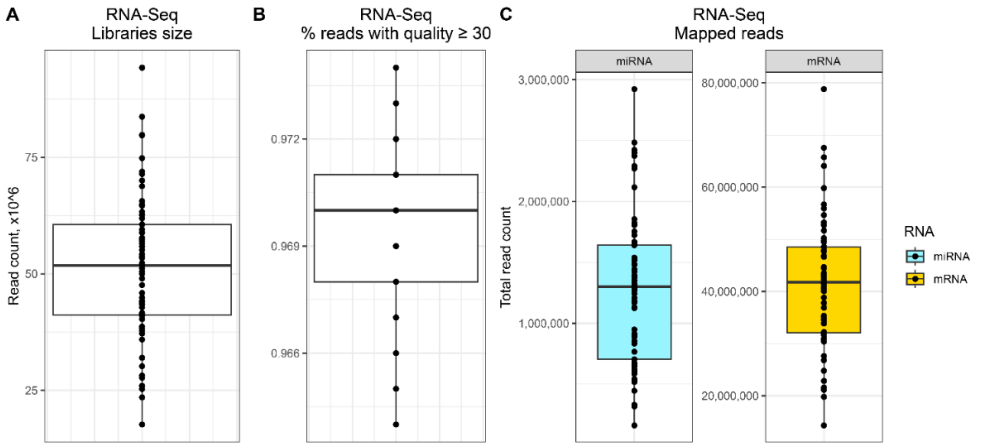


Figure 2. RNA-Seq dataset metrics. (A) Number of sequenced reads. (B) Percentage of sequenced reads with quality score Q30 > =30. (C) Number of reads mapped to miRNAs (light blue) or mRNAs (yellow). Each dot represents a sample.

As explained in the Methods section, we performed 16 cycles for some RNA-Seq libraries to reach the required concentration for sequencing (1.6 nM) (Supplementary table 1). In consequence, this increased the percentage of snoRNA-mapping reads (Figure 3A). As the read count of the protein-coding genes would be underestimated during DESeq2 normalization, we removed the snoRNA-mapping reads from the main dataset and performed the analysis of snoRNA genes separately. Boxplot of mapped read distributions per gene biotype after snoRNA removal revealed that one DMSO control replicate (DMSO_ctrl_1) was a clear outlier in multiple biotypes (Figure 3B). As the DMSO samples would be used as control for all comparisons, this outlier would have had a major impact in all downstream analyses, and importantly in the most important biotypes ("protein coding" and "miRNA" in particular). Although other samples were flagged as outliers in other biotypes ("processed pseudogene" or "rRNA"), this was less consistent and did not warrant further samples removal.

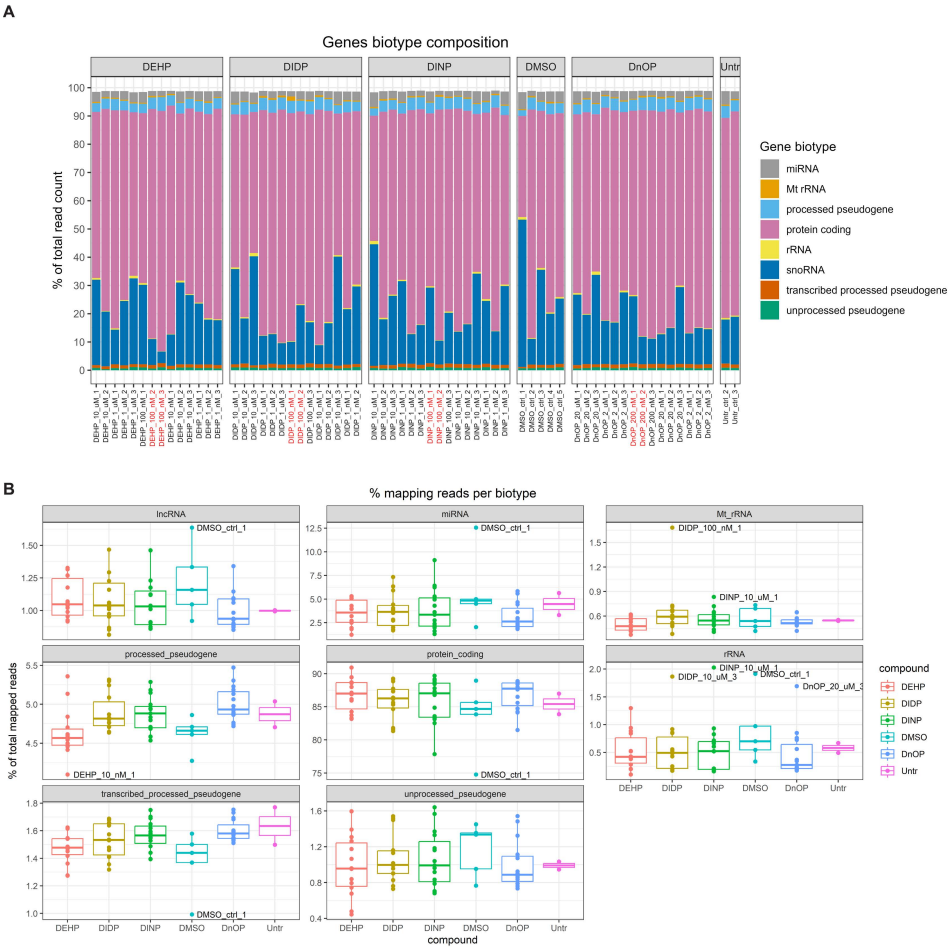


Figure 3. (A) Genes biotype composition of RNA-Seq samples expressed as percentage of total normalized read count. For clarity and to reduce noise, only the biotypes that make up at least 1% of the total in at least one sample are reported. The sample names for which 14 cycles during the library prep protocol were used are reported in red. (B) Biotype distribution after snoRNA removal. The biotypes that make up at least 1% of the mapped reads are reported. The labelled samples are the ones flagged as outliers.

3.1.2 MaSigPro analysis

The MaSigPro R package, initially developed to identify changes in gene expression along a time series, can also be used to analyze the evolution of the gene expression level across a dose range exposure. We then investigated using MaSigPro whether some genes would show dose regulation across our six doses (untreated plus five doses). We then allowed the significant genes to be grouped into 9 clusters, which include the genes that have a similar trend in change in expression over the dose series. For every compound, we observed some clusters with a non-monotonic dose-response curve (DEHP: 1, 3-5; DIDP: 1, 3-9; DINP: 1, 3-5, 9; DnOP: clusters

1, 3, 6, 8) (Figure 4). The genes belonging to the various clusters are reported in Table 1. In those indicated clusters, the highest dose (10 µM or 20 µM) was consistently showing to be different from the other four. In this cellular system, this dose could be used for phthalates to derive a point of departure (PoD) metric, which in the toxicology field represents a dose at which a biological response is first observed and can be used to make extrapolations for risk assessment (74).

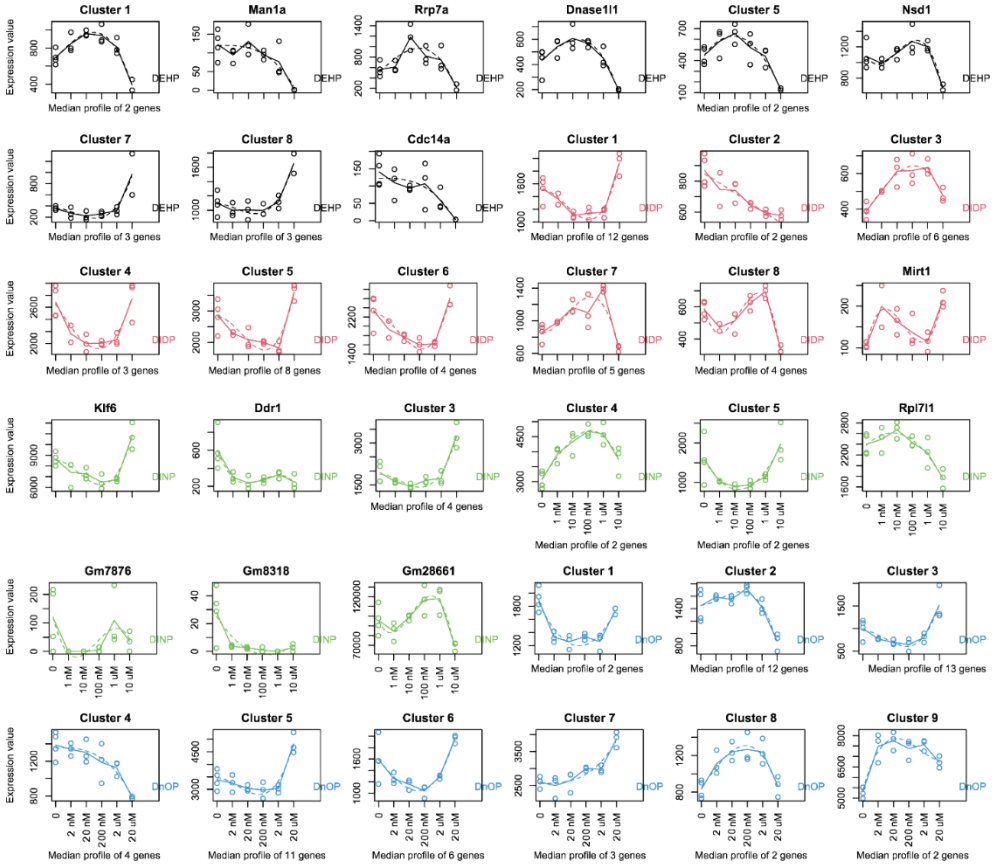


Figure 4. MaSigPro analysis of the gene expression over the dose series employed in the study (0-1-10-100 nM-1-10 µM for DEHP, DIDP and DINP, and 0-2-20-200 nM-2-20 µM for DnOP). The curves for each compound are color-coded and the compound name is reported in every plot. Each gene was analyzed to fit up to a cubic polynomial regression model. Fits that passed multiple testing correction (FDR = 0.05) were selected and clustered in 9 groups using hierarchical clustering. If a cluster comprises only one gene, the gene name is indicated on top of the plot, otherwise the cluster name is reported. The x-axis reports the dose range used (nM = nanomoles per liter, µM = micromoles per liter). The dots represent the expression values of each replicate (or the average of each gene if a cluster comprises more genes) and the dotted line shows the fit. The genes belonging to the various clusters are reported in Table 1.

Table 1. Genes belonging to the different clusters identified by maSigPro analysis (degree = 3, FDR < 0.05). Note that the same cluster number in the different series does not necessarily refer to the same curve shape.

	DEHP	DIDP	DINP	DnOP
Cluster 1	*Abcf3, Blmh*	*Brf2, Brip1os, Capns1, Ccdc84, Dpf2, Kank1, Kat7, Reep6, Scamp2, Tgfbi, Yars, Zswim4*	*Klf6*	*Ccnd2, Lor*
Cluster 2	*Man1a*	*Celf2, Gm8318*	*Ddr1*	*Cbx1, Ccnb1, Hspbap1, Kif23, Larp7, Lbr, Ndc80, Nfyc, Slc7a7, Srsf3, Uba2, Zzz3*
Cluster 3	*Rrp7a*	*1600014C10Rik, Dnajc8, Eda2r, Gm10323, Naa20, Zfp87*	*Dnajb1, Mxi1, Net1, Tbl2*	*4833411C07Rik, Gm25395, Gm7908, Lman2l, Pank4, Pcsk1n, Pi4k2a, Ptpn23, Rnasek, Slc25a34, Slc25a39, Slc38a1, Slc39a7*
Cluster 4	*Dnase1l1*	*Aplp2, Gorasp2, Ppp1r18*	*Acaa2, Ing5*	*Asf1b, Basp1, Gm32885, Ncoa4*
Cluster 5	*Lipt1, Serpinb9b*	*Cxcl14, Mrps34, Nop53, Prpf8, Prpsap1, Rab4a, Selenow, Slc25a3*	*Pnrc1, Setd5*	*Dap, Ddit3, Gstm1, Luc7l, Map1b, Pde4d, Slc3a2, Slc48a1, Srxn1, Tmbim1, Ttc3*
Cluster 6	*Nsd1*	*Lbp, Parm1, Sel1l, Tspan7*	*Rpl7l1*	*Gpr137b, Mid1, Nat9, Polk, Sel1l, Zbtb20*
Cluster 7	*Higd2a, Secisbp2l, Slc6a2*	*2610021A01Rik, 4930430F08Rik, Rab5c, Trmt10c, Xrcc6*	*Gm7876*	*Cyb5b, Mtres1, Sys1*
Cluster 8	*Mark4, Pkm, Ubap1*	*Cd2bp2, Ddx6, Mgme1, Snx1*	*Gm8318*	*Exoc6, Mrps22*
Cluster 9	*Cdc14a*	*Mirt1*	*Gm28661*	*Acaa2, Dnajc8*

3.1.3 Differentially expressed genes and miRNAs

Considering the divergent nature of the highest dose compared to the other four, we decided to exclude it from the differential expression analysis. Given that for each dose we had

triplicates or duplicates, by consolidating all the remaining doses together and comparing them to the solvent control, we aimed at increasing the statistical power and detecting gene, miRNA and snoRNA expression alterations specifically attributable to phthalate treatment. By doing so, we could focus on identifying changes at the compound level while accounting for the different response observed with the highest dose.

Differential expression analysis revealed how all the treatments had moderate effects on the cells in terms of number of differentially expressed genes (DEGs), miRNAs and snoRNAs (Figure 5): the number of DEGs compared to the control was 5, 5, 10 and 49 for DEHP, DIDP, DINP and DnOP, respectively (FDR < 0.01) (Table 2). Only DIDP treatment influenced miRNA expression, with mmu-miR-143-3p being downregulated. No effect was observed on snoRNA expression as well as on thyroid markers (Supplementary figure 1).

Table 2. List of differentially expressed genes (DEGs) and miRNAs (DE miRNAs) in every phthalate vs DMSO comparison (FDR < 0.01). Downregulated genes are reported with blue text, upregulated ones in red. DEGs that appear in more than one comparison are in bold.

	DEHP	DIDP	DINP	DnOP			
DEGs	*Gpd1* ***Ing5*** *Myh14* ***Acaa2*** *Gm15516*	***Plekha3*** *Cxcl14* *Zgpat* *Rpl38-ps2* *Gm10323*	***Acaa2*** ***Cops5*** ***Idh3g*** ***Ing5*** ***Mid1*** ***Rab5a*** *Srrm1* *Tnrc6b* *Ttc32* ***Zfp960***	***Acaa2*** *App* *Arhgef10l* *Ccnd2* *Ccni* ***Cops5*** *Dnajc8* *Exoc6* *Fbp2* *Gm21293* *Gm5641* *Hbb-bs* *Herpud2*	*Hsd17b10* *Hspa1b* ***Idh3g*** *Ifitm3* ***Ing5*** ***Mid1*** *Mtpn* *Mzt1* *Npm1* *Nrip1* *Nucb2* *Phlda1*	***Plekha3*** *Pole2* *Ppp1r7* *Ptp4a1* ***Rab5a*** *Rars* *Rnf128* *Rpl19-ps11* *Shroom3* *Sox4* *Tjp1* *Tmed7*	*Tmem80* *Trmt61b* *Tspan1* *Ube2g2* *Uqcrc2* *Vps25* *Wasf2* *Zc3h11a* *Zfand1* *Zfp330* ***Zfp960*** *Zswim6*
DE miRNAs			mmu-miR-143-3p				

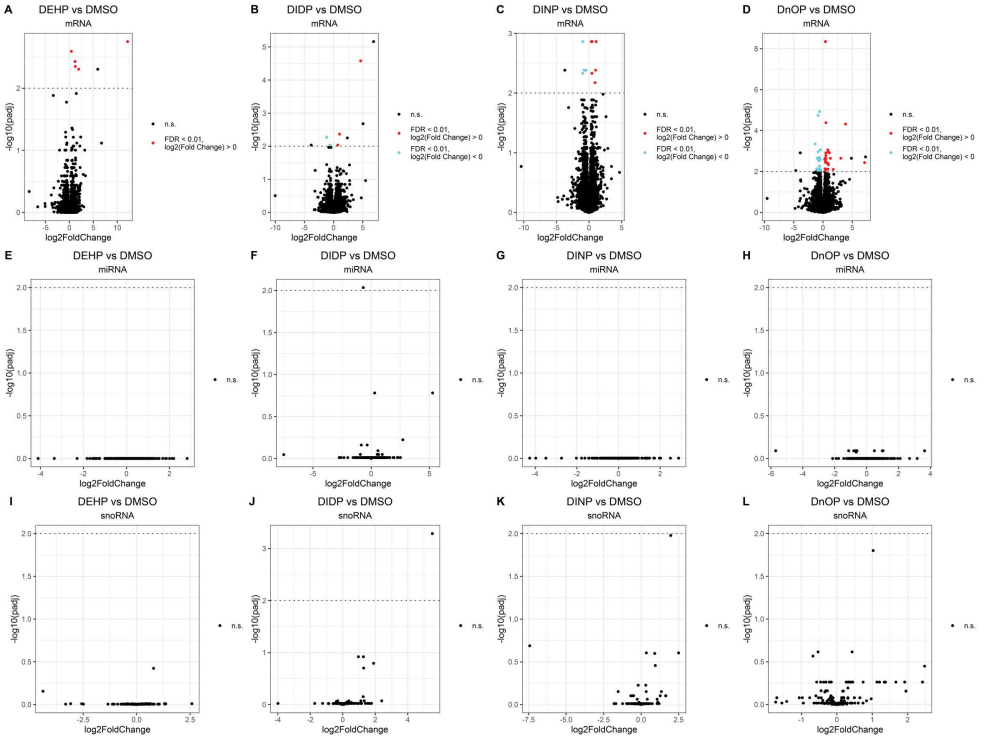


Figure 5. Volcano plots of the expressed genes (A-D), miRNAs (E-H) and snoRNAs (I-L) in each phthalate vs DMSO control. every dot represents a gene. Element not differentially expressed ('n.s.') are in black. The False Discovery Rate (FDR) threshold of 0.01 is indicated as a dotted line on the y axis. Upregulated elements ('FDR < 0.01, log2(Fold Change) > 0') are indicated in red, downregulated ones ('FDR < 0.01, log2(Fold Change) < 0') in cyan. The genes above this line colored in black are the ones that fail to pass the 'spurious spike' filter as described in the Methods section.

Interestingly, despite the weak effects on gene expression, the Inhibitor of growth protein 5 (*Ing5*) gene was consistently upregulated in three out of four treatments (FDR: DEHP vs DMSO = 4.44e-3, DIDP vs DMSO = 0.14, DINP vs DMSO = 1.37e-3, DnOP vs DMSO = 1.23e-3) (Figure 6).

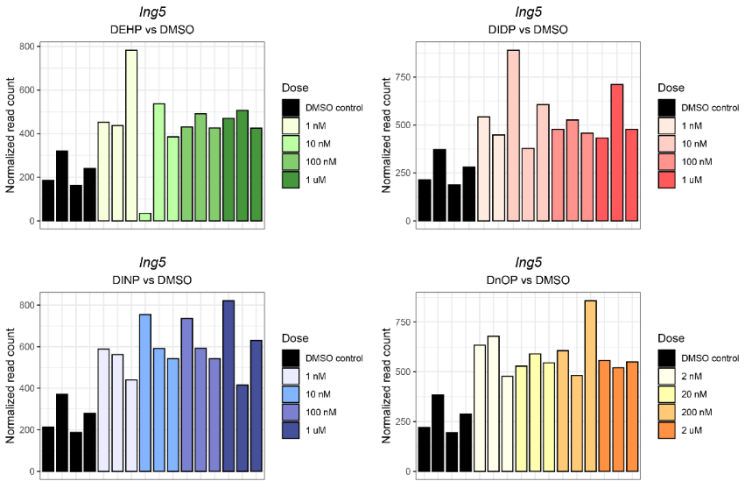


Figure 6. *Ing5* normalized expression in each phthalate and DMSO control samples. The different doses are reported in the legend. The darkest shade corresponds to the highest concentration (10 or 20 µM), while the lightest shades to the lowest (1 or 2 nM). The DMSO control samples are reported in black.

Other genes differentially expressed in more than one condition were identified: *Acaa2* (DEHP, DINP and DnOP vs DMSO), *Plekha3* (DIDP and DnOP vs DMSO) and five genes (*Cops2*, *Idh3g*, *Mid1*, *Rab5a*, *Zpf960*) dysregulated in DINP and DnOP vs DMSO.

3.1.4 Gene set enrichment analysis (GSEA)

GSEA was performed using the Reactome database using a q-value threshold of 0.05. We identified 123 enriched pathways in the DEHP vs DMSO comparison, 79 in DIDP vs DMSO, 173 in DINP vs DMSO and 311 in DnOP vs DMSO (Figure 7).

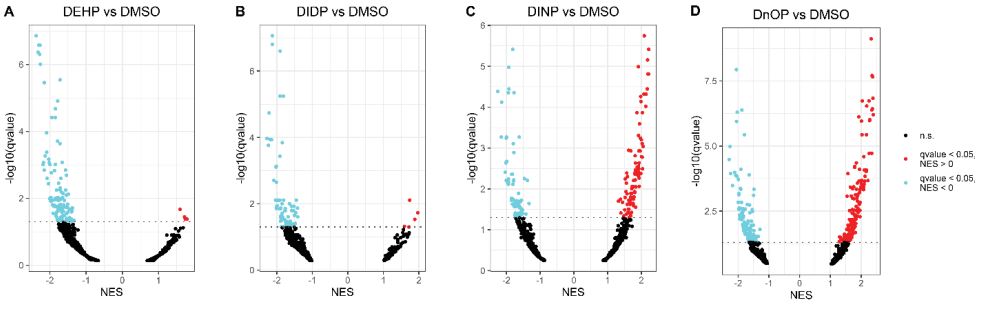


Figure 7. Volcano plots of the enriched Reactome pathways identified by GSEA analysis in each phthalate vs DMSO control (A: DEHP vs DMSO; B: DIDP vs DMSO; C: DINP vs DMSO; D: DnOP vs DMSO). Every dot represents a pathway. Pathways with a q-value < 0.05 and normalized enrichment score (NES) > 0 are in red. Pathways with a q-value < 0.05 and NES < 0 are in cyan. Pathways with q-value >= 0.05 are in black ('n.s.'). The q-value threshold of 0.05 is indicated as a dotted line on the y-axis.

In both DEHP and DIDP vs DMSO comparisons, most pathways were enriched in the control (normalized enrichment score NES < 0). Conversely, we observed a balance between pathways enriched in the treatment (NES > 0) and in the control in the DINP and DnOP vs DMSO comparisons. To identify common effects across the treatments, we focused on the pathways which appeared in all comparisons (Figure 8), thus retrieving 23 terms, 1 enriched in the treatment, 22 enriched in the DMSO control.

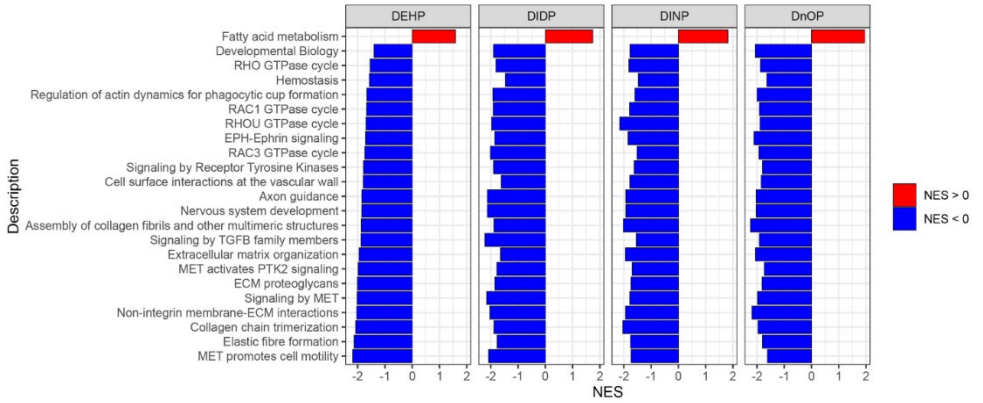


Figure 8. Results of gene set enrichment analysis (GSEA) on Reactome pathways. The reported pathways have a q-value < 0.05 and appear in all four phthalates vs DMSO comparisons.

Among the selected pathways with NES < 0, we identified several terms related to signal transduction and extracellular matrix (ECM) organization. The only term with NES > 0 was "Fatty acid metabolism". In Figure 9 the terms are reported with their respective position in the Reactome terms hierarchy, for better understanding of their relationships.

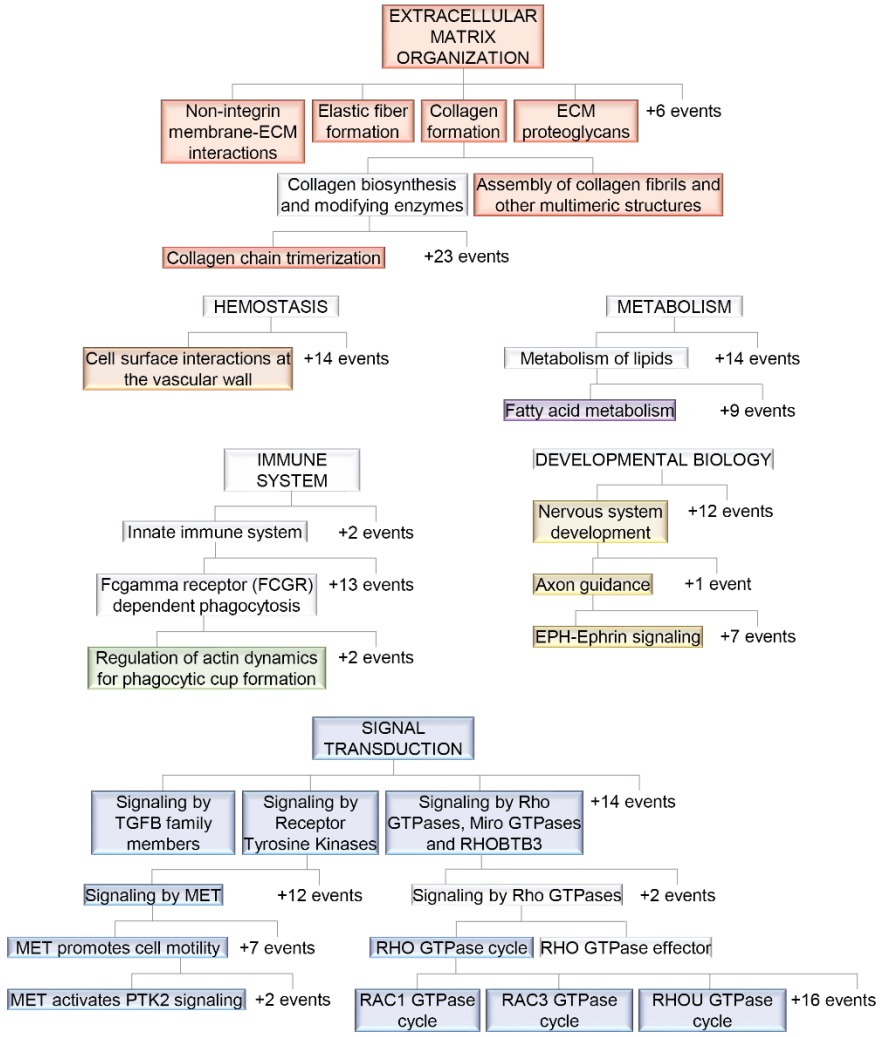


Figure 9. Common enriched Reactome pathways in all four phthalates identified by GSEA. The top term, in uppercase, represents the most parent term in the hierarchy. The colored terms are the ones reported in Figure 4, while the terms in a white rectangle are reported for better understanding the relationships among terms. If a branch has more children events, it is indicated as "+N events".

3.2 ATAC-Seq results

As explained in the previous section, *Ing5* upregulation was observed in 3 phthalate exposures out of 4. ING5 is a component of the histone acetyltransferase complexes HBO1-JADE, that mediates histone H4 acetylation *in vivo*, and HBO1-BRPF1 and MOZ/MORF that mediate histone H3 acetylation (43, 44). For this reason, we investigated whether phthalate treatment could have an impact on the chromatin status with ATAC-Seq. To this end, we

selected two of the four phthalates, DEHP and DINP, and the highest dose included in the differential expression analysis (1 µM). The exposure was increased to 5 days, to allow time to any chromatin rearrangements, if any, to take place, accounting for any delay between gene upregulation of *Ing5* and an actual observable effect on the epigenome. For ATAC-Seq library preparation, a viability of at least 90% was required. Unfortunately, we were not able to recover enough cells from our thyroid follicles model with this viability. For this reason, we selected the human epithelial thyroid cell line Nthy-ori 3-1 (Figure 10). In the next sections some quality control metrics of the ATAC-Seq libraries and the results of the differential accessibility analysis are reported.

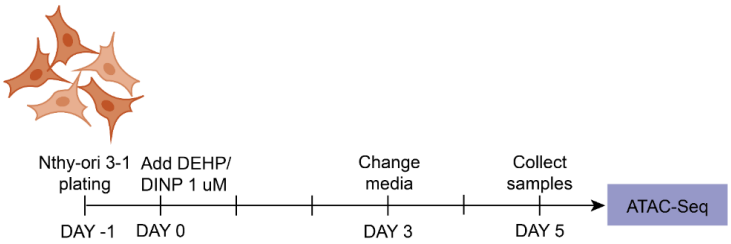


Figure 10. Graphical representations of the exposure regimens of Nthy-ori 3-1 cells to the phthalates.

3.2.1 ATAC-Seq QC

ATAC-Seq libraries had a median 82.2 million (M) reads per sample (min = 26.5, max = 237.9 M) (Figure 11A), with a median 77.36% (min = 75.46%, max = 79.08%) of sequenced reads being successfully aligned to the GRCh38 nuclear genome (Figure 11B). The Transcription Start Site (TSS) enrichment score had a distribution between 10.7 and 19.5 (median 18.1) (Figure 11C). The distribution of nucleosome free regions (NFR), mono-, di-, tri- or poly-nucleosome regions was consistent across samples (Figure 11D) and the read length distribution profiles typical of ATAC-Seq libraries (Figure 11E). The library complexity metrics were within the accepted values recommended by the Encode Project (Figure 11F) (75).

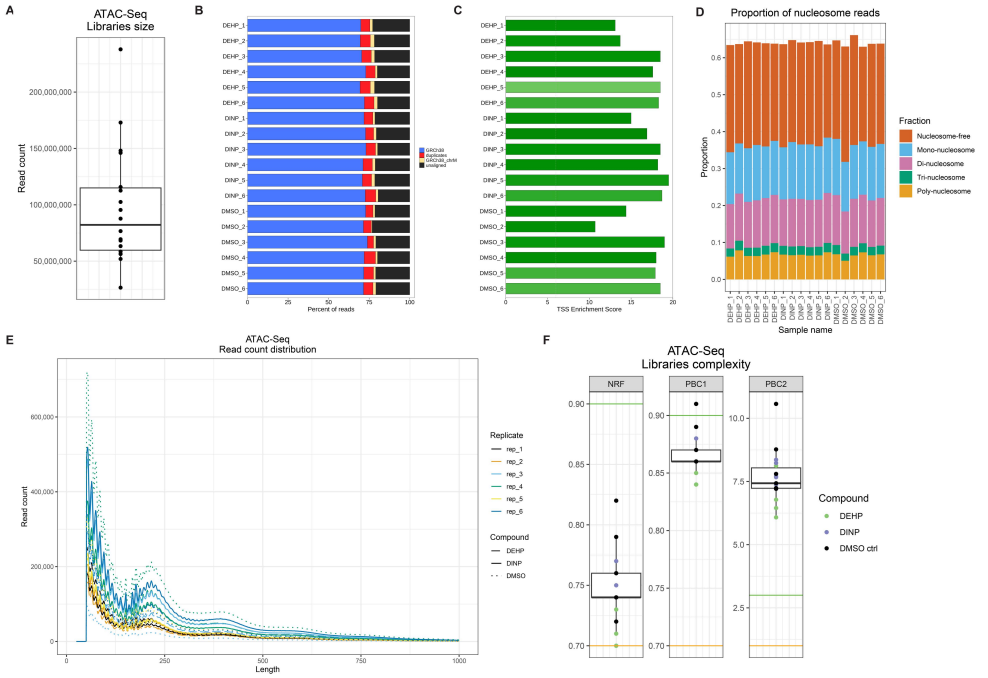


Figure 11. ATAC-Seq dataset metrics. (A) Libraries size. (B) Percentage of reads aligned to GRCh38 nuclear or chromosomal genome, duplicated reads and unaligned reads. (C) Transcription Start Site (TSS) enrichment score. (D) Proportion of nucleosome-free, mono-, di-, tri- and poly-nucleosome reads. (E) Read length distribution. Each line represents a sample. (F) Boxplots of ATAC-Seq library complexity metrics. Each dot represents a sample (green = DEHP, violet = DINP, black = DMSO control). The green and orange lines represent the ENCODE thresholds for an acceptable (orange) or ideal (green) metric. (NRF = Non-Redundant Fraction; PBC1 = PCR Bottlenecking Coefficient 1; PBC2 = PCR Bottlenecking Coefficient 2). Plots (B) and (C) are output by the PEPATAC pipeline.

3.2.2 Differential accessibility analysis by ATAC-Seq

We identified 111,133 genomic regions when comparing DEHP-treated and DMSO samples, and 118,855 regions in the DINP vs DMSO comparison tested for differential accessibility. In both treatments, we observed a general increase in accessibility compared to the control, but none of the regions passed multiple-testing correction (Figure 12).

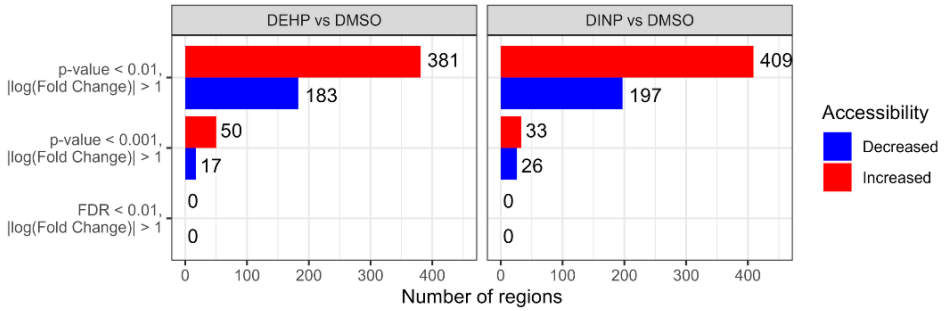


Figure 12. Number of differentially accessible regions with p-value < 0.001 or 0.01 and |log(Fold Change)| > 1 in the DEHP- or DINP-treated samples vs DMSO control. After multiple-testing correction, none of the regions had an FDR < 0.01.

We annotated the regions with p-value < 0.01 with HOMER to identify the closest gene to each region and looked for overlap between DEHP vs DMSO and DINP vs DMSO. We identified 4 regions with decreased accessibility and 17 with increased accessibility that overlap with regulatory regions (Table 3).

Table 3. Differentially accessible (DA) regions identified both in the DEHP vs DMSO and DINP vs DMSO comparisons that overlap with regulatory regions. DA regions were annotated with HOMER to identify the closest gene. A description of the gene region where the DA is located is reported. The regulatory regions that overlap with the identified DA regions were retrieved from the ENCODE Candidate Cis-Regulatory Elements (cCREs) registry (dELS = distal enhancer-like signature, pELS = proximal enhancer-like signature, PLS = promoter-like signature).

cCRE ID	cCRE type	Closest gene	Gene type	Gene description	Annotation	Accessibility compared to control
EH38D2971615	dELS	*ASPSCR1*	Protein coding	Alveolar soft part sarcoma chromosome region, candidate 1	Intron 3 of 14	Increased
EH38D2949029	dELS,CTCF-bound	*BCAS3*	Protein coding	Breast carcinoma amplified sequence 3	Intron 5 of 24	Increased
EH38D3590998	dELS,CTCF-bound	*CCNI*	Protein coding	Cyclin I	TTS	Increased
EH38D4073480 EH38D4073481 EH38D4073482 EH38D4073483 EH38D4073484 EH38D4073485 EH38D4073486 EH38D4073487	pELS,CTCF-bound pELS pELS,CTCF-bound pELS,CTCF-bound PLS,CTCF-bound pELS,CTCF-bound pELS pELS	*CDCA2*	Protein coding	Cell division cycle associated 2	Promoter-TSS	Increased

cCRE ID	cCRE type	Closest gene	Gene type	Gene description	Annotation	Accessibility compared to control
EH38D3714835	dELS	*CDK7*	Protein coding	Cyclin-dependent kinase 7	Intron 11 of 11	Increased
EH38D3539140 EH38D3539141 EH38D3539142 EH38D3539143 EH38D3539144 EH38D3539145 EH38D3539146 EH38D3539147 EH38D3539148	pELS,CTCF-bound pELS,CTCF-bound pELS,CTCF-bound PLS DNase-H3K4me3 DNase-H3K4me3,CTCF-bound DNase-H3K4me3 DNase-H3K4me3 pELS	*CPLX1*	Protein coding	Complexin 1	Promoter-TSS	Increased
EH38D3817447	pELS	*DTNBP1*	Protein coding	Dystrobrevin binding protein 1	TTS	Increased
EH38D3721344	dELS	*LHFPL2*	Protein coding	Lipoma HMGIC fusion partner-like 2	Intron 2 of 4	Increased
EH38D4221907	dELS,CTCF-bound	*FKTN*	Protein coding	Fukutin	Intron 7 of 11	Increased
EH38D4071656	dELS,CTCF-bound	*LOC100507156*	ncRNA	Uncharacterized LOC100507156	Intron 7 of 13	Increased
EH38D3794981	dELS,CTCF-bound	*MIR8056*	ncRNA	microRNA 8056	Intergenic	Increased
EH38D2141456	pELS,CTCF-bound	*PAFAH2*	Protein coding	Platelet-activating factor acetylhydrolase 2, 40kDa	Intergenic	Increased
EH38D4190148	dELS,CTCF-bound	*SLC25A51*	Protein coding	Solute carrier family 25, member 51	Intron 5 of 5	Increased
EH38D3296811 EH38D3296812 EH38D3296813 EH38D3296814	dELS,CTCF-bound dELS,CTCF-bound dELS dELS,CTCF-bound	*SRC*	Protein coding	SRC proto-oncogene, non-receptor tyrosine kinase	Intergenic	Increased
EH38D3526429	CTCF-only,CTCF-bound	*VPS8*	Protein coding	Vacuolar protein sorting 8 homolog (*S. cerevisiae*)	Intron 20 of 46	Increased
EH38D3015415	dELS	*CCBE1*	Protein coding	Collagen and calcium binding EGF domains 1	Intron 2 of 10	Decreased
EH38D3079246	dELS,CTCF-bound	*PXDN*	Protein coding	Peroxidasin homolog	Intergenic	Decreased

cCRE ID	cCRE type	Closest gene	Gene type	Gene description	Annotation	Accessibility compared to control
				(Drosophila)		
EH38D3369502	CTCF-only,CTCF-bound	*TEX33*	Protein coding	Testis expressed 33	Intron 5 of 5	Decreased
EH38D2338583	dELS	*TMEM72*	Protein coding	Transmembrane protein 72	Intergenic	Decreased

We looked at which of these regions fall within the transcription start site (TSS) or transcription termination site (TTS). We identified four regions, all with increased accessibility, two located at the TSS of the *CDCA2* (Cell Division Cycle Associated 2) and *CPLX1* (Complexin 1) genes, and two located at the TTS of *CCNI* (Cyclin I) and *DTNBV1* (Dystrobrevin Binding Protein 1). However, when inspecting the normalized read coverage on the Integrative Genomics Viewer (IGV), there did not seem to be a clear increase in accessibility compared to the control (Table 4, Supplementary figure 2).

Table 4. Differentially accessible (DA) regions identified both in the DEHP vs DMSO and DINP vs DMSO comparisons that overlap with regulatory regions and fall within the Transcription Start Site (TSS) or Transcription Termination Site (TTS) of the closest gene. DA regions were annotated with HOMER to identify the closest gene. The regulatory regions that overlap with the identified DA regions were retrieved from the ENCODE Candidate Cis-Regulatory Elements (cCREs) registry (dELS = distal enhancer-like signature, pELS = proximal enhancer-like signature, PLS = promoter-like signature).

cCRE ID	cCRE type	Closest gene	Gene description	Annotation	Distance from TSS, nt	Accessibility compared to control
EH38D3590998	dELS,CTCF-bound	*CCNI*	cyclin I	TTS	28,721	Increased
EH38D4073480	pELS,CTCF-bound	*CDCA2*	cell division cycle associated 2	promoter-TSS	-221	Increased
EH38D4073481	pELS					
EH38D4073482	pELS,CTCF-bound					
EH38D4073483	pELS,CTCF-bound					
EH38D4073484	PLS,CTCF-bound					
EH38D4073485	pELS,CTCF-bound					
EH38D4073486	pELS					
EH38D4073487	pELS					
EH38D3539140	pELS,CTCF-bound	*CPLX1*	complexin 1	promoter-TSS	-519 (DEHP vs DMSO); -94 (DINP vs DMSO)	Increased
EH38D3539141	pELS,CTCF-bound					
EH38D3539142	pELS,CTCF-bound					
EH38D3539143	PLS					
EH38D3539144	DNase-H3K4me3					
EH38D3539145	DNase-H3K4me3,CTCF-bound					
EH38D3539146	DNase-H3K4me3					
EH38D3539147	DNase-H3K4me3					
EH38D3539148	pELS					
EH38D3817447	pELS	*DTNBP1*	dystrobrevin binding protein 1	TTS	139,732	Increased

4. Discussion

In this work, we analyzed the response of mouse embryonic stem cell-derived thyroid follicles after exposure to the phthalates DEHP, DIDP, DINP and DnOP in a range of concentrations from 1 nM to 10 µM (2 nM to 20 µM for DnOP) for 24 hours. The low dose range was selected to reflect the low daily intake of phthalates measured in the general population (76), and the relatively short exposure time to detect the initial reaction to phthalates exposure by identifying the early changes in the transcriptome. In this way, we aimed at testing whether our model would be able to capture the molecular initiating event (MIE) of these phthalates which would then be followed by the key events (KEs), to ultimately lead to an adverse outcome (integrated in the concept of an adverse outcome pathway (AOP)) (77). We performed RNA-Seq analysis, and simultaneously analyzed both mRNA and small RNAs from the same samples. The dose series analysis showed how most of the identified genes either increased or decreased sharply in expression at the highest dose, setting it apart from the others and possibly indicating it as a dose to determine a point of departure (PoD) for those genes, which is used in toxicology to establish a threshold dose for risk assessment (74, 78).

The compounds showed a modest effect on the cells at the time and doses of exposure in terms of number of differentially expressed genes and miRNAs, while no effect was observed on snoRNAs expression. DIDP was the only compound where a microRNA (mmu-miR-143-3p) was downregulated. This microRNA, together with mmu-miR-143-5p, has been observed to be downregulated in several cancers and is thought to have tumour-suppressing activity and being a negative regulator of cell proliferation (79-81). Despite the low number of DEGs, we observed a partial overlap across treatments (*Acaa2* and *Plekha3* in 3 treatments, *Cops2, Idh3g, Mid1, Rab5a* and *Zpf960* in 2 treatments). It is possible that the higher number of DEGs in the DnOP vs DMSO comparison could be explained by the doses used, being twice as high than the other phthalates, though still within the same order of magnitude. Acaa2 (Acetyl-CoA Acyltransferase 2) is one of the enzymes that catalyses the last step of the mitochondrial beta-oxidation pathway. Plekha3 (Pleckstrin Homology Domain Containing A3) is involved in the regulation of vesicular cargo transport from the trans-Golgi network to the plasma membrane and is predicted to be involved in ceramide transport and intermembrane lipid transfer (82, 83). Cops2 is a member of the COP9 signalosome complex (CSN), which is involved in decreasing the ubiquitin ligase activity of the SCF-type E3 ligase complexes. Idh3g is an enzyme that takes part in the Krebs cycle and performs the decarboxylation of isocitrate into alpha-ketoglutarate. Mid1 (Midline 1) is likely involved in the formation of multiprotein structures acting as anchor points to microtubules. It has also E3 ubiquitin ligase activity towards the protein Igbp1, promoting its degradation. Rab5a is a member of the RAS oncogene family and is a small GTPase which, in its active form, recruits proteins responsible for vesicle formation, movement, tethering and fusion (82).

Via gene set enrichment analysis (GSEA), we looked for enriched pathways shared by the four treatments to try and identify effects that could be attributed to the "phthalates" EDC class. Only the pathway "Fatty acid metabolism" was enriched in all treatments. Interestingly, phthalates have been shown to increase the metabolism of fatty acids in the liver (84, 85), but also in non-liver tissue such as cardiomyocytes, where increased used of fatty acids for energy production was suggested (86). To our knowledge, our analysis is the first observation of the conservation of these mechanisms in an *in vitro* thyroid model. Additionally, among the pathways downregulated in the treatment groups we found many related to cell extracellular matrix (ECM) organization and receptor signaling. It is also interesting to note that despite the low number of DE genes, due to the low doses used, we were still able to detect relevant enriched pathways using GSEA.

Taken together, the results of differential gene expression gene analysis and GSEA seem to point to an effect of phthalates on energy production, with genes involved in the cellular respiration being dysregulated, and lipid metabolism increasing.

Ing5 was upregulated in three treatments (DEHP, DINP, DnOP) compared to the control. The ING family comprises 5 genes (*ING1* to *ING5*), which have a role in cell cycle regulation and cell proliferation by interacting with several partners, such as p53, p300 and histone acetylation complexes (43, 87). *ING5* is a tumor-suppressor gene which is downregulated in several types of cancer including thyroid (88), colorectal (89), breast (90) and lung (91). Its protein is a component of the histone acetyltransferase HBO1-JADE, which acetylates histone H4 at Lys residues 5, 8 and 12 (H4K5ac, H4K8ac, H4K12ac), MOZ/MORPH, that performs histone H3 acetylation, and HBO1-BRPF (H3K14ac) (44, 92).

In consequence of *Ing5* overexpression, we hypothesized that phthalate treatment could have an impact on the chromatin status. For this reason, we exposed the human thyroid follicular epithelial cell line Nthy-ori 3-1 to DEHP or DINP 1 µM for 5 days and analyzed the genomic accessibility by ATAC-Seq. We reasoned that, since *Ing5* is not a thyroid-specific gene and its expression is not limited to the thyroid, we would be able to observe changes also in a different cell model since we would be investigating a general phthalate mechanism rather than a model-specific response. Differential accessibility analysis resulted in a general increase in accessibility in the treatment group, but none of the identified regions passed multiple-testing correction. Among the regions with a p-value < 0.01, we identified four common ones with an increased accessibility in the DEHP vs DMSO and DINP vs DMSO comparison localized on regulatory regions in the TSS or TTS. However, the signal did not seem to reflect a real change in accessibility.

In this work, we showed that even with a stem cell-derived *in vitro* thyroid model exposed to a range of low, biologically relevant concentrations of four phthalates, we were able to detect some of the effects that have been previously reported *in vivo*. Our analysis demonstrates

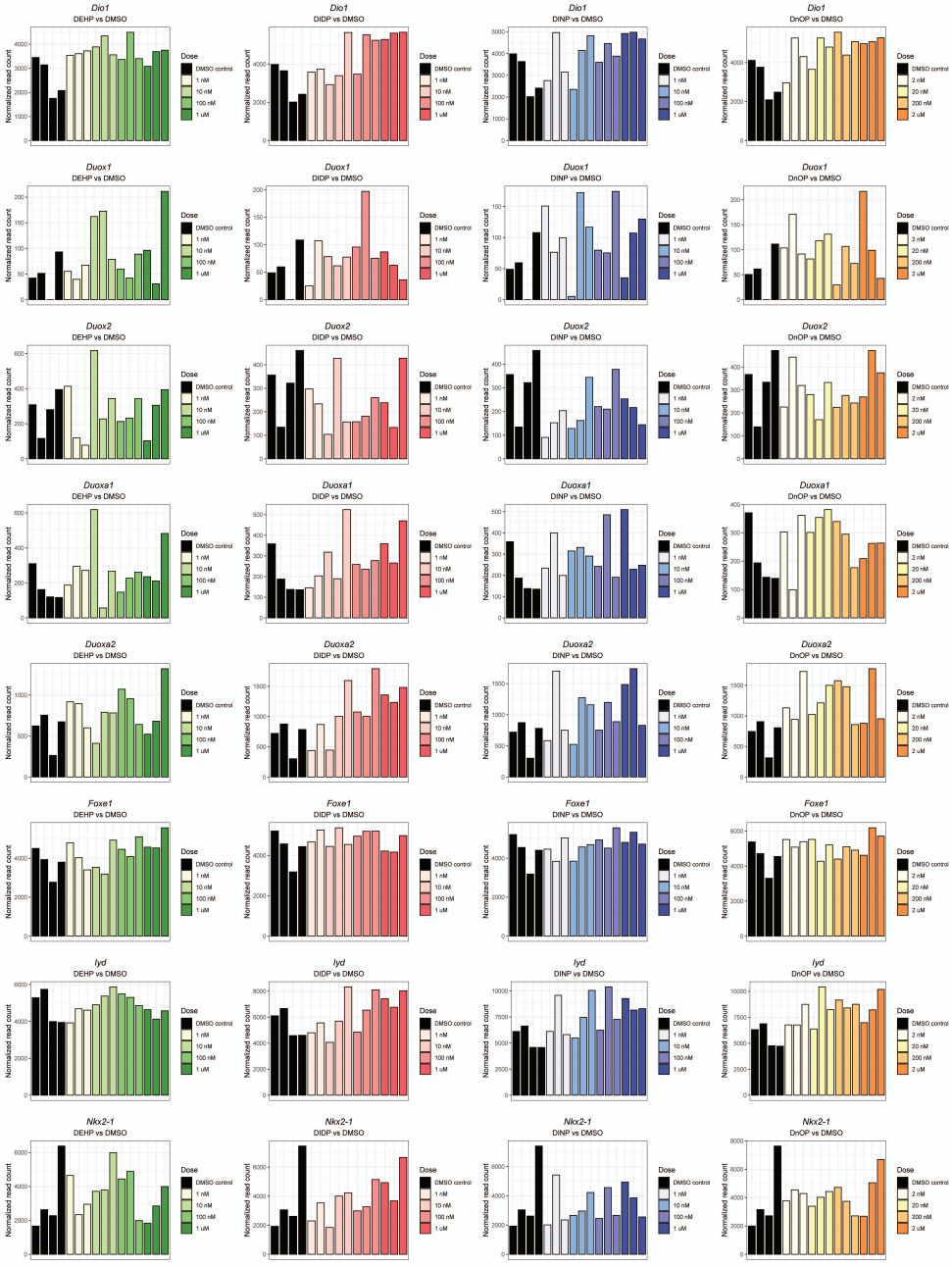
that it is not necessary to use cytotoxic doses in toxicological experiments to obtain observable results and that low dose exposure can be analyzed without lowering the statistical stringency. We are convinced that 3D *in vitro* systems, such as organoids, can be a valid alternative to animal studies even for EDCs, provided that enough dataset are generated to allow regulators to infer risk assessment thresholds.

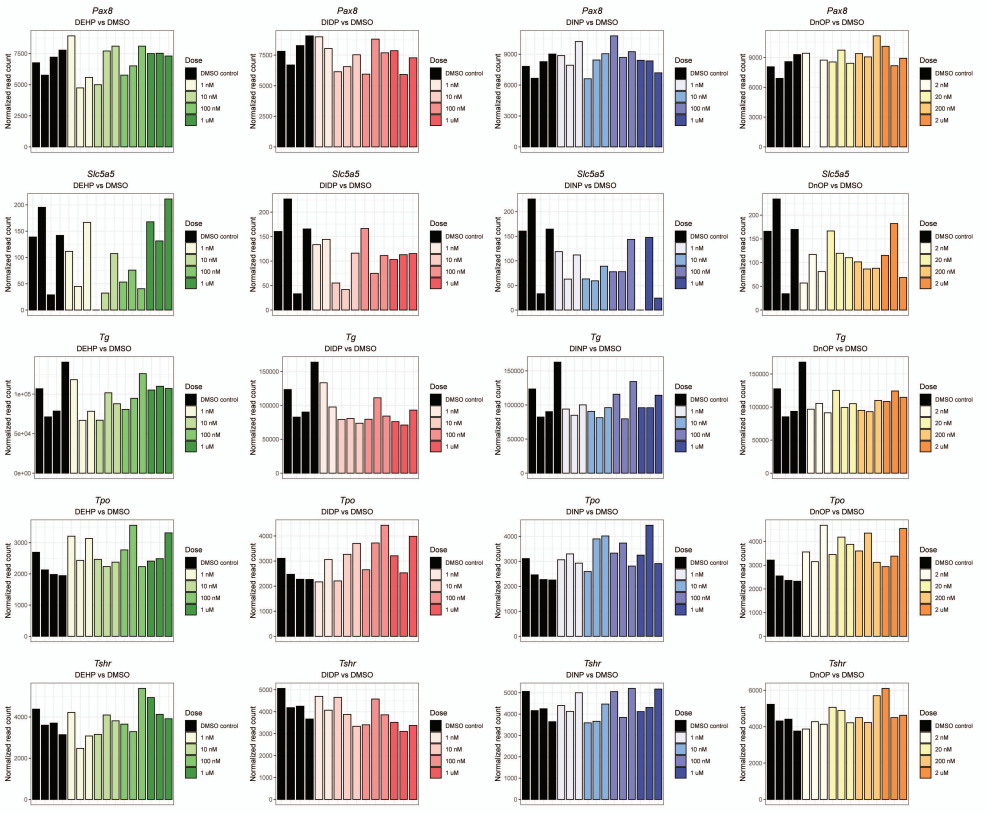
5. Supplementary Material

Supplementary table 1. RNA-Seq samples produced in the study and the number of PCR cycles performed during Combo-Seq library preparation.

Same name	# PCR cycles
DEHP_10_uM_1	16
DEHP_10_uM_2	16
DEHP_1_uM_1	16
DEHP_1_uM_2	16
DEHP_1_uM_3	16
DEHP_100_nM_1	16
DEHP_100_nM_2	14
DEHP_100_nM_3	14
DEHP_10_nM_1	16
DEHP_10_nM_2	16
DEHP_10_nM_3	16
DEHP_1_nM_1	16
DEHP_1_nM_2	16
DEHP_1_nM_3	16
DIDP_10_uM_1	16
DIDP_10_uM_2	16
DIDP_10_uM_3	16
DIDP_1_uM_1	14
DIDP_1_uM_2	14
DIDP_1_uM_3	14
DIDP_100_nM_1	16
DIDP_100_nM_2	16
DIDP_100_nM_3	16
DIDP_10_nM_1	14
DIDP_10_nM_2	16
DIDP_10_nM_3	16
DIDP_1_nM_1	16
DIDP_1_nM_2	16
DMSO_ctrl_1	16
DMSO_ ctrl_2	16
DMSO_ ctrl_3	16
DMSO_ ctrl_4	16
DMSO_ ctrl_5	16

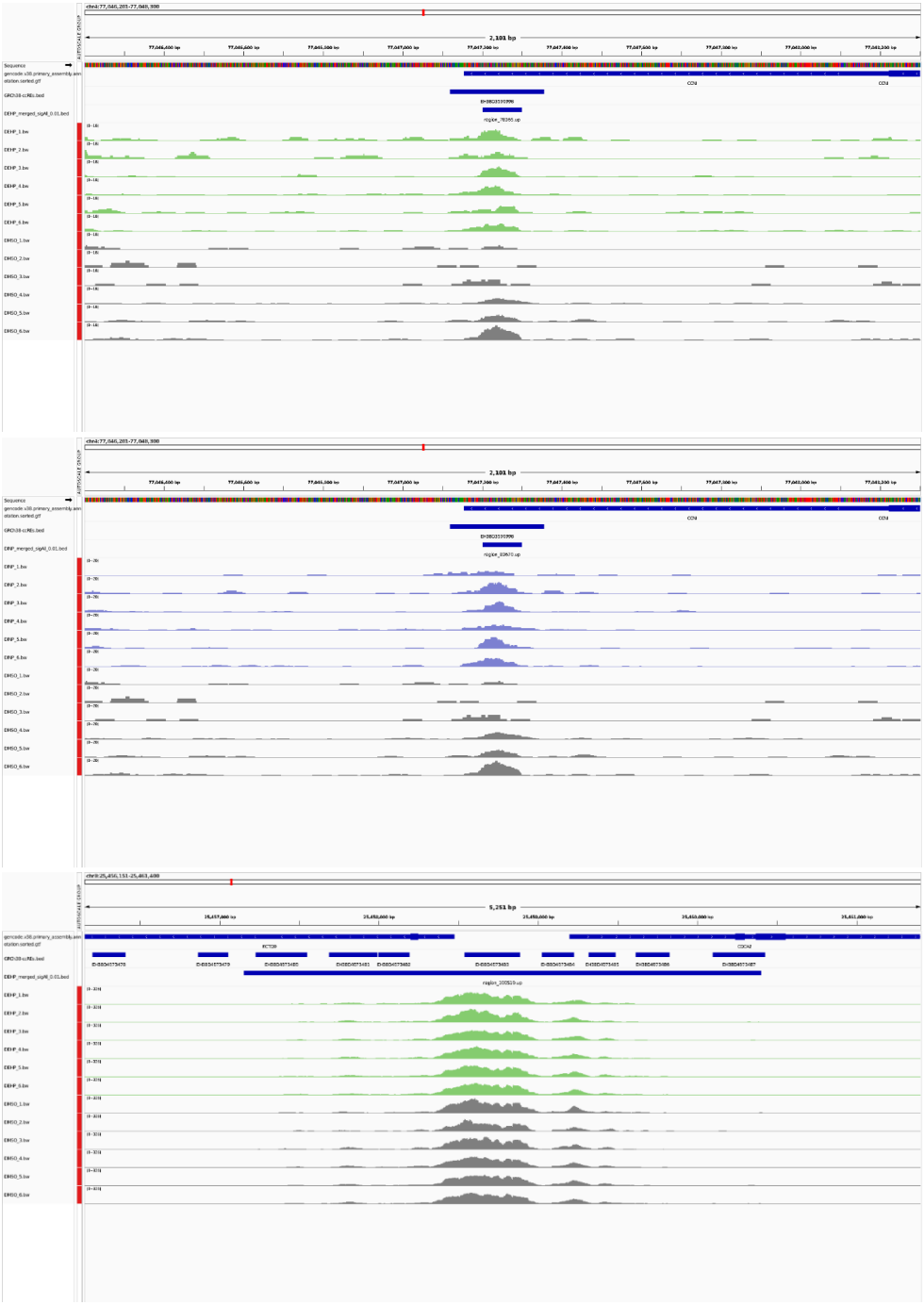
Same name	# PCR cycles
DINP_10_uM_1	16
DINP_10_uM_2	16
DINP_10_uM_3	16
DINP_1_uM_1	16
DINP_1_uM_2	16
DINP_1_uM_3	16
DINP_100_nM_1	16
DINP_100_nM_2	14
DINP_100_nM_3	16
DINP_10_nM_1	14
DINP_10_nM_2	16
DINP_10_nM_3	16
DINP_1_nM_1	16
DINP_1_nM_2	16
DINP_1_nM_3	16
DnOP_20_uM_1	16
DnOP_20_uM_2	16
DnOP_20_uM_3	16
DnOP_2_uM_1	16
DnOP_2_uM_2	16
DnOP_2_uM_3	16
DnOP_200_nM_1	16
DnOP_200_nM_2	16
DnOP_200_nM_3	14
DnOP_20_nM_1	16
DnOP_20_nM_2	14
DnOP_20_nM_3	16
DnOP_2_nM_1	16
DnOP_2_nM_2	16
DnOP_2_nM_3	16
Untr_2	16
Untr_3	16

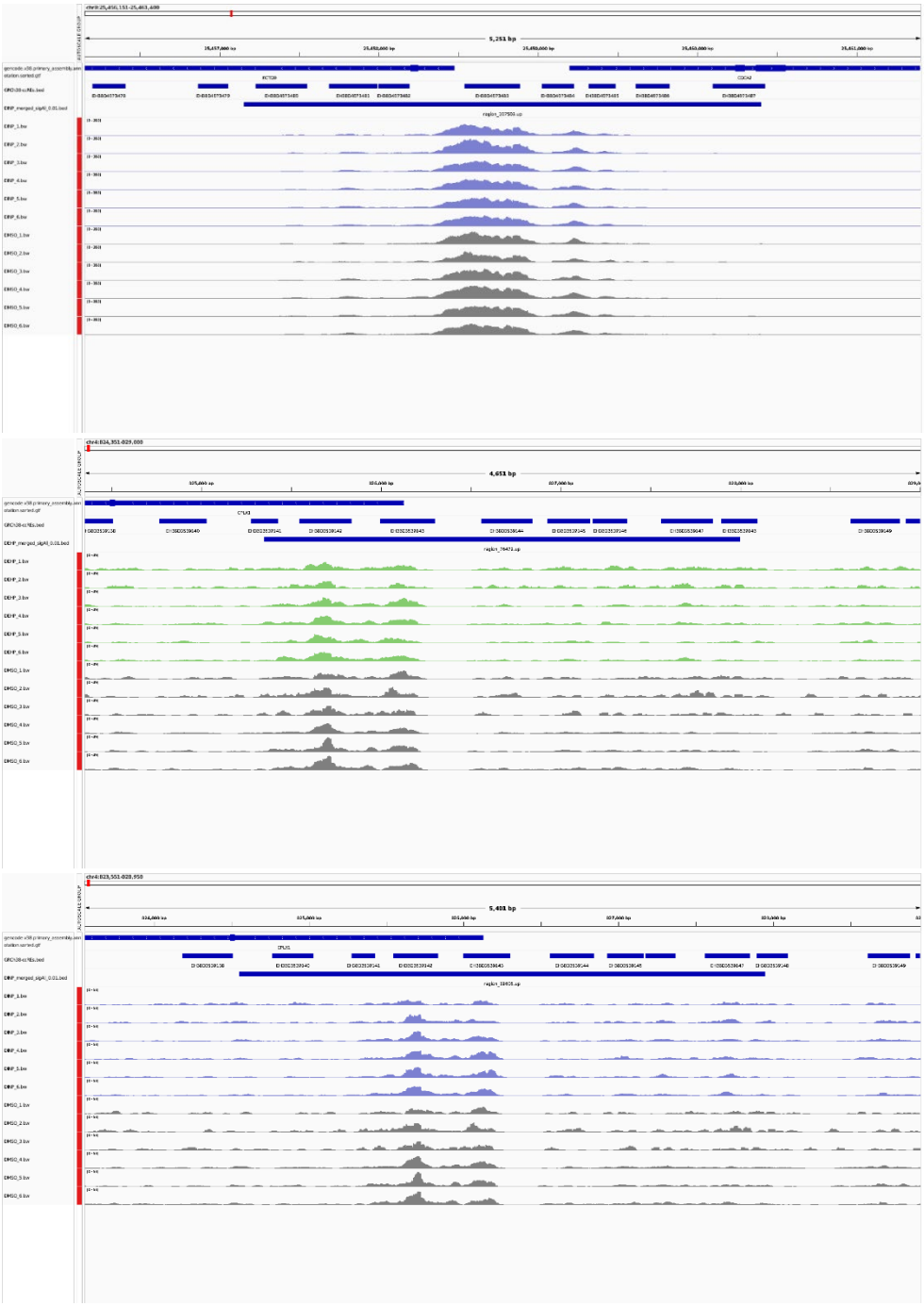

Dio1
DEHP vs DMSO
DIDP vs DMSO
DINP vs DMSO
DnOP vs DMSO
Duox1
Duox2
Duoxa1
Duoxa2
Foxe1
Iyd
Nkx2-1
Normalized read count
Dose
DMSO control
1 nM
10 nM
100 nM
1 uM
2 nM
20 nM
200 nM
2 uM

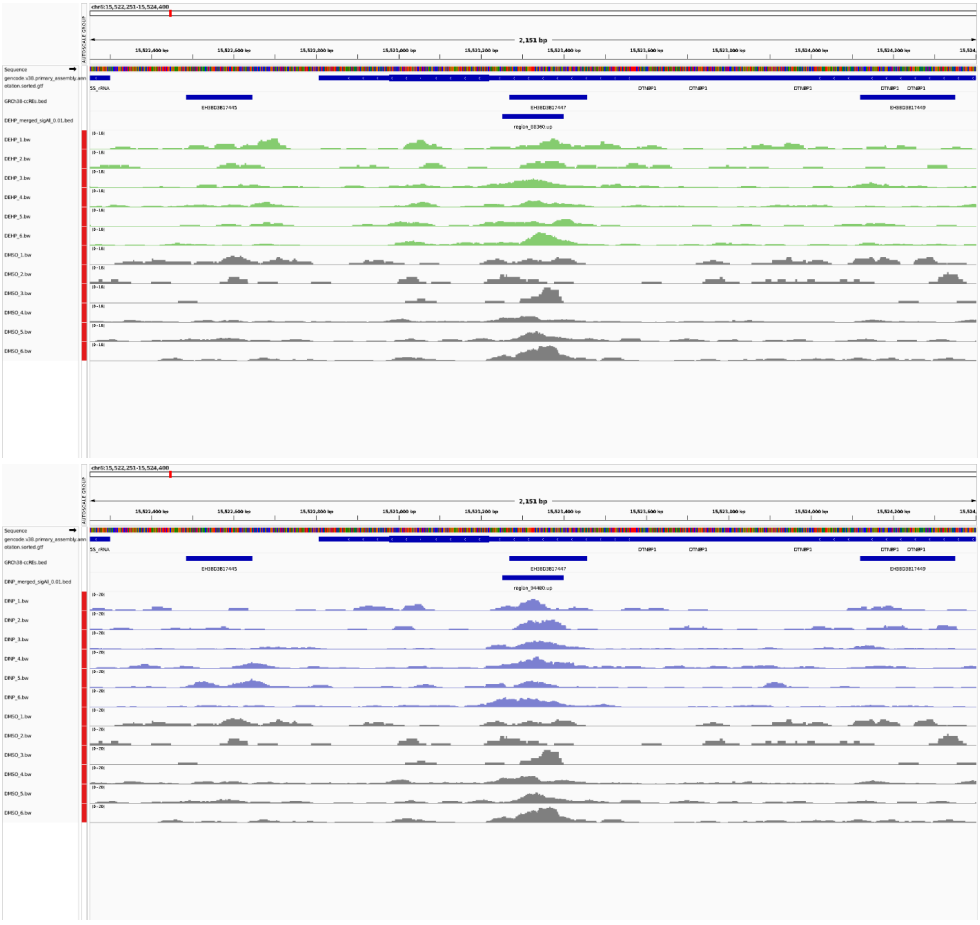


Supplementary figure 1. Barplots of normalized expression of some thyroid markers in each phthalate and DMSO control samples. the gene name is indicated on top of each plot and the different doses are reported in the legend. The darkest shade corresponds to the highest concentration (1 or 2 μM), while the lightest shades to the lowest (1 or 2 nM). The DMSO control samples are in black.

Supplementary figure 2. IGV screenshots of the four regions identified in Table 4. The tracks have been normalized for visualization with BeCorrect (green = DEHP, violet = DINP, grey = DMSO).







6. Data Availability

The sequencing data used in this manuscript has been deposited in BioStudies under the accession E-MTAB-12830.

7. Funding

The SCREENED project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825745.

8. Ethical Standards

The manuscript does not contain clinical studies or patient data.

9. Financial Interests

The authors have no relevant financial or non-financial interests to disclose.

10. References

1. Babich MA. Overview of phthalates toxicity. US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/phthalover.pdf.
2. Carlson KR. Toxicity review of Di(2-ethylhexyl) Phthalate (DEHP). US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/ToxicityReviewOfDEHP.pdf.
3. U.S EPA. Phthalates Action Plan2012. Available from: https://www.epa.gov/sites/default/files/2015-09/documents/phthalates_actionplan_revised_2012-03-14.pdf.
4. Holland M. Socio-economic assessment of phthalates. 2018, doi:https://doi.org/10.1787/a38a0e34-en.
5. Phthalates NRCUCotHRo. 2 - Phthalate Exposure Assessment in Humans. Phthalates and Cumulative Risk Assessment: The Tasks Ahead. Washington (DC): National Academies Press (US); 2008. <https://www.ncbi.nlm.nih.gov/books/NBK215044/>.
6. Wittassek M, Koch HM, Angerer J, Bruning T. Assessing exposure to phthalates - the human biomonitoring approach. Mol Nutr Food Res. 2011;55(1):7-31, doi:https://doi.org/10.1002/mnfr.201000121.
7. Versar Inc. Review of Exposure Data and Assessments for Select Dialkyl Ortho-PhthalatesFebruary 2010. Available from: https://www.cpsc.gov/s3fs-public/pthalexp.pdf.
8. (CDC) CfDCaP. Fourth National Report on Human Exposure to Environmental Chemicals, Updated Tables. January 2019;1, doi:https://doi.org/10.15620/cdc75822.
9. Haines DA, Saravanabhavan G, Werry K, Khoury C. An overview of human biomonitoring of environmental chemicals in the Canadian Health Measures Survey: 2007-2019. Int J Hyg Environ Health. 2017;220(2 Pt A):13-28, doi:https://doi.org/10.1016/j.ijheh.2016.08.002.
10. Wittassek M, Angerer J. Phthalates: metabolism and exposure. Int J Androl. 2008;31(2):131-8, doi:https://doi.org/10.1111/j.1365-2605.2007.00837.x.
11. Koch HM, Angerer J. Chapter 3A. Phthalates: Biomarkers and Human Biomonitoring. Biomarkers and Human Biomonitoring. Issues in Toxicology. Cambridge: Royal Society of Chemistry; 2011. p. 179-233.
12. Net S, Sempere R, Delmont A, Paluselli A, Ouddane B. Occurrence, fate, behavior and ecotoxicological state of phthalates in different environmental matrices. Environ Sci Technol. 2015;49(7):4019-35, doi:https://doi.org/10.1021/es505233b.
13. Consumer Product Safety Improvement Act of 2008, Title I—Children's Product Safety, Sec. 108, Public Law 110-314 (2008). <https://www.cpsc.gov/s3fs-public/pdfs/blk_pdf_cpsia.pdf>.
14. OJ. L 344, 27.12.2005, p. 43.
15. Babich MA, Osterhout CA. Toxicity review of Diisononyl Phthalate (DINP). US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/ToxicityReviewOfDINP.pdf.

16. Carlson KR. Toxicity review of Di-n-Octyl Phthalate (DnOP). US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. March 2010. Available from: https://www.cpsc.gov/s3fs-public/pdfs/blk_media_toxicityDNOP.pdf.
17. Osterhout CA. Toxicity Review of Di(isodecyl) Phthalate. US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/toxicityDIDP.pdf.
18. Phthalates NRCUCotHRo. 3 - Toxicity Assessment. Phthalates and Cumulative Risk Assessment: The Tasks Ahead. Washington (DC): National Academies Press (US); 2008. <https://www.ncbi.nlm.nih.gov/books/NBK215030/>.
19. Fletcher EJ, Santacruz-Márquez R, Mourikes VE, Neff AM, Laws MJ, Flaws JA. Effects of Phthalate Mixtures on Ovarian Folliculogenesis and Steroidogenesis. Toxics. 2022;10(5):251, doi:https://doi.org/10.3390/toxics10050251.
20. OECD. Detailed Review Paper on the State of the Science on Novel In Vitro and In Vivo Screening and Testing Methods and Endpoints for Evaluating Endocrine Disruptors. OECD Series on Testing and Assessment, No 178, OECD Publishing, Paris. 2012, doi:https://doi.org/10.1787/9789264221352-en.
21. Radke EG, Braun JM, Meeker JD, Cooper GS. Phthalate exposure and male reproductive outcomes: A systematic review of the human epidemiological evidence. Environ Int. 2018;121(Pt 1):764-93, doi:https://doi.org/10.1016/j.envint.2018.07.029.
22. Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, et al. EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. Endocr Rev. 2015;36(6):E1-E150, doi:https://doi.org/10.1210/er.2015-1010.
23. Lind PM, Lind L. Endocrine-disrupting chemicals and risk of diabetes: an evidence-based review. Diabetologia. 2018;61(7):1495-502, doi:https://doi.org/10.1007/s00125-018-4621-3.
24. Tuculina MJ, Perlea P, Gheorghita M, Cumpata CN, Dascalu IT, Turcu A, et al. Diabetes mellitus: Plasticizers and nanomaterials acting as endocrine-disrupting chemicals (Review). Exp Ther Med. 2022;23(4):288, doi:https://doi.org/10.3892/etm.2022.11217.
25. Cheng SY, Leonard JL, Davis PJ. Molecular aspects of thyroid hormone actions. Endocr Rev. 2010;31(2):139-70, doi:https://doi.org/10.1210/er.2009-0007.
26. Hiller-Sturmhöfel S, Bartke A. The endocrine system: an overview. Alcohol Health Res World. 1998;22(3):153-64.
27. Benvenga S, Tuccari G, Ieni A, Vita R. Thyroid gland: anatomy and physiology. Encyclopedia of Endocrine Diseases. 2018;4:382-90, doi:https://doi.org/10.1016/B978-0-12-801238-3.96022-7.
28. Carvalho DP, Dupuy C. Thyroid hormone biosynthesis and release. Mol Cell Endocrinol. 2017;458:6-15, doi:https://10.1016/j.mce.2017.01.038.
29. Bereketoglu C, Pradhan A. Plasticizers: negative impacts on the thyroid hormone system. Environ Sci Pollut Res Int. 2022;29(26):38912-27, doi:https://doi.org/10.1007/s11356-022-19594-0.
30. Ghisari M, Bonefeld-Jorgensen EC. Effects of plasticizers and their mixtures on estrogen receptor and thyroid hormone functions. Toxicol Lett. 2009;189(1):67-77, doi:https://doi.org/10.1016/j.toxlet.2009.05.004.

31. Kim MJ, Kim HH, Song YS, Kim OH, Choi K, Kim S, et al. DEHP Down-Regulates Tshr Gene Expression in Rat Thyroid Tissues and FRTL-5 Rat Thyrocytes: A Potential Mechanism of Thyroid Disruption. Endocrinol Metab (Seoul). 2021;36(2):447-54, doi:https://doi.org/10.3803/EnM.2020.920.
32. Wu H, Zhang W, Zhang Y, Kang Z, Miao X, Na X. Novel insights into di-(2-ethylhexyl)phthalate activation: Implications for the hypothalamus-pituitary-thyroid axis. Mol Med Rep. 2021;23(4), doi:https://doi.org/10.3892/mmr.2021.11930.
33. Manikkam M, Tracey R, Guerrero-Bosagna C, Skinner MK. Plastics derived endocrine disruptors (BPA, DEHP and DBP) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. PLoS One. 2013;8(1):e55387, doi:https://doi.org/10.1371/journal.pone.0055387.
34. Martinez-Arguelles DB, Papadopoulos V. Prenatal phthalate exposure: epigenetic changes leading to lifelong impact on steroid formation. Andrology. 2016;4(4):573-84, doi:https://doi.org/10.1111/andr.12175.
35. Guida N, Laudati G, Galgani M, Santopaolo M, Montuori P, Triassi M, et al. Histone deacetylase 4 promotes ubiquitin-dependent proteasomal degradation of Sp3 in SH-SY5Y cells treated with di(2-ethylhexyl)phthalate (DEHP), determining neuronal death. Toxicol Appl Pharmacol. 2014;280(1):190-8, doi:https://doi.org/10.1016/j.taap.2014.07.014.
36. Jacobs MN, Marczylo EL, Guerrero-Bosagna C, Rüegg J. Marked for Life: Epigenetic Effects of Endocrine Disrupting Chemicals. Annual Review of Environment and Resources. 2017;42(1):105-60, doi:https://doi.org/10.1146/annurev-environ-102016-061111.
37. Ogundipe VML, Plukker JTM, Links TP, Coppes RP. Thyroid Gland Organoids: Current Models and Insights for Application in Tissue Engineering. Tissue Eng Part A. 2022;28(11-12):500-10, doi:https://doi.org/10.1089/ten.tea.2021.0221.
38. Li Y, Tang P, Cai S, Peng J, Hua G. Organoid based personalized medicine: from bench to bedside. Cell Regen. 2020;9(1):21, doi:https://doi.org/10.1186/s13619-020-00059-z.
39. Sondorp LHJ, Ogundipe VML, Groen AH, Kelder W, Kemper A, Links TP, et al. Patient-Derived Papillary Thyroid Cancer Organoids for Radioactive Iodine Refractory Screening. Cancers (Basel). 2020;12(11), doi:https://doi.org/10.3390/cancers12113212.
40. Caipa Garcia AL, Arlt VM, Phillips DH. Organoids for toxicology and genetic toxicology: applications with drugs and prospects for environmental carcinogenesis. Mutagenesis. 2022;37(2):143-54, doi: https://doi.org/10.1093/mutage/geab023.
41. W. M. S. Russell, Burch RL. The Principles of Humane Experimental Technique. Medical Journal of Australia. 1960;1(13):500-, doi:https://doi.org/10.5694/j.1326-5377.1960.tb73127.x.
42. Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao XH, Dumitrescu AM, et al. Generation of functional thyroid from embryonic stem cells. Nature. 2012;491(7422):66-71, doi:https://doi.org/10.1038/nature11525.
43. Doyon Y, Cayrou C, Ullah M, Landry AJ, Cote V, Selleck W, et al. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell. 2006;21(1):51-64, doi:https://doi.org/10.1016/j.molcel.2005.12.007.

44. Lalonde ME, Avvakumov N, Glass KC, Joncas FH, Saksouk N, Holliday M, et al. Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. Genes Dev. 2013;27(18):2009-24, doi:https://doi.org/10.1101/gad.223396.113.
45. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods. 2017;14(10):959-62, doi:https://doi.org/10.1038/nmeth.4396.
46. Nazzari M, Hauser D, van Herwijnen M, Romitti M, Carvalho DJ, Kip AM, Caiment F. CODA: a combo-Seq data analysis workflow. Brief Bioinform. 2022;24(1):1-14, doi:https://doi.org/10.1093/bib/bbac582.
47. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1): 10-2, doi:https://doi.org/10.14806/ej.17.1.200.
48. PerkinElmer Inc. NEXTFLEX® Combo-Seq Analysis Guidelines 2020 [Available from: https://perkinelmer-appliedgenomics.com/wp-content/uploads/2020/06/NOVA-5139-AG_v01_NEXTFLEX-Combo-seq-Analysis-Guideline.pdf.
49. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323, doi:https://doi.org/10.1186/1471-2105-12-323.
50. Li B. rsem-prepare-reference documentation page [Available from: https://deweylab.github.io/RSEM/rsem-prepare-reference.html.
51. Patil AH, Halushka MK. miRge3.0: a comprehensive microRNA and tRF sequencing analysis pipeline. NAR Genom Bioinform. 2021;3(3):lqab068, doi:https://doi.org/10.1093/nargab/lqab068.
52. Smith JP, Corces MR, Xu J, Reuter VP, Chang HY, Sheffield NC. PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. NAR Genom Bioinform. 2021;3(4):lqab101, doi:https://doi.org/10.1093/nargab/lqab101.
53. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9, doi:https://doi.org/10.1038/nmeth.1923.
54. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):1-4, doi:https://doi.org/10.1093/gigascience/giab008.
55. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
56. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184-91, doi:https://doi.org/10.1038/nprot.2009.97
57. Conesa A, Nueda MJ. maSigPro: Significant Gene Expression Profile Differences in Time Course Gene Expression Data. R package version 1.70.0. 1.70.0 ed2022.
58. Conesa A, Nueda MJ. maSigPro User's Guide4 September 2017. Available from: https://www.bioconductor.org/packages/release/bioc/vignettes/maSigPro/inst/doc/maSigProUsersGuide.pdf.

59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550, doi:https://doi.org/10.1186/s13059-014-0550-8.
60. CEFIC C4 team. Omics Data Analysis Framework for Regulatory application (R-ODAF) 2021 [Available from: https://github.com/R-ODAF/Main.
61. Verheijen MC, Meier MJ, Asensio JO, Gant TW, Tong W, Yauk CL, Caiment F. R-ODAF: Omics data analysis framework for regulatory application. Regul Toxicol Pharmacol. 2022;131:105143, doi:https://10.1016/j.yrtph.2022.105143.
62. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014;32(9):896-902, doi:https://doi.org/10.1038/nbt.2931.
63. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12(2):477-9, doi:https://doi.org/10.1039/C5MB00663E.
64. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50(D1):D687-D92, doi:https://doi.org/10.1093/nar/gkab1028.
65. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5, doi:http://doi.org/10.1093/nar/gkw257.
66. Lun AT, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. 2016;44(5):e45, doi:https://doi.org/10.1093/nar/gkv1191.
67. Sheikh AA, Blais A. Improved sensitivity and resolution of ATAC-seq differential DNA accessibility analysis. bioRxiv. 2022:2022.03.16.484118, doi:https://doi.org/10.1101/2022.03.16.484118.
68. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40, doi:https://doi.org/10.1093/bioinformatics/btp616.
69. Lun ATL, Smyth GK. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. Nucleic Acids Res. 2014;42(11):e95, doi:https://doi.org/10.1093/nar/gku351.
70. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89, doi:https://doi.org/10.1016/j.molcel.2010.05.004.
71. Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Ai R, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583(7818):699-710, doi:https://doi.org/10.1038/s41586-020-2493-4.
72. Gontarz P, Fu S, Xing X, Liu S, Miao B, Bazylianska V, et al. Comparison of differential accessibility analysis strategies for ATAC-seq data. Scientific Reports. 2020;10(1):10150, doi:https://doi.org/10.1038/s41598-020-66998-4.

73. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nature Biotechnology. 2011;29(1):24-6, doi:https://doi.org/10.1038/nbt.1754.
74. Johnson GE, Soeteman-Hernandez LG, Gollapudi BB, Bodger OG, Dearfield KL, Heflich RH, et al. Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen. 2014;55(8):609-23, doi:https://doi.org/10.1002/em.21870.
75. Terms and Definitions 2023 [Available from: https://www.encodeproject.org/data-standards/terms/.
76. Fromme H, Gruber L, Schlummer M, Wolz G, Bohmer S, Angerer J, et al. Intake of phthalates and di(2-ethylhexyl)adipate: results of the Integrated Exposure Assessment Survey based on duplicate diet samples and biomonitoring data. Environ Int. 2007;33(8):1012-20, doi:https://doi.org/10.1016/j.envint.2007.05.006.
77. OECD. ENV/JM/MONO(2013)6 Revised Guidance Document on Developing and Assessing Adverse Outcome Pathways Series on Testing & Assessment, No 184, 27-Jul-2017. <https://one.oecd.org/document/env/jm/mono(2013)6/en/pdf>.
78. Sturla SJ. Point of Departure. Chem Res Toxicol. 2018;31(1):2-3, doi:https://doi.org/10.1021/acs.chemrestox.7b00341.
79. Wada M, Goto Y, Tanaka T, Okada R, Moriya S, Idichi T, et al. RNA sequencing-based microRNA expression signature in esophageal squamous cell carcinoma: oncogenic targets by antitumor miR-143-5p and miR-143-3p regulation. J Hum Genet. 2020;65(11):1019-34, doi:https://doi.org/10.1038/s10038-020-0795-x.
80. Sanada H, Seki N, Mizuno K, Misono S, Uchida A, Yamada Y, et al. Involvement of Dual Strands of miR-143 (miR-143-5p and miR-143-3p) and Their Target Oncogenes in the Molecular Pathogenesis of Lung Adenocarcinoma. Int J Mol Sci. 2019;20(18):4482, doi:https://doi.org/10.3390/ijms20184482.
81. Li D, Hu J, Song H, Xu H, Wu C, Zhao B, et al. miR-143-3p targeting LIM domain kinase 1 suppresses the progression of triple-negative breast cancer cells. Am J Transl Res. 2017;9(5):2276-85.
82. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54:1 30 1-1 3, doi: https://doi.org/10.1002/cpbi.5.
83. Godi A, Di Campli A, Konstantakopoulos A, Di Tullio G, Alessi DR, Kular GS, et al. FAPPs control Golgi-to-cell-surface membrane traffic by binding to ARF and PtdIns(4)P. Nat Cell Biol. 2004;6(5):393-404, doi:https://doi.org/10.1038/ncb1119.
84. Hinton RH, Mitchell FE, Mann A, Chescoe D, Price SC, Nunn A, et al. Effects of phthalic acid esters on the liver and thyroid. Environ Health Perspect. 1986;70:195-210, doi:https://doi.org/10.1289/ehp.8670195.
85. Howarth JA, Price SC, Dobrota M, Kentish PA, Hinton RH. Effects on male rats of di-(2-ethylhexyl) phthalate and di-n-hexylphthalate administered alone or in combination. Toxicol Lett. 2001;121(1):35-43, doi:https://doi.org/10.1016/S0378-4274(01)00313-7.

86. Posnack NG, Swift LM, Kay MW, Lee NH, Sarvazyan N. Phthalate exposure changes the metabolic profile of cardiac muscle cells. Environ Health Perspect. 2012;120(9):1243-51, doi:https://doi.org/10.1289/ehp.1205056.
87. Shiseki M, Nagashima M, Pedeux RM, Kitahama-Shiseki M, Miura K, Okamura S, et al. p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity. Cancer Res. 2003;63(10):2373-8.
88. Gao W, Han J. Overexpression of ING5 inhibits HGF-induced proliferation, invasion and EMT in thyroid cancer cells via regulation of the c-Met/PI3K/Akt signaling pathway. Biomed Pharmacother. 2018;98:265-70, doi:https://doi.org/10.1016/j.biopha.2017.12.045.
89. Xin H, Wang C, Chi Y, Liu Z. MicroRNA-196b-5p promotes malignant progression of colorectal cancer by targeting ING5. Cancer Cell International. 2020;20(1):119, doi:https://doi.org/10.1186/s12935-020-01200-3.
90. Xu J, Zhao J, Jiang M, Yang L, Sun M, Wang H. MiR-193 promotes cell proliferation and invasion by ING5/PI3K/AKT pathway of triple-negative breast cancer. Eur Rev Med Pharmacol Sci. 2020;24(6):3122-9, doi:https://doi.org/10.26355/eurrev_202003_20679.
91. Liu XL, Meng J, Zhang XT, Liang XH, Zhang F, Zhao GR, Zhang T. ING5 inhibits lung cancer invasion and epithelial-mesenchymal transition by inhibiting the WNT/beta-catenin pathway. Thorac Cancer. 2019;10(4):848-55, doi:https://doi.org/10.1111/1759-7714.13013.
92. Han J, Lachance C, Ricketts MD, McCullough CE, Gerace M, Black BE, et al. The scaffolding protein JADE1 physically links the acetyltransferase subunit HBO1 with its histone H3-H4 substrate. J Biol Chem. 2018;293(12):4498-509, doi:https://doi.org/10.1074/jbc.RA117.000677

Chapter 4

Multiomics Analysis of the Effects of Endocrine Disrupting Chemicals on Mouse Embryonic Stem Cell-Derived Thyroid Organoids

Marta Nazzari, James C. Waddington, Mírian Romitti, Anna M. Kip, Daniel J. Carvalho, Duncan Hauser, Stephen Pennington, Florian Caiment

Abstract

Endocrine disrupting chemicals are pollutants of human origin that can negatively impact endocrine organs, including the thyroid. Due to the central role of the thyroid hormone in growth and development and body metabolism, the effect of these chemicals on the thyroid is a major environmental concern. However, there is still an incomplete understanding of the direct action of endocrine disrupting chemicals on the thyroid functioning at the cellular level.

We used a range of doses to expose mouse embryonic stem cell-derived thyroid organoids for 24 hours to compounds from four endocrine disrupting chemicals classes (OPFRs: organophosphate flame retardants, phthalates, PCBs: polychlorinated biphenyls, PAHs: polycyclic aromatic hydrocarbons), to methimazole and sodium perchlorate and evaluated gene, miRNA, and protein expression patterns using transcriptomic and proteomic analyses. We used the transcriptomics and proteomics data alone and in combination to perform differential expression analysis followed by gene ontology analysis, and to build a random forest model that would allow us to classify unknown samples if belonging to any of the selected EDC class.

Dose-analysis of the gene and miRNA expression identified the point of departure for several genes, such as Aryl Hydrocarbon Receptor targets after PAHs and PCBs treatment. We additionally observed non-monotonic dose-responses. Mmu-miR-142a-3p was induced by PAHs and we observed a concomitant downregulation of the target protein HMGB1. TPO was upregulated by diisodecyl phthalate. Differential gene expression analysis of the class effects showed an upregulation by PCBs of *Dio1*, *Tpo* and *Tshr*. PAHs dysregulated pathways related to energy production, transcription and translation. There was no overlap among the pathway affected by PCBs. We observed enrichment for only two OPFRs, and both affected mRNA processing and small molecule catabolism. Similarly to OPFRs, only two phthalates showed pathway enrichment, but the terms did not overlap. The response to sodium perchlorate and methimazole was affected by the organoid differentiation batch. Despite the weak overlapping signals observes for 2 of the classes, machine learning applied to the integrated datasets of transcriptomics and proteomics were able to also classify each EDC classes with accuracies between 0.997 and 1.

This study integrates proteomic and transcriptomic data to understand the molecular impact of endocrine disruptors on an *in vitro* thyroid model. We also evaluate the feasibility of using methimazole and perchlorate in short-term omics-based studies.

1. Introduction

Endocrine disrupting chemical (EDCs) is a very broad category of environmental contaminants that can interfere with the endocrine system. They are widespread in the environment and in everyday and household materials, causing the daily exposure of the general population. They are known to cause several health effects and are a cause for concern, due to their ability to affect reproduction and (neuro)development and to be involved in development of cancer, cardiovascular and metabolic diseases such as obesity and diabetes (1).

Among the hormone-producing organs, the thyroid plays a central role in metabolism regulation, homeostasis and growth (2). Thyroid hormone (TH) production starts in the thyroid when the thyroid stimulating hormone (TSH) is released by the adenohypophysis in the bloodstream and binds to the TSH receptors (TSHR) on the thyrocytes surface. TSHR activation induces the synthesis of the two forms of TH: the active form of the TH is triiodothyronine (T3), while tetraiodothyronine (T4) is the inactive form. The thyroid mostly produces T4 (around 80% T4 and 20% T3 (3)), reflected in the very different total plasma concentration of the two types of TH (1.8 nM for T3, 100 nM for T4). After synthesis, the TH is released in the bloodstream, where most of it is bound by proteins to prevent degradation. Thus, the serum concentration of free TH is very low (5 pM for T3 and 20 pM for T4) (4). In the target organs, the TH binds to the thyroid receptor (TR). Once T4 reaches the target organs, it is converted to T3 via removal of one iodine atom by deiodinases (DIO). In human, type 1 deiodinases (DIO1) are located on the plasma membrane and are expressed in the liver, kidney, thyroid and hypophysis, while type 2 deiodinases (DIO2) are located on the endoplasmic reticulum membrane and have a more widespread (but variable in its level) expression in thyroid, heart, brain, spinal cord, skeletal muscle, placenta, skin, retina, cochlea, kidney, brown adipose tissue and pancreas (5-9).

Several classes of compounds are considered EDCs, including phthalates, organophosphate flame retardants (OPFRs), polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs), which will be the focus of this work. Phthalates are alkyl or dialkyl esters of phthalic acid and are used primarily as plasticizers for polyvinylchloride (PVC) product or as solvents (10, 11). As they are not covalently bound to the matrix they are added into, they can leave the material by direct release, evaporation, leaching or abrasion (12). In consequence, human exposure to phthalates can occur via ingestion, inhalation or dermal absorption (12), ingestion via foodstuff being the most prominent way of exposure in the general population (13). OPFRs include a wide range of halogenated and non-halogenated compounds containing phosphorus and are found in engineering plastics, coatings, polyurethane foams and textiles [20]. PAHs comprise a large group of organic compounds composed of two or more benzene rings and containing only carbon and hydrogen (14). They are formed during the incomplete combustion of organic material, including materials employed in energy production

at the industrial or household level, and tobacco smoke. The greatest source of exposure in the general population is thought to be contaminated or burnt food (15). PCBs are organic synthetic compounds that were produced and used until the 1970s as coolants and lubricants in many types of electrical equipment, both in industrial and consumer contexts. Due to their high chemical stability, they break down slowly and remain in the environment for a very long time. In human, PCB contamination mainly occurs via contaminated food and air and when ingested they bioaccumulate due to their lipophilicity (16). For all these four EDC classes there is evidence of interference with the TH system, directly affecting thyrocytes and TH production, but also TH signaling and clearance, competition with TH binding in the serum and interference with DIO activity. Some publications, like biomonitoring studies, report negative associations between exposure to EDCs and decreased TH serum levels, without mechanistic hypotheses ((17) for phthalates, (18, 19) for OPFRs, (20-22) for PAHs, (23-25) for PCBs).

To provide biological insight into the direct effects of EDCs on the thyroid, we exposed mouse embryonic stem-cell (mESC) derived thyroid organoids to 16 EDCs belonging to the classes discussed above (phthalates, PAHs, OPFRs and PCBs) for 24 hours. We selected 5 increasing doses including those identified in biomonitoring studies (1, 10, 100 nM, 1, 10 uM) and analyzed the gene and miRNA content. Additionally, we selected the 10 uM dose to also perform untargeted proteomics analysis. We first analyzed the transcriptomic and proteomic data separately and then integrated them to gain insight on the compounds' effects performing differential expression and gene ontology analyses. We built a random forest (RF) classification model to classify unknown samples to any of these classes. We also exposed the organoids to two compounds able to interfere at with the production of the TH the protein level (methimazole and sodium perchlorate) and assessed their efficacy in evaluating the transcriptomic and proteomic impacts of short-term exposures.

2. Materials and Methods

2.1 Chemical Compounds

The following compounds were used: phthalates: bis(2-ethylhexyl) phthalate (DEHP, CAS 117-81-7, 67261 Sigma-Aldrich), di-iso-decylphthalate (DIDP, CAS 26761-40-0, 80135 Supelco), di-iso-nonylphthalate (DINP, CAS 28553-12-0, 376663 Sigma-Aldrich), di-n-octylphthalate (DnOP, CAS 117-84-0, D201154 Sigma-Aldrich); organophosphate flame retardants (OPFRs) – phosphate esters: tris(1,3-dichloroisopropyl) phosphate (TDCPP, CAS 13674-87-8, 32951 Sigma-Aldrich), triphenyl phosphate (TPP, CAS 115-86-6, 442829 Sigma-Aldrich), bisphenol A bis(diphenyl phosphate) (BADP, CAS 5945-33-5, BD117437 Chempure GMBH); OPFRs – phosphonates: dimethyl methyl phosphonate (DMMP, CAS 756-79-6, D169102 Sigma-Aldrich); polychlorinated biphenyls (PCBs): PCB-118 (CAS 31508-00-6, DRE-C20011800 LGS Standards), PCB-126 (CAS 57465-28-8, AMB22734910 Ambinter), PCB-138 (CAS 35065-28-2, 35494 Sigma-Aldrich), PCB-153 (CAS 35065-27-1, DRE-C20015300 LGS Standards); polycyclic aromatic hydrocarbons (PAHs): benz[a]anthracene (BAA, CAS 56-55-3, B2209 Sigma-Aldrich), benzo[a]pyrene (BAP, CAS 50-32-8, B-1760 Sigma-Aldrich), benzo[k]fluoranthene (BKF, CAS 207-08-9, 392251 Sigma-Aldrich), dibenzo[a,h]anthracene (DAHA, CAS 53-70-3, 33530 Fluka); methimazole (MMI, CAS 60-56-0, M8506 Sigma-Aldrich); sodium perchlorate ($NaClO_4$, CAS 7791-07-3, 71853-M Sigma-Aldrich). Each compound was dissolved in 100% DMSO (1029521000, Merck Millipore), aliquoted in single-use vials and stored at -80 C.

2.2 Organoids Differentiation

2.2.1 Embryonic Stem Cells Culture and Thyroid Organoids Differentiation

Thyroid follicles were differentiated from the A2Lox.Cre_TRE-Nkx2-1/Pax8_Tg-EGFP mouse ESC as previously described [42,47]. Briefly, cells were initially cultured on gamma-irradiated mouse embryonic fibroblasts (MEFs) feeder using mouse stem cell medium [42,47] and incubated at 37 °C, 5% CO2 and > 95% humidity. For differentiation into thyroid, embryoid bodies (EBs) were generated by hanging drops culture of ESCs (1000 cells per drop) for 4 days. They were then collected and embedded in growth factor restricted Matrigel (354230, Corning). 50 mL Matrigel drops containing around 20 EBs were plated into 12-well plates. Cells were differentiated using a differentiation medium composed of DMEM (31966021, Gibco) supplemented with 15% FBS, vitamin C (50 µg/mL) (A4544, Sigma), nonessential amino acids (0.1 mM) (11140035, Gibco), sodium pyruvate (1 mM) (11360039, Gibco), penicillin and

streptomycin (50 U/mL) (15140122, Gibco), 2-mercaptoethanol (0.1 mM) (31350010, Gibco). The differentiation medium was supplemented with 1 µg/mL of doxycycline (D9891-1G, Sigma) for 3 days for *Nkx2-1* and *Pax8* induction, followed by two weeks of maturation by using the differentiation medium containing 8-Br-cAMP (0.3 nM) (B 007-500, BioLog).

2.2.2 Thyroid Follicles Enrichment Protocol

At day 21, after complete thyroid maturation, Matrigel drops containing the thyroid follicles were washed twice with Hanks's balanced salt solution (HBSS, containing calcium and magnesium) (14025050, Gibco) and incubated in a HBSS solution (1 mL per well) containing 10 U/mL of dispase® II (4942078001, Roche) and 125 U/mL of collagenase type IV (Sigma) for 30-45 min at 37 °C. The enzymes were then inactivated by adding 10% FBS. Organoids were centrifuged at 1200 rpm for 3 minutes. After rinsing twice with HBSS, the follicles were enriched by filtration using 30 mm (to remove single cells) (43-50030, pluriSelect Life Science GmbH) to and 100 mm reverse strainer (to remove big follicles aggregates) (43-50100, pluriSelect Life Science GmbH).

Resuspended follicles were cultured in subsequent experiments in the differentiation medium described above and supplemented with 8-Br-cAMP (10 µM) and TGF-βRI inhibitor SB431542 (10 µM) (1614, Tocris), hereafter termed "supplemented differentiation medium".

2.3 Exposures to Endocrine Disrupting Chemicals (EDCs)

Due to the limited availability of follicles that can be obtained in a single differentiation batch, the exposures were performed in three separate experiments (Experiment 1, 2 and 3), in which one or more EDCs classes would be tested (Experiment 1: phthalates; Experiment 2: OPFRs, PAHs, MMI and PER; Experiment 3: PCBs, MMI and PER).

For each exposure, differentiated follicles were plated on 48-well cell culture plates at a density of 1,000 follicles per well for RNA extraction and transcriptomic analysis, and of ~10,000 follicles per well for protein extraction and proteome analysis. Phthalates, OPFRs, PAHs and PCBs solutions were prepared in supplemented differentiation medium at 1 nM, 10 nM, 100 nM, 1 uM or 10 uM for EDCs (2 nM, 20 nM, 200 nM, 2 uM and 20 uM for DnOP) and 10 uM, 100 uM and 1 mM for MMI and PER (n = 3 for each concentration, final [DMSO] = 0.5% of the volume). DMSO solvent (n = 5) and untreated (n = 3) controls were included in each experiment. Due to the greater amount of follicles required for proteomics analysis, samples dedicated to protein collection were exposed only at the highest doses (10 uM for EDCs, 1 mM for MMI and PER) (n = 3, also for DMSO and untreated controls). The experimental setup is

summarized in Figure 1. The plated follicles were incubated at 37 °C, 5% CO_2 and >95% humidity for 24 hours.

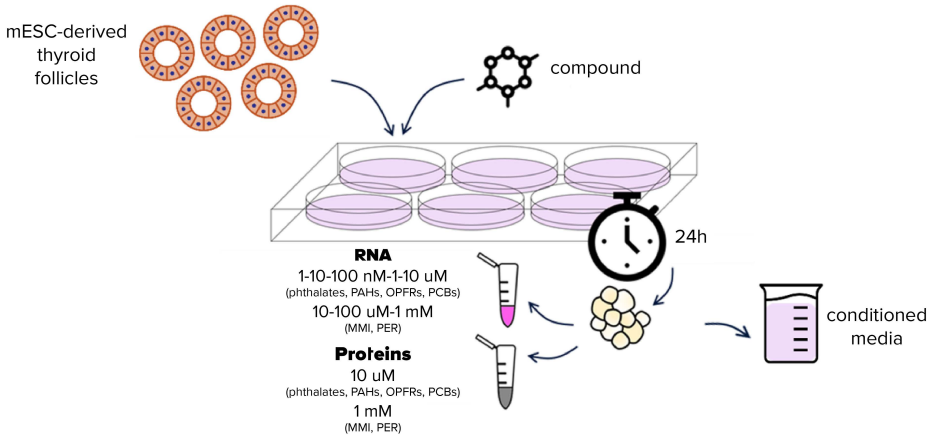


Figure 1. Graphical representation of the exposure regimens of thyroid follicles for transcriptomics proteomics (note: a 48-well plate as used in the experiments). "Compound" can indicate any of the compounds tested, DMSO solvent or media only.

2.4 Transcriptomics Sample Preparation

2.4.1 RNA Isolation

After 24 hours, the follicles were collected, washed once with PBS, and lysed in QIAzol Lysis Reagent (79306, Qiagen). Total RNA was extracted using the miRNAeasy Micro Kit (217084, Qiagen). All samples had a RIN (RNA Integrity Number) of 8 or higher.

2.4.2 RNA-Seq Libraries Preparation

Twenty (20) ng of total RNA were used to prepare RNA-Seq libraries with the NEXTFLEX® Combo-Seq™ mRNA/miRNA Kit (NOVA-5139-53, PerkinElmer). To deplete tRNA fragments and Y RNA fragments, the NEXTFLEX® tRNA/YRNA Blocker (NOVA-51312, PerkinElmer) was used during the library preparation following the manufacturer's instructions. Fourteen (14) cycles of PCR were performed during the protocol for samples collected in Experiment 1. For some samples, the final library concentration was below the pooling concentration used for sequencing (1.6 nM). In these cases, the library was prepared again performing 16 cycles. For the samples collected in Experiments 2 and 3, we performed 16 PCR cycles. Some samples could not be sequenced (Table 1). The prepared libraries were sequenced on an S4 Illumina flowcell 35 cycles (v1.5) (Illumina) in single-end mode.

Table 1. Samples that could not be sequenced, their respective experiment and the reason why they failed.

Sample	Experiment	Reason
DEHP_10_uM_3	1	Low library concentration
DIDP_1_nM_3	1	Low library concentration
Untr_1	1	Low library concentration
BAA_1_nM_3	2	Sequenced, but mislabeled
BAP_10_uM_3	2	Sequenced, but mislabeled
BAP_100_nM_2	2	Sequenced, but mislabeled
BKF_1_nM_3	2	Not enough RNA for library prep
TPP_10_nM_2	2	Sequenced, but mislabeled
DMMP_1_uM_3	2	Sequenced, but mislabeled
DMMP_10_nM_2	2	Sequenced, but mislabeled
BADP_1_uM_1	2	Sequenced, but mislabeled
BADP_10_nM_2	2	Not enough RNA for library prep
BADP_1_nM_2	2	Sequenced, but mislabeled
TDCPP_100_nM_3	2	Sequenced, but mislabeled
PCB138_100_nM_3	3	Low library concentration

2.5 Transcriptomics Data Analysis

2.5.1 RNA-Seq Data Processing

The fastq files were processed according to our previously published CODA pipeline (26). Briefly, reads were trimmed using Cutadapt (v3.7) (27) as indicated in the Combo-Seq manual guidelines (28). To retrieve genes counts, the trimmed reads were aligned both to the mouse transcriptome (GRCm39 v27) and quantified using RSEM (v1.3.3) with the "--STAR" parameter (v2.7.10a), following the ENCODE3's STAR-RSEM pipeline (29, 30). To retrieve miRNAs, miRge3.0 (v0.0.9) (31) was run by inputting the trimmed files using the miRBase mouse annotation (v22).

2.5.2 Biotype Mapping and Outliers Identification

The quantified RNA species were mapped to their respective biotypes using the R (32) package biomaRt (33). We removed the reads mapping to snoRNAs (which were analyzed separately) and calculated the percentage of mapped reads per biotype. We retained only those constituting at least 1% in at least one sample. Outliers for each biotype were identified per treatment group (any compound, DMSO and untreated) as being 1.5 times the interquartile range (IQR) below the 25th percentile or above the 75th percentile:

$$biotype\ x\ in\ sample\ y\ <\ 25th\ percentile\ (biotype\ x\ in\ group\ z) - 1.5 * IQR\ (biotype\ x\ in\ group\ z)$$

or

$$biotype\ x\ in\ sample\ y\ >\ 75th\ percentile\ (biotype\ x\ in\ group\ z) + 1.5 * IQR(biotype\ x\ in\ group\ z).$$

2.5.3 Gene Filtering and Dose Series Analysis

To filter out low expressed genes, we used a modified version of the Omics Data Analysis Framework for regulatory application (R-ODAF) pipeline developed by our group (34, 35). Briefly, genes/miRNAs/snoRNAs ("features") were considered expressed if their normalized expression was ≥1 count per million (CPM) in at least 75% of the samples in either group (i.e. treatment versus control) ("relevance threshold"). The RUVs function from the RUVSeq package (36) with k = 2 was used on the features passing the relevance threshold filter to remove unwanted variation.

MaSigPro (37) was used for dose series analysis using read counts normalized with DESeq2 (38) following the steps highlighted in the maSigPro User's Guide for Next-Generation Sequencing data (39) and with a single series course experiment. The "tetha" (θ) value was set to 10 (default), the FDR to 0.05 and the "degree" parameter to 3 (which corresponds to a cubic polynomial regression model). The variable "Time" with values 0, 1, 2, 3, 4, 5 was used in the model as a mock variable to represent the "Dose" values of 0 (DMSO control), 1 nM, 10 nM, 100 nM, 1 uM and 10 uM (2 nM, 20 nM, 200 nM, 2 uM, 20 uM for DnOP). In the case of MMI and PER analysis, "Time" assumed the values 1, 2, 3 for the doses 10 uM, 100 uM and 1 mM, respectively.

2.5.4 miRNA Targets Identification

The targets of the DE miRNA were retrieved using the miRTarBase release 9.0 for *Mus musculus* (40).

2.5.5 Gene Differential Expression Analysis

Differentially expressed (DE) features were identified by grouping the samples of a single EDC class exposed to 10 uM dose and comparing to the DMSO solvent controls of the corresponding Experiment. We used the R package DESeq2 [56] and applied the "3rd quartile rule" and "spurious spikes" filters described in the R-ODAF pipeline. After identifying the "expressed" features using the "relevance" filter described in paragraph 2.5.3, differential expression analysis was performed setting the FDR to 0.01. We applied the "spurious spikes" and "3rd quartile" filters to the identified DE features. The former was applied to identify those cases in which a very high expression value in only one replicate in a group is responsible for a certain feature to have an FDR < 0.01: we applied the following formula to every DE feature for both treatment and control group:

$$\frac{read\ count\ feature_i}{total\ read\ count\ feature_i\ in\ group_j} < 1.4 \times \left(number\ of\ replicates\ in\ group_j\right)^{-0.66},$$

where *i* refers to any feature, and *j* to either the treatment or control group. The "3rd quartile"

rule retained only those DE features for which the median of either condition (treatment or control) is higher than the 3rd quartile of the other condition, and aimed at replacing a fold change threshold. After DE features identification, the barplots of the normalized counts were manually inspected to determine whether the signal could be a technical artefact or be instead a biologically relevant signal.

2.6 Proteomics Sample Preparation

2.6.1 Protein Samples Processing

At the end of the exposure period, cells were collected in 1.5 mL Eppendorf tubes and pelleted by centrifugation. To each cell pellet, 50 uL lysis buffer (7 M Urea, 2 M Thiourea, 4% CHAPS, 30 mM Tris, pH 8.5) was added and after vortex mixing samples were frozen at -80°C. Samples were shipped to Atturos (ATT) on dry ice, where they were subsequently stored at -80°C until processed further.

Sample preparation of the follicles was performed in two batches, the first containing lysates from Experiment 1 (phthalates) and the second containing lysates from Experiments 2 and 3 (PAHs, OPFRs, PCBs, MMI and PER).

Cells lysates were digested using trypsin on the S-Trap (Protifi) 96-well plate platform. Briefly, lysates were thawed from -80°C and proteins solubilized and denatured using a 10% SDS, 100 mM TEAB (pH 7.55) solution, added at a 1:1 ratio. Proteins were reduced and alkylated using final concentrations of 10 mM DTT and 40 mM Iodoacetamide, respectively. S-Trap binding buffer (9:1 ratio of methanol:TEAB) was added to each sample and loaded onto a 96-well S-Trap plate. After multiple wash steps, sequencing grade modified trypsin (Promega) was used to digest the proteins overnight (16 h) at 37 °C. After digestion peptides were recovered and dried down using a vacuum centrifuge.

Samples were next prepared for discovery with liquid chromatography with tandem mass spectrometry (LC-MS/MS) (Bruker timsTOF Pro coupled to an EvoSep One LC system). Briefly, all follicles cell digests were resuspended in 2% acetonitrile/0.1% formic acid and peptides were quantified using A215 on a DeNovix DS-11 spectrophotometer. Five hundred nanograms of peptide were loaded onto EvoSep EvoTips using the manufacturers' protocol. Samples were introduced onto a Bruker timsTOF in pre-determined worklist run orders. Worklists were flanked by commercially available HeLa cell line digests to monitor instrument performance over time. Each EDC class was run in batches, separated by blank injections, with DMSO control biological replicates injected at the start of each batch (and experiment).

2.6.2 LC-MS/MS Data Acquisition

LC-MS/MS data acquisition of the follicles was also performed in two batches, the first containing peptide digests from Experiment 1 (phthalates) and the second containing peptide digests from Experiments 2 and 3 (PAHs, OPFRs, PCBs, MMI and PER).

The 30 samples per day (SPD) EvoSep method (44-minute gradient) was used for these experiments. The timsTOF Pro mass spectrometer was operated in positive ion polarity with TIMS (Trapped Ion Mobility Spectrometry) and PASEF (Parallel Accumulation Serial Fragmentation) modes enabled. The accumulation and ramp times for the TIMS were both set to 100 ms, with an ion mobility (1/k0) range from 0.62 to 1.46 Vs/cm. Spectra were recorded in the mass range from 100 to 1,700 m/z. The precursor (MS) Intensity Threshold was set to 2,500 and the precursor Target Intensity set to 20,000. Each PASEF cycle consisted of one MS ramp for precursor detection followed by 10 PASEF MS/MS ramps, with a total cycle time of 1.16 s.

Pierce™ HeLa Protein Digest Standard (88329, ThermoFisher) was injected in triplicate at the start and end of both batches worklist to monitor instrument performance over the duration of the run. Follicles samples were run in between the HeLa standards (the full worklists are reported in Supplementary table 1).

2.7 Proteomics Data Analysis

2.7.1 Processing Pipeline

Data acquired from the LC-MS/MS runs were searched using Max Quant v1.6.17.0, again in two batches. The first batch contained MS data from Experiment 1, while the second batch contained MS data from Experiments 2 and 3. Data from all three experiments was not searched together as the "match between runs" feature was used to enhance the number of proteins identified: searching all three experiments together would increase the risk of identifying false positives. The full list of search parameters is shown in Supplementary table 2. HeLa quality control samples were searched using the same parameters, with the human UniProt fasta file substituted with the mouse equivalent. Further data processing was performed using Perseus v1.6.15.0 software, where protein data was Filtered according to internal workflows. First, proteins identified as "potential contaminants" by Max Quant were removed, followed by those that appeared in the reverse protein database. Remaining proteins were exported and analyzed further.

2.7.2 Differential Protein Expression

Univariate statistical analysis was performed using a Student's t-test on DMSO (control) cells vs each compound exposure group. Multiple testing correction was done according

to the Benjamini-Hochberg method and an adjusted p-value of 0.01 was used as a cut off to determine significance. DMSO controls were contained within each Experiment (1, 2 and 3) to compare to their respective class of EDCs. Proteins were only considered valid if they appeared in at least three samples in one of the test groups (i.e., control or exposed).

3. Results

3.1 Transcriptomics and Proteomics Datasets Quality Control (QC)

3.1.1 Transcriptomics QC

Transcriptomics libraries had good coverage, with the median library size being 64.9 million (M) reads (min = 17.7 M, max = 171.6 M) (Figure 2A), and high quality (median 96.1% of reads with quality score >= 30; min = 85.8%, max = 97.4%) (Figure 2B). The distribution of trimmed reads reflected the number of sequenced reads, showing how the trimming was homogeneous across samples (median count of trimmed reads 58.5 M; min = 15.8 M, max = 154.4 M) (Figure 2C). The median number of reads mapping to mRNA (Gencode annotation) and miRNAs (miRBase annotations) was 49.2 M and 1 M, respectively (mRNA: min = 14.3 M, max = 138.3 M; miRNA: min = 0.16 M, max = 4.3 M) (Figure 2D-E).

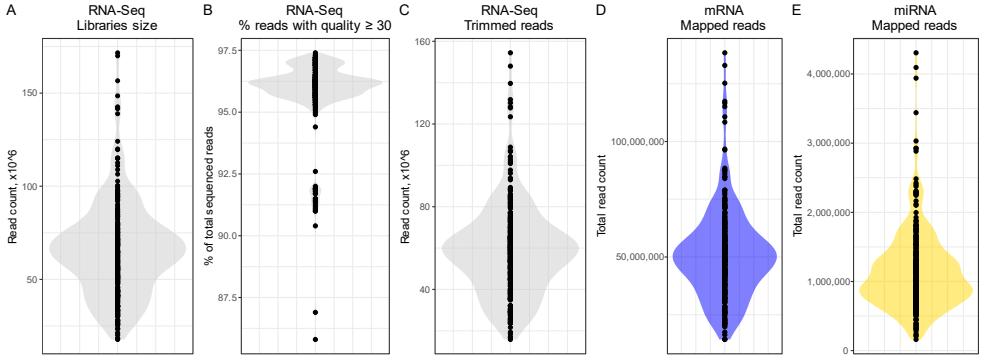
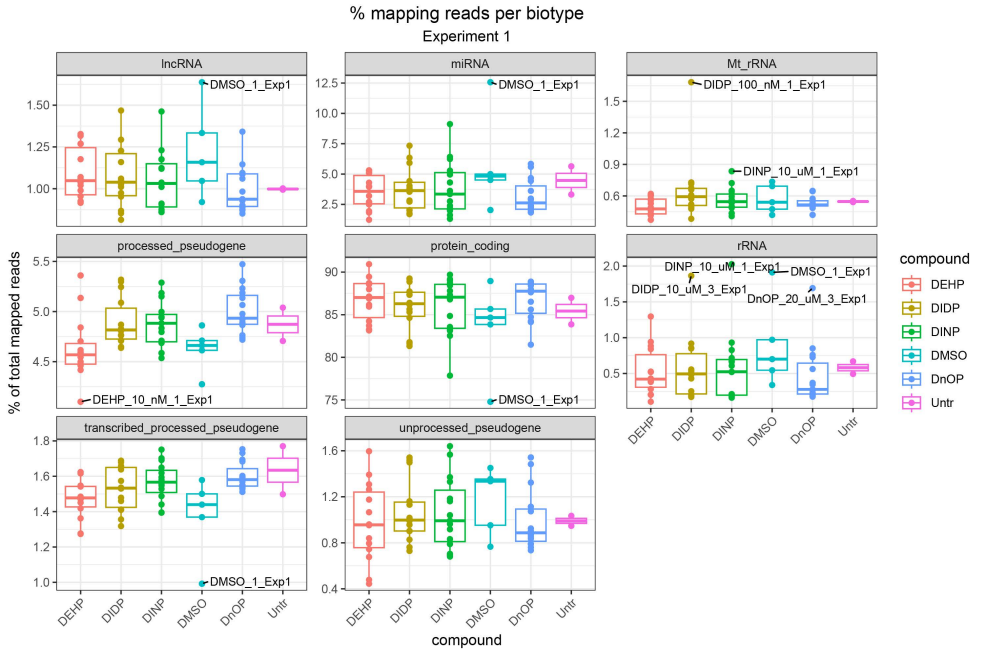


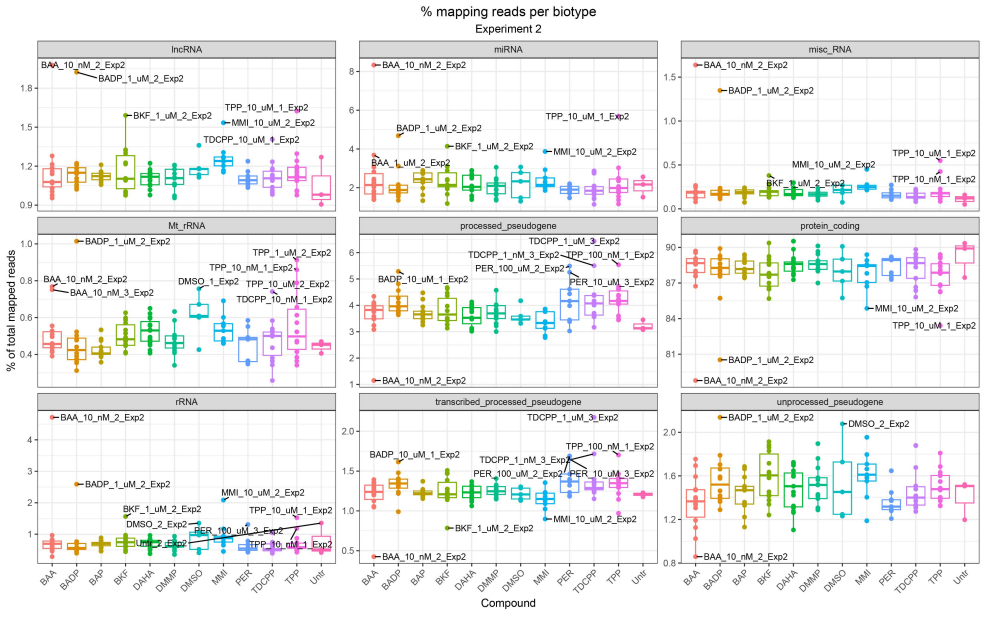
Figure 2. RNA-Seq quality control metrics: (A) libraries size, (B) percentage of sequenced reads with quality score >= 30; (C) number of trimmed reads; (D) reads mapped to mRNA using the mouse Gencode annotations; (E) reads mapped to miRNAs using the mouse miRBase annotations.

To identify possible outliers, we analyzed the percentage of RNA biotype composition in the mapped reads after small nucleolar RNAs (snoRNAs) removal (which were analyzed separately) (Figure 3). Among the outliers, we identified DMSO replicate 1 (Experiment 1) and BAA 10 nM replicate 2 (Experiment 2) to be outliers in multiple biotypes, so we removed them from our analyses. Although other samples were flagged as outliers in some biotypes, it was not frequent and for this reason they were not removed.

A


% mapping reads per biotype
Experiment 1
lncRNA
miRNA
Mt_rRNA
processed_pseudogene
protein_coding
rRNA
transcribed_processed_pseudogene
unprocessed_pseudogene
% of total mapped reads
compound
DEHP
DIDP
DINP
DMSO
DnOP
Untr

B


% mapping reads per biotype
Experiment 2
lncRNA
miRNA
misc_RNA
Mt_rRNA
processed_pseudogene
protein_coding
rRNA
transcribed_processed_pseudogene
unprocessed_pseudogene
% of total mapped reads
Compound

C

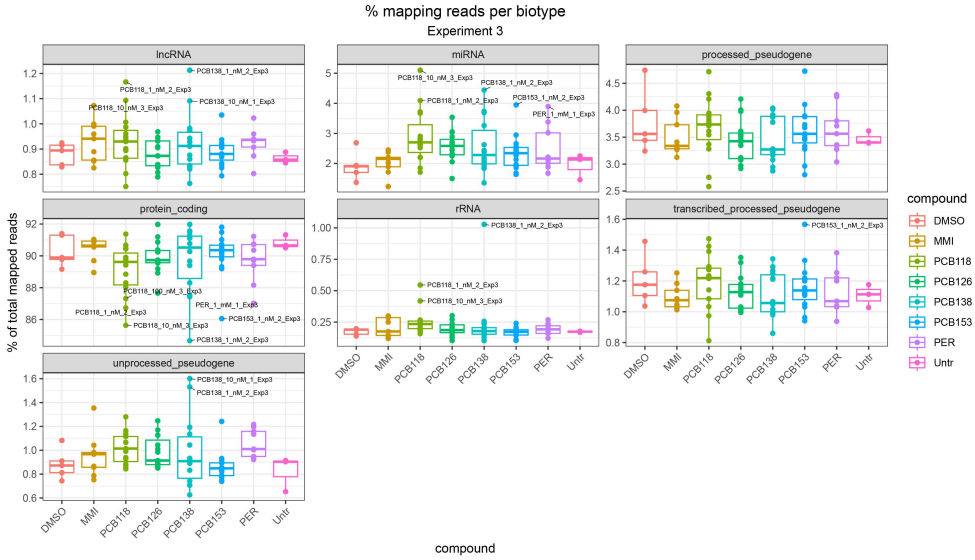


Figure 3. Biotype distribution after snoRNA removal in the samples of (A) Experiment 1, (B) Experiment 2 and (C) Experiment 3. The biotypes that make up at least 1% of the reads are reported. The labelled samples are the ones flagged as outliers according to the formula presented in the Materials and Methods section.

3.1.2 Proteomics QC

The number of proteins identified in all replicates of the HeLa technical standards at the beginning and end of each run and of experimental samples is reported in Table 2. In run 1, 3,452 proteins were identified in all HeLa technical replicates (86.39% of all identified proteins), 3,360 in the experimental samples and 47 in the blank. In run 2, 2,582 proteins were identified in all HeLa technical replicates (79.99% of all identified proteins), 2,648 in the experimental samples and 104 in the blank. For both runs reproducibility in terms of protein identification was consistently high in all groups, which confirmed that instrument stability and performance were maintained throughout the worklist runs. The percentage of overlapping proteins varied between each group and was lower than that of the HeLa digests. This suggests the variation was due to the biological replicates, rather than technical variability. To facilitate a more robust analysis, filtering of the proteins was performed prior to any statistical analysis, as described in Materials and Methods.

Table 2. Number of proteins identified in all HeLa technical replicates (3 at the start, 3 at the end of each run) and experimental samples from the LC-MS/MS data acquisition for Experiments 1, 2 and 3. As explained in the Methods section 2.6.2, samples from Experiments 2 and 3 were analyzed during a single run.

LC-MS/MS data acquisition run	Experiment	Sample	Proteins identified in all replicates (N and % of total)	CV
1	Experiment 1 (phthalates)	HeLa standards	3,452 (86.39%)	1.23%
		Experimental samples	3,630	1.93%
		Blank	47	
2	Experiment 2 (PAHs, OPFRs) and Experiment 3 (PCBs)	HeLa standards	2,582 (79.99%)	3.78%
		Experimental samples	2,648	13.80%
		Blank	104	

3.2 Dose Series Analysis of Gene Expression

We report here the results of dose series analysis on gene expression. The complete list of genes dysregulated by all compounds is reported in Supplementary table 3.

Treatment with phthalates mostly caused a non-monotonic response on gene expression, with the highest dose (10 uM) possibly representing the point of departure (PoD) for several clusters (Supplementary figure 1). Among the induced genes, we identified *Ing5* as being consistently upregulated in all conditions. We have previously performed a thorough analysis on this dataset which also includes a follow-up study of phthalate treatment using ATAC-Seq, and we refer the reader to our other publication (41).

For the polycyclic aromatic hydrocarbons, we identified a significant signature for 7, 7, 9 and 4 genes for BAA, BAP, BKF and DAHA treatments, respectively (Supplementary figure 2). PAHs can bind and activate the aryl hydrocarbon receptor (AhR) a ligand-dependent transcription factor (42). Indeed, we observed a steady increase in expression with increase of dose of some AhR target genes, namely the Aryl Hydrocarbon Receptor Repressor (*Ahrr*) (BKF), the Cytochrome P450 members *Cyp1a1* and *Cyp1b1* (BKF, DAHA), NAD(P)H Quinone Dehydrogenase 1 (*Nqo1*) (43, 44) (BKF, DAHA) and the TCDD Inducible Poly(ADP-Ribose) Polymerase (*Tiparp*) (BKF). It is also interesting to observe how BAA and BAP treatment drastically decreased *mt-Nd6* expression.

Regarding organophosphate flame retardants (OPFRs), we did not see any effect on gene expression by TDCPP. BADP decreased the expression of *Lonfr3* and *Nkd1* and increased the expression of *Glb1l2* and *Tmem199* in a dose-dependent manner. *H3f3b* was repressed only by the highest dose. For the other genes we observed a non-monotonic response (Supplementary figure 3A). DMMP had an effect only a non-monotonic effect (U-shaped) on the expression of

three genes (*Arpc5l*, *Cldn18*, *Gm42826*) with the minimum expression between 10 nM and 1 uM (Supplementary figure 3B). Treatment with TPP increased the expression of *Nif3l1* and had mostly a non-linear effect on the identified genes (Supplementary figure 3C).

Regarding polychlorinated biphenyls (PCBs), we observed a sharp reduction of *Spop* and *Yrdc* and an induction of *Myo5c* after PCB118 treatment (Supplementary figure 4A). For PCB126 we observed a cluster composed of 4 genes (*Aldh3a1*, *Cyp1a1*, *Cyp1b1*, *Tiparp*) that were strongly induced at the 1 uM dose (Supplementary figure 4B). These are Ahr targets, and indeed planar PCBs such as PCB126 can act like Ahr agonists (45). PCB138 had only a non-linear inhibitory effect on *Gprc5a* and *Trim45* (Supplementary figure 4C), while PCB153 reduced or induced the expression of most of the identified genes in a linear fashion (Supplementary figure 4D).

Methimazole (MMI) and sodium perchlorate (PER) are two compounds that interfere with normal thyroid physiology by inhibiting the catalytic activity of Tpo (46) and interfering with I- uptake by Nis (47), respectively. MMI is used as a drug to treat hyperthyroidism, while PER is classified as an endocrine disrupting chemical by the European Chemicals Agency (ECHA) (48). We observed a noticeable overlap between the signatures of MMI and PER in Experiment 3 (Supplementary figure 5A-B): 14 out of 22 mitochondrial tRNA genes belonged to clusters in which the expression decreased in the two lowest doses (10 uM, 100 uM), only to increase, sometimes above the levels of the DMSO-treated controls, in the highest dose (1 mM) (Supplementary figure 8C-D). The other genes were *Gm15191*, *Gm6069*, *Gsn*, *Mir1291*, *Nop56*, *Snhg5* and *Tmem38b*. There was no overlap for the same treatment (MMI or PER) between the two experiments. PER exposure in Experiment 3 was the only treatment that dysregulated a gene important in the thyroid for the synthesis of thyroid hormone, namely *Tshr*, with the expression increasing at the 10 uM and 100 uM doses and decreasing at 1 mM.

3.3 Dose Series Analysis of microRNA Expression

Similarly to the genes, we performed a dose series analysis to miRNA expression to see whether miRNA expression could be altered in a dose-dependent way (Figure 4). Treatment with the four EDCs classes showed interesting dose-effects (Figure 4), with BKF treatment inducing mmu-miR-143-3p and PCB118 inducing mmu-miR-499-5p in a dose-dependent way. BAP and DMMP treatment reduced the expression of mmu-miR-1249-3p and mmu-miR-215-5p, respectively. DnOP treatment increased the expression of mmu-miR-335-3p only at the highest dose. Although mmu-miR-582-3p and mmu-miR-598-3p resulted dysregulated by DEHP and DINP, respectively, the effects were less consistent and we would not confidently consider these changes biologically relevant.

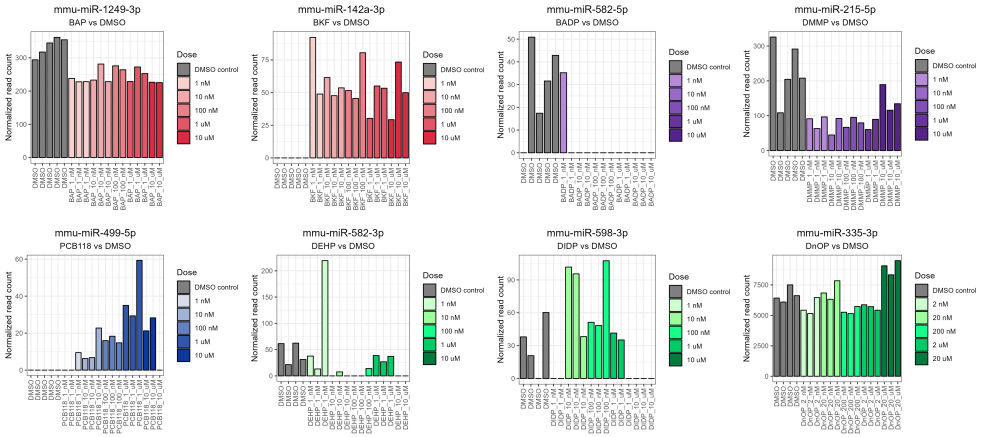


Figure 4. Results of dose series analysis of on miRNA expression. Barplots of expression values of miRNAs that resulted dysregulated in a dose-dependent way as identified by maSigPro (FDR < 0.05). Each bar is a replicate, and the different doses are reported in different shades of color. The grey color represents the DMSO control, which we considered dose 0 (red = PAHs, purple = OPFRs, blue = PCBs, green = phthalates).

3.4 Differential Expression Analysis

3.4.1 Differentially Expressed Proteins at 10 uM Dose

The number of differentially expressed proteins (DEPs) was consistent in the PAHs, PCBs and phthalates classes. In the OPFRs, TDCPP treatment induced around four times (207) the number of DEPs of BADP (69) and DMMP (85), while TPP had around half (29) (Figure 5A). Within the OPFRs, most proteins were differentially expressed by a single compound. Lsm12 was dysregulated by all four compounds (Figure 5B) but in different directions (upregulated by TDCPP, downregulated by the others). Among PAHs, BKF and DAHA shared the highest number of DEPs, followed by BKF, DAHA and BAP. This is interesting, as BKF is the most potent PAH of the ones select, followed by DAHA, then BAP and lastly BAA. Fourteen DEPs were common to all PAHs (Cgn, Ckap4, Eif3i, Fth1, Ist1, Ak4, Pkm, Ndufa12, Parvb, Pbdc1, Psmd2, Rs27l, Tra2b, Ubc9) (Figure 5C). Like OPFRs, most proteins were differentially expressed by individual PCBs, with multiple compounds sharing 5 or less DEPs and Aldh1a3 being downregulated by all treatments (Figure 5D). There was little overlap among the other phthalates and no DEP was common to all compounds (Figure 5E). Regarding the reference compounds MMI and PER, both induced more DEPs in Experiment 2 than 3. The greatest overlap was observed between MMI and PER in Experiment 2 (34 DEPs), while 4 DEPs overlapped between MMI and PER in Experiment 3. The DEPs overlap between the two experiments was 2 for MMI (Itgb1, Ndufa12) and 1 for PER (Sf3a2), respectively (Figure 5F).

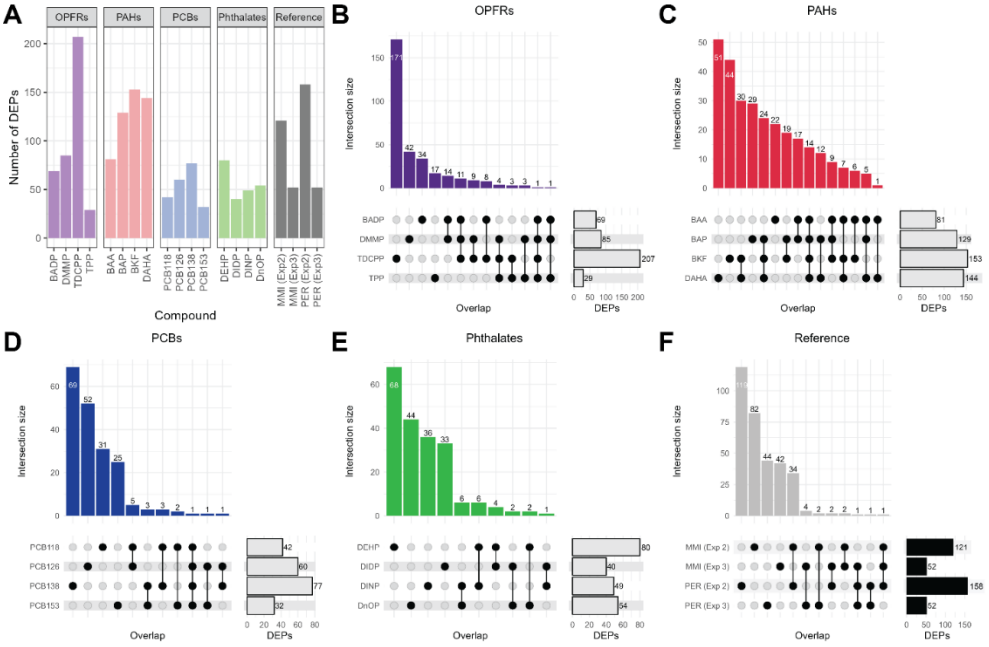
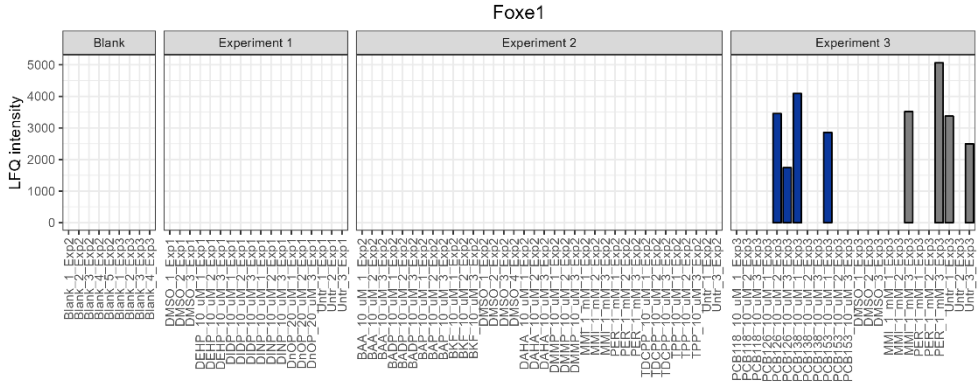
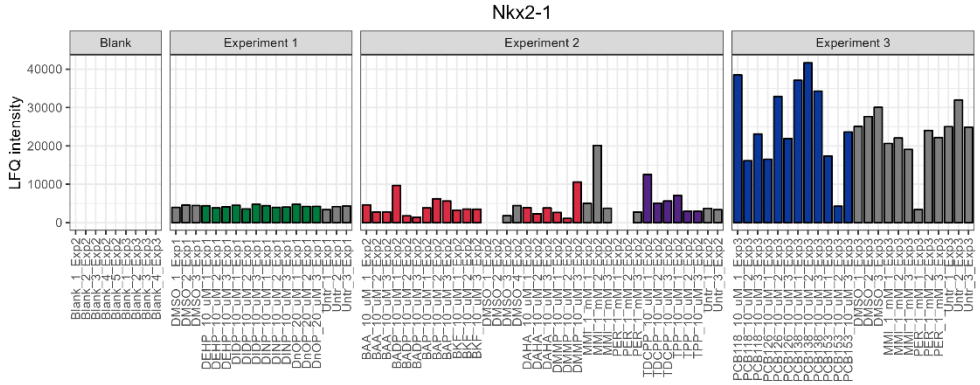
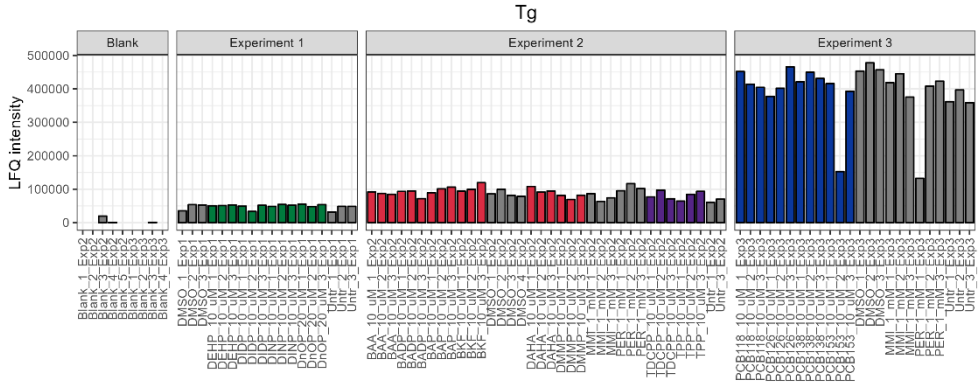


Figure 5. Number of differentially expressed proteins (DEPs) and their overlap within EDC classes. (A) Barplot reporting the number of DEPs. The compounds are grouped per class and the DEPs were determined comparing the compound-treated samples (n = 3) to the DMSO control (n = 3) of the corresponding experiment (1, 2 or 3) (p-value < 0.01). For the reference compounds methimazole (MMI) and sodium perchlorate (PER) the experiment ("Exp2" or "Exp3") is also indicated. (B-F) Upset plot reporting the overlap of DEPs within a compound class (B = OPFRs; C = PAHs; D = PCBs; E = Phthalates; F = reference compounds).

Of note, most thyroid markers or proteins involved in the synthesis of the thyroid hormone were inconsistently detected across all experiments (Foxe1, Nkx2-1, Tg, Tpo, Tshr) (Figure 6), while others (Hhex, Dio1, Dio2, Duox1, Duox2, Duoxa1, Duoxa2, Nis) were not detected at all. Among all treatments, DIDP was the only one that upregulated a thyroid gene, namely TPO (Figure 7).


Foxe1
Blank
Experiment 1
Experiment 2
Experiment 3
LFQ intensity
0
1000
2000
3000
4000
5000


Nkx2-1
Blank
Experiment 1
Experiment 2
Experiment 3
LFQ intensity
0
10000
20000
30000
40000


Tg
Blank
Experiment 1
Experiment 2
Experiment 3
LFQ intensity
0
100000
200000
300000
400000
500000

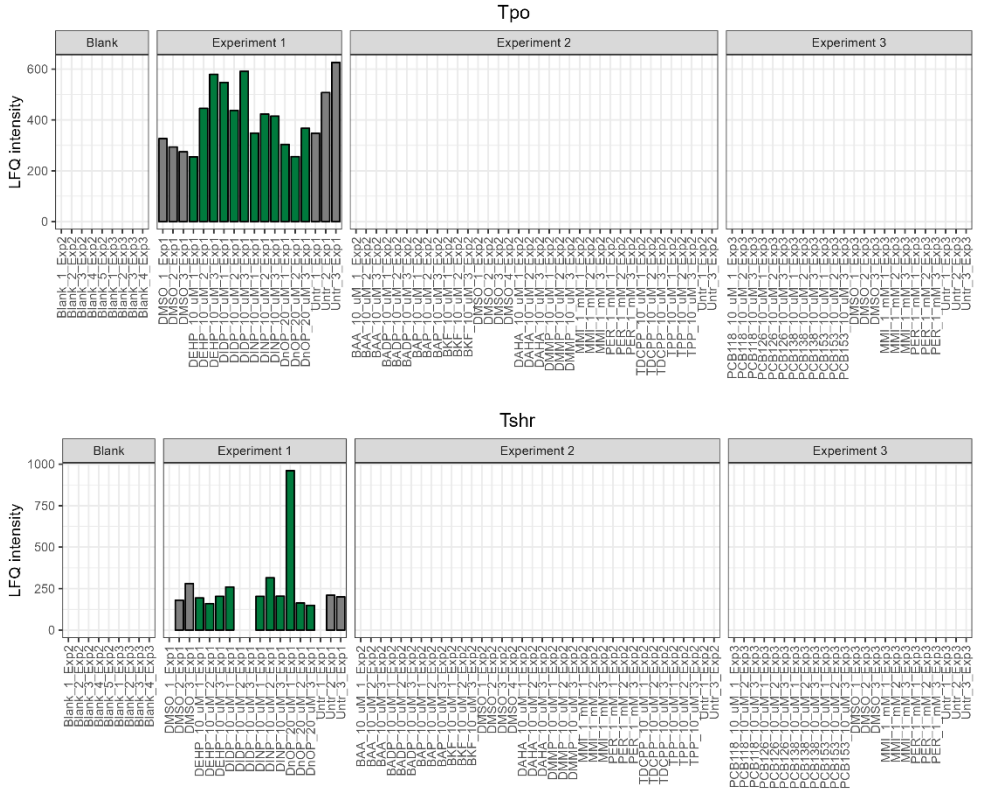


Figure 6. Expression of thyroid markers or proteins important for the synthesis of the thyroid hormone. The samples are grouped per experiment and the blanks (no injection of sample in the mass spectrometer) are also reported (LFQ = Label-Free Quantification).

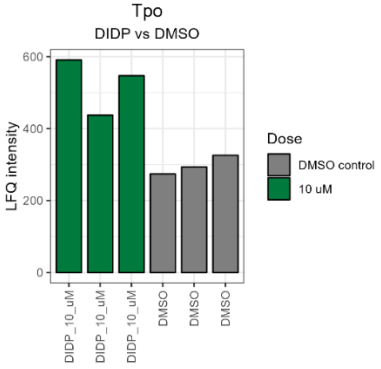
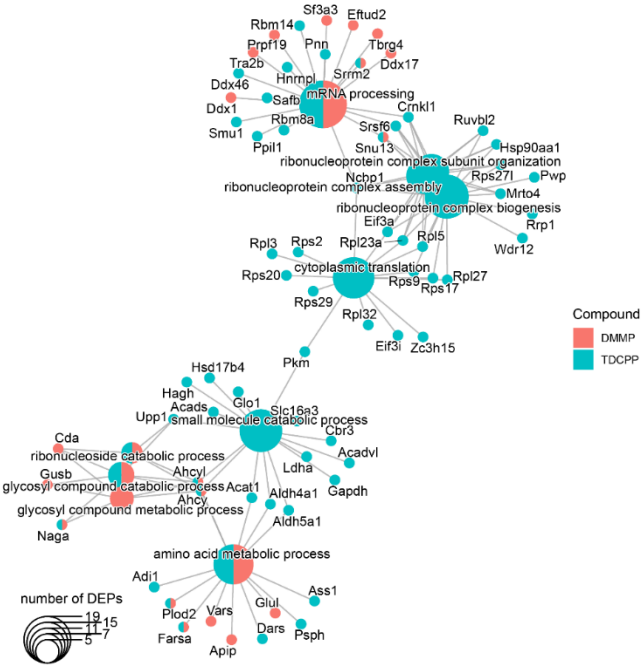


Figure 7. Tpo protein expression after DIDP treatment. The barplot reports the expression of the Tpo protein as LFQ intensity in the DIDP and DMSO control samples in Experiment 1 (LFQ = label-free quantitation).

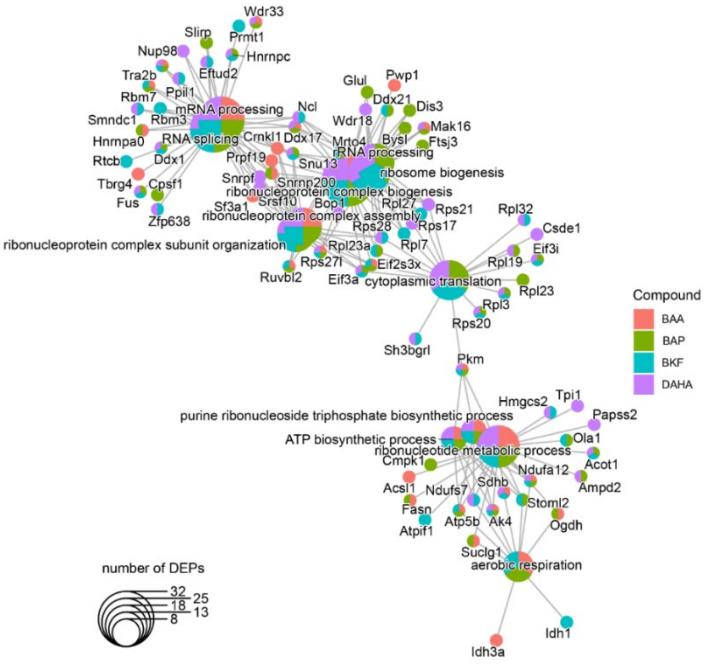
3.4.1.1 Gene Ontology Analysis of Differentially Expressed Proteins

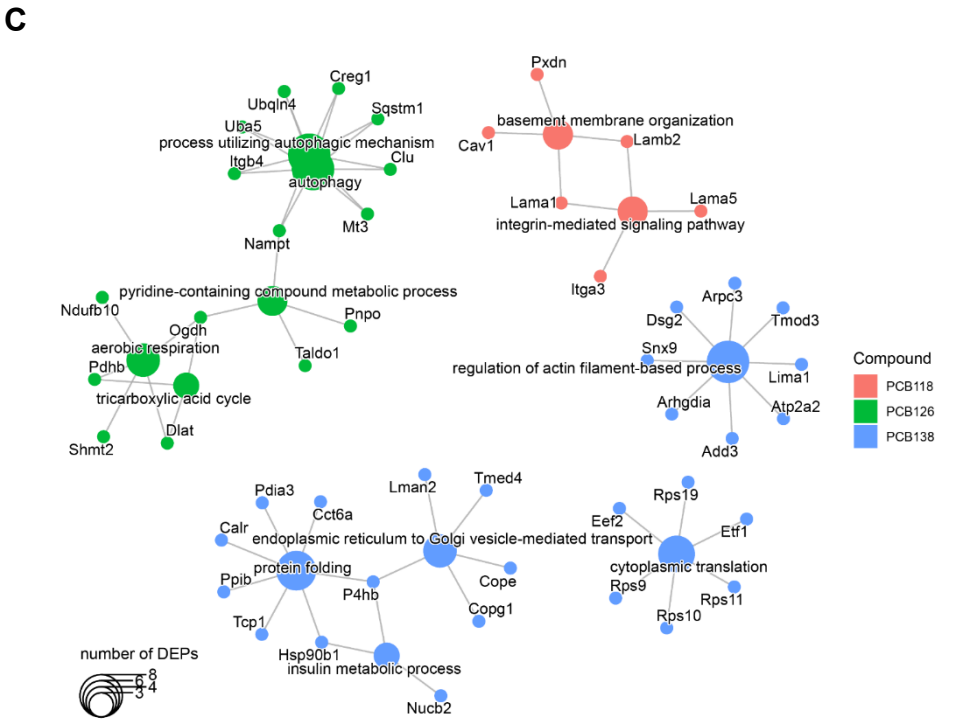
Gene ontology (GO) analysis on the Biological Pathway (BP) (GO BP) ontology showed an enrichment only for DMMP and TDCPP treatment for OPFRs. TDCPP had the most enriched terms, belonging to the ontologies of ribonucleoprotein biogenesis, assembly and organization, mRNA processing, and small molecule catabolic process. The last two terms were shared by DMMP as well (Figure 8A). Most PAHs shared terms related to the themes of energy production (ribonucleotide metabolism, ATP biosynthesis and aerobic respiration) and transcription and translation (ribosome assembly and biogenesis, mRNA processing and splicing, cytoplasmic translation) (Figure 8B). Regarding PBCs, there was no overlap of enriched terms: PCB118 dysregulated pathways related to basement membrane organization and integrin-mediated signaling pathway; PCB126 to autophagy, aerobic respiration, tricarboxylic acid cycle and pyridine-containing compound metabolic process; PCB138 to protein folding and transport from endoplasmic reticulum to Golgi, cytoplasmic translation and regulation of actin filament-based process (Figure 8C). Similarly, we observed enrichment for two phthalates out of four (DEHP and DINP) but the terms did not overlap: DEHP dysregulated pathways involved in protein polymerization and positive regulation of cell growth and nuclear transport. DINP affected instead the pathways of purine nucleotide and ribose phosphate biosynthetic process, regulation of translation initiation and non-coding RNA (ncRNA) catabolism (Figure 8D). GO BP analysis on the reference compounds showed how the cell differentiation batch affected the enriched pathways the most, with terms related to RNA splicing and mRNA processing, ribonucleoprotein biogenesis, assembly and organization and translation being enriched in samples treated with MMI and PER in Experiment 2. For PER (Experiment 3) we saw instead an enrichment for terms related to protein localization to cell junction (Figure 8E). There was no enrichment for MMI (Experiment 3).

A


Sf3a3
Eftud2
Rbm14
Prpf19
Pnn
Tbrg4
Tra2b
Ddx17
Srrm2
Hnrnpl
Ddx46
mRNA processing
Safb
Ddx1
Rbm8a
Crnkl1
Smu1
Srsf6
Ruvbl2
Ppil1
Snu13
Hsp90aa1
ribonucleoprotein complex subunit organization
Nchp1
Rps27l
Pwp
ribonucleoprotein complex assembly
Mrto4
ribonucleoprotein complex biogenesis
Eif3a
Rrp1
Rpl3
Rps2
Rpl23a
Rpl5
Wdr12
cytoplasmic translation
Rps20
Rps9
Rpl27
Rps17
Rps29
Rpl32
Eif3i
Zc3h15
Pkm
Compound
DMMP
TDCPP
Hsd17b4
Hagh
Glo1
Acads
Upp1
Slc16a3
small molecule catabolic process
Cbr3
Cda
Acadvl
ribonucleoside catabolic process
Ahcyl
Ldha
Gusb
glycosyl compound catabolic process
Acat1
Gapdh
Ahcy
Aldh4a1
glycosyl compound metabolic process
Aldh5a1
Naga
amino acid metabolic process
Adi1
Ass1
number of DEPs
19
15
11
7
5
Plod2
Vars
Glul
Farsa
Dars
Psph
Apip

B


Wdr33
Slirp
Prmt1
Nup98
Hnrnpc
Tra2b
Eftud2
Glul
Pwp1
Rbm7
Ppil1
Ddx21
Dis3
mRNA processing
Ncl
Smndc1
Rbm3
Wdr18
Mak16
Hnrnpa0
RNA splicing
Crnkl1
Ddx17
Mrto4
Bysl
Ftsj3
rRNA processing
Rtcb
Ddx1
Prpf19
Snu13
ribosome biogenesis
Tbrg4
Cpsf1
Snrpf
Snrnp200
ribonucleoprotein complex biogenesis
Fus
Srsf10
Bop1
Rpl27
Rps21
Rpl32
Zfp638
Sf3a1
ribonucleoprotein complex assembly
Rps28
Rps17
Csde1
ribonucleoprotein complex subunit organization
Rpl23a
Rpl7
Eif3i
Rps27l
Eif2s3x
Rpl19
Ruvbl2
Eif3a
cytoplasmic translation
Rpl23
Rpl3
Rps20
Compound
BAA
BAP
BKF
DAHA
Sh3bgrl
Pkm
Hmgcs2
Tpi1
purine ribonucleoside triphosphate biosynthetic process
Papss2
ATP biosynthetic process
Ola1
ribonucleotide metabolic process
Cmpk1
Ndufa12
Acot1
Acsl1
Ndufs7
Sdhb
Ampd2
Stoml2
Fasn
Atp5b
Ak4
Ogdh
Atpif1
Suclg1
aerobic respiration
number of DEPs
32
25
18
13
8
Idh1
Idh3a


C
Creg1
Ubqln4
Sqstm1
Uba5
process utilizing autophagic mechanism
Itgb4
Clu
autophagy
Mt3
Nampt
Pxdn
basement membrane organization
Cav1
Lamb2
Lama1
Lama5
integrin-mediated signaling pathway
Itga3
pyridine-containing compound metabolic process
Ndufb10
Pnpo
Ogdh
aerobic respiration
Taldo1
Pdhb
tricarboxylic acid cycle
Dlat
Shmt2
Arpc3
Dsg2
Tmod3
Snx9
regulation of actin filament-based process
Lima1
Arhgdia
Atp2a2
Add3
Compound
PCB118
PCB126
PCB138
Pdia3
Lman2
Tmed4
Cct6a
Calr
endoplasmic reticulum to Golgi vesicle-mediated transport
protein folding
Ppib
P4hb
Cope
Copg1
Tcp1
Hsp90b1
insulin metabolic process
Nucb2
Rps19
Eef2
Etf1
cytoplasmic translation
Rps9
Rps11
Rps10
number of DEPs
8
6
4
3
D
Alyref
Ipo11
Nup93
Hspa9
Ppp1r12a
nuclear transport
Nup133
nucleocytoplasmic transport
Nmd3
Nsun2
ncRNA catabolic process
Dis3
Zc3h4
Stmn2
Cfl1
protein polymerization
Ube2s
regulation of protein polymerization
Capza2
positive regulation of cell growth
Rhoa
Dbnl
Mtpn
Sh3glb1
Exosc9
Prpsap2
Nme2
purine nucleotide biosynthetic process
ribose phosphate biosynthetic process
Ndufa9
purine-containing compound biosynthetic process
Ppcdc
Nudt2
Compound
DEHP
DINP
Eif5
Eif2s1
Eif4g2
regulation of translational initiation
Rpl13a
number of DEPs
7
6
4
3

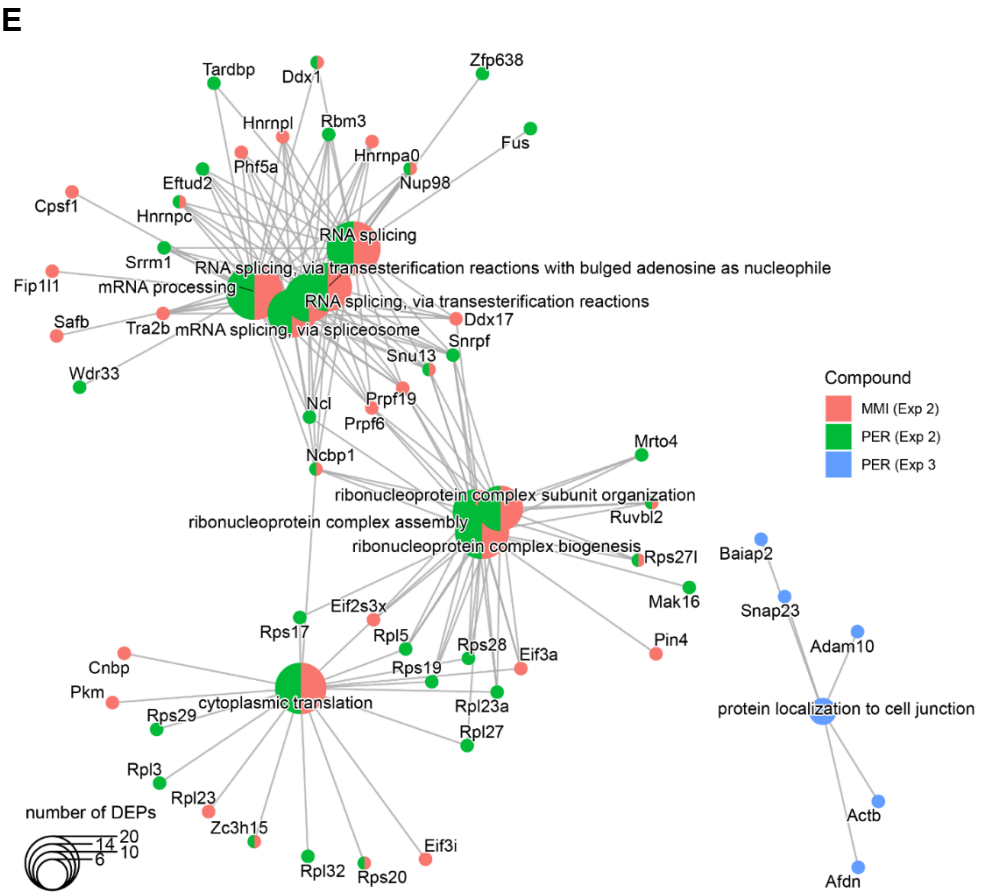


Figure 8. Gene Ontology Biological Process (GO) analysis of differentially expressed proteins (DEPs). GO BP enrichment analysis for (A) OPFRs, (B) PAHs, (C) PCBs, (D) phthalates and (E) reference compounds. The enriched terms and the DEPs referring to them are reported (q-value < 0.01). The node size represents the number of DEPs participating to the term.

3.4.2 Differentially Expressed Genes and miRNAs at 10 uM Dose per Class

To identify genes and miRNAs that were differentially expressed at the class level and compare them to the identified differentially expressed proteins, we performed differential expression analysis grouping the samples of a compound class exposed to the 10 uM dose to their respective DMSO control. In this way, we aimed at identifying genes and miRNAs which may represent a biomarker of exposure to a certain class.

3.4.2.1 Differentially Expressed Genes

We identified 2, 3, 7 and 88 genes as being differentially expressed after phthalate, PAH, OPFR and PCB treatment, respectively (Figure 9A-D). Of these, 4 genes in the PCB group overlapped with those identified in the maSigPro analysis: *Aldh3a1* was upregulated and induced by PCB126 in the maSigPro analysis, *Myo5c* upregulated and induced by PCB118, *Fam20a* upregulated and *Nectin3* downregulated and dysregulated by PCB153 in the same direction. Additionally, *Dio1*, *Tshr* and *Tpo* were upregulated after PCB treatment compared to the control.

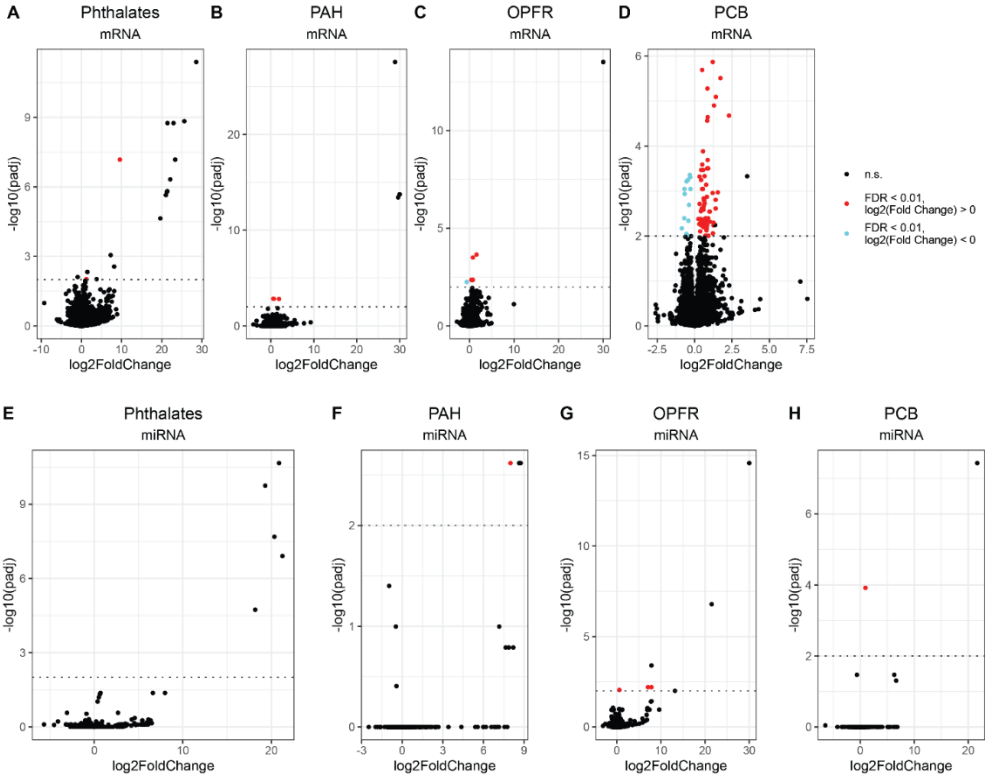
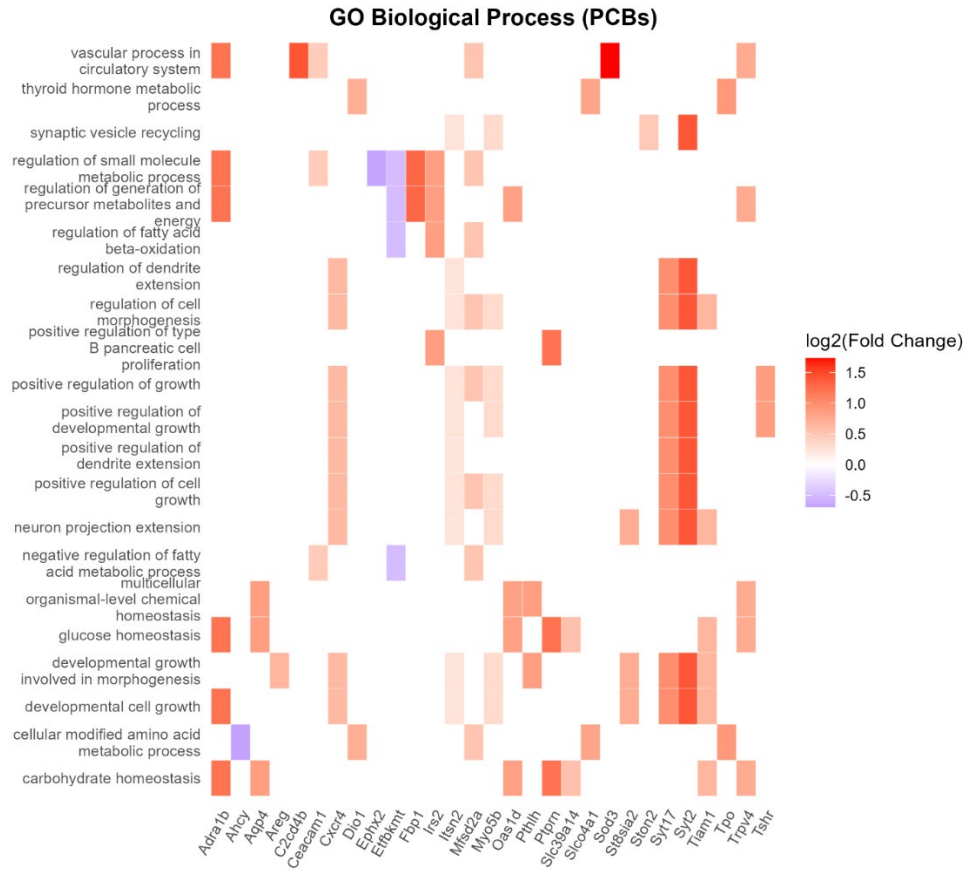


Figure 9. Volcano plots of differentially expressed genes (A-D) and miRNAs (E-H) in each class 10 uM vs DMSO control comparison. Every dot represents a gene/miRNA. Features not differentially expressed ('n.s.') are in black. The False Discovery Rate (FDR) threshold of 0.01 is indicated as a dotted line on the y-axis. Upregulated features ('FDR < 0.01, log2FoldChange > 0') are indicated in red, downregulated ones ('FDR < 0.01, log2FoldChange < 0') in cyan. The features above this line colored in black are the ones that fail to pass the 'spurious spike' or '3rd quartile' filters as described in the Methods section.

GO BP analysis of the DEGs showed enrichment only for PCBs (q-value < 0.01). The 21 enriched pathways were mostly upregulated (Figure 10A) and related to the biological themes

of glucose homeostasis, regulation of fatty acid metabolism, cell growth and development and thyroid hormone metabolic process (Figure 10B).

A



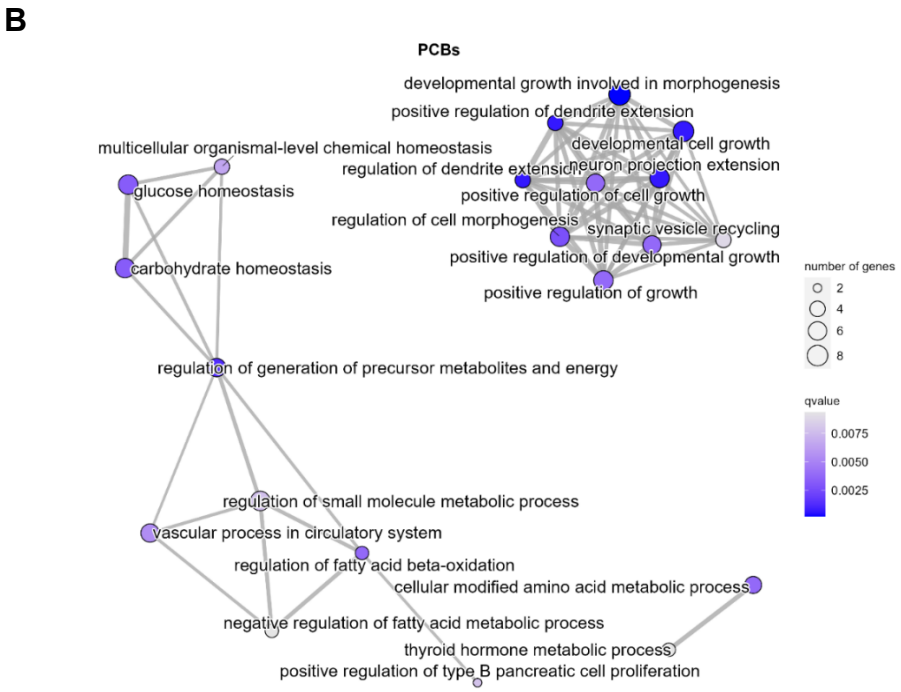


Figure 10. Results of the Gene ontology – Biological Process analysis of DEGs in the PCB vs DMSO comparison (q-value < 0.01). (A) Heatplot of the enriched terms. (B) Emap plot of the enriched terms, connecting the ones that share some of the DEGs.

3.4.2.2 Differentially Expressed miRNAs

The global expression of miRNAs was not highly affected by treatments (Figure 9E-H). All were induced compared to the control, some at high level and in all treated samples (mmu-miR-182-5p by OPFRs and mmu-miR-335-3p by PCBs), while others at a relative lower level and not in 100% of treated samples (mmu-miR-142a-3p by PAHs, mmu-miR-3076-3p and mmu-miR-6939-3p by OPFRs) possibly due to a sampling effect driven by the low read count (Figure 11). No snoRNA was found differentially expressed.

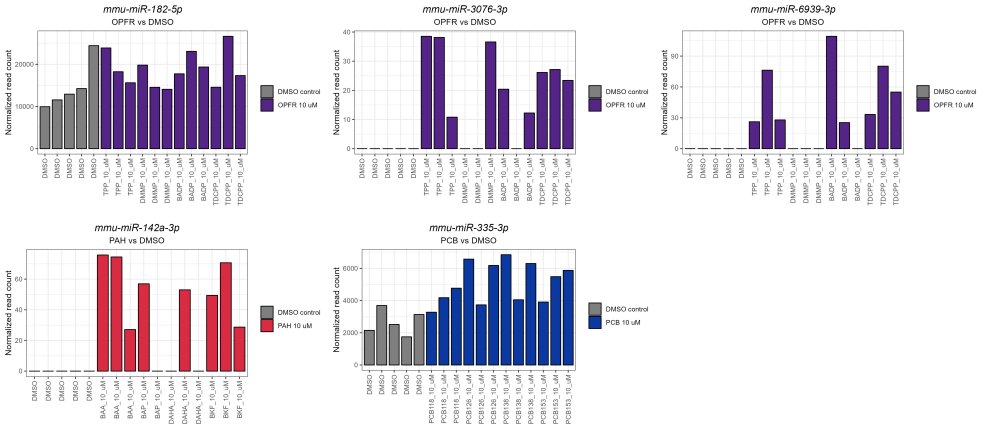


Figure 11. Barplot of normalized read count of differentially expressed miRNAs for each EDC class 10 uM vs DMSO control comparison (FDR < 0.01). Every bar represents a sample. The different treatments are color-coded (grey = DMSO, color = 10 uM dose). The EDC class is reported at the top of the plots.

3.5 Integration of Transcriptomics and Proteomics

To see whether any alteration of gene or miRNA expression reflected on the differential expression of proteins, we checked for any overlap between differentially expressed or dose-dependent changes in gene or miRNA expression and protein expression.

3.5.1 Overlap between targets of affected miRNAs and DEPs

Since, in mammals, miRNAs can interfere with protein translation by binding to the 3' end of the mRNA (49), we looked for an overlap between the dysregulated miRNAs or DE miRNAs and DEPs whose transcript is reported to be their target.

Interestingly, we identified Hmgb1 as being downregulated after BKF treatment (Figure 12) and a target of mmu-miR-143-3p. Of note, mmu-miR-143-3p was detected also in BAA treatment, while it did not pass the relevance filter in BAP and DAHA treatments. Additionally, Hmbg1 was not differentially expressed in the other PAHs comparisons in a statistically significant way.

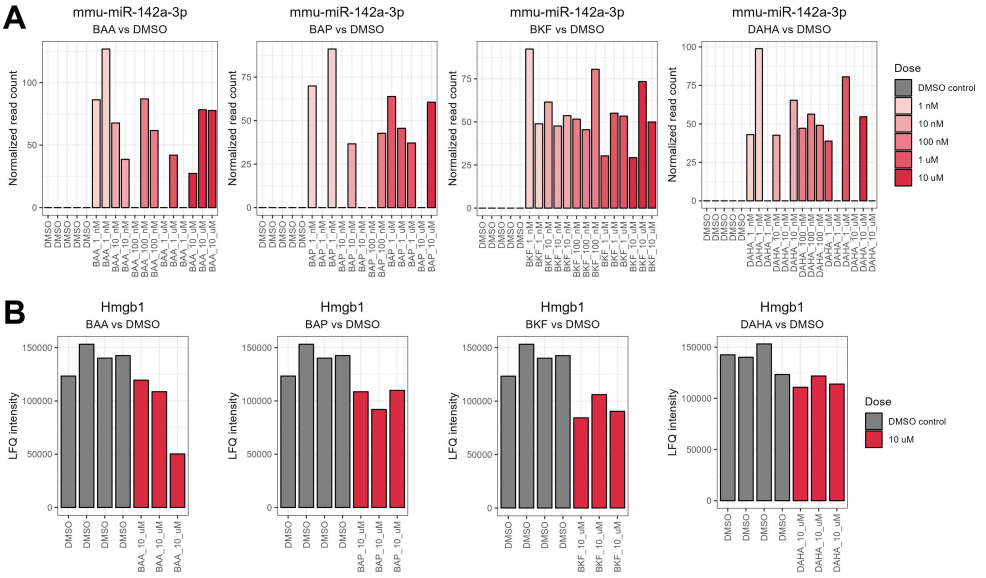


Figure 12. Expression of mmu-miR-142-3p and Hmgb1 in the samples treated with polycyclic aromatic hydrocarbons. Barplots reporting the expression of (A) mmu-miR-142-3p and (B) Hmgb1 protein. The gene expression is reported as normalized read count, while the protein expression as LFQ intensity. Each dose or control has a different color and the DMSO samples refer to Experiment 2. The treatment to which each plot refers is reported below the miRNA/protein name (LFQ = label-free quantitation).

3.5.3 Random Forest Model for EDC Class Features Identification

We applied a random forest (RF) classification algorithm to the combined gene, miRNA and protein expression to see whether we could identify features that would allow us to distinguish between a certain EDC class and the other samples.

To select the genes or miRNAs to use for the model, we used the read count in transcript per million (TPM) value and retained the features whose average count across samples was greater than 0.5 TPM. Regarding proteins, we initially selected those detected in all triplicates, and no further filtering was applied. This resulted in 8,024 features being selected for training the model. The trained model was able to accurately classify the 4 EDC classes with surprisingly high accuracy (1 for PAHs and OPFRs and 0.997 for phthalates and PCBs). For all classes, the Area Under the Receiver Operating Characteristic curve (AUC) was 1 (Figure 13A, Table 3).

Table 3. Accuracy and AUC for each random forest binary classification model trained to distinguish between a certain EDC class and the rest (AUC = Area Under the Receiver Operating Characteristic curve) calculated after cross-validation. The mean and standard error are reported.

	Accuracy		AUC	
	Mean	Standard error	Mean	Standard error
Phthalates	0.997	0.00270	1	0
PAH	1	0	1	0
OPFR	1	0	1	0
PCB	0.997	0.00263	1	0

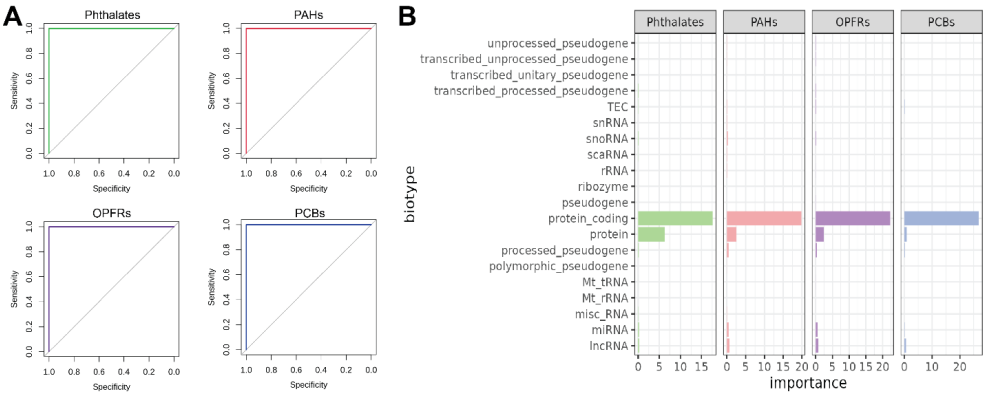


Figure 13. (A) Receiver Operating Characteristic (ROC) curves for the trained random forest (RF) classification model for phthalates, PAHs, OPFRs and PCBs. (B) Gini importance per biotype for each EDC class.

We calculated the sum of the Gini importance (which measures the importance of a feature in the decision tree) per biotype for each model (Table 4) and saw how protein coding genes had the most weight in all cases, followed by protein expression. For phthalates, proteins had a relatively higher importance than the other classes, while for PCBs they had a relatively lower importance. Other noteworthy biotypes were miRNAs, long non-coding RNAs (lncRNAs), processed pseudogenes, TEC ("To be Experimentally Confirmed") and snoRNAs (Figure 13B).

Table 4. Importance of biotypes per EDC class in each random forest classification model. The biotypes with importance >= 0.1 are reported.

Phthalates	Importance
protein coding	17.600
protein	6.270
lncRNA	0.180

PAHs	Importance
protein coding	19.900
protein	2.520
lncRNA	0.635

miRNA	0.167
processed pseudogene	0.126
snoRNA	0.104

miRNA	0.444
processed pseudogene	0.443
snoRNA	0.263

OPFRs	Importance
protein coding	22.200
protein	2.490
lncRNA	0.853
miRNA	0.597
processed pseudogene	0.405
snoRNA	0.137
TEC	0.116

PCBs	Importance
protein coding	26.800
protein	0.889
lncRNA	0.577
processed pseudogene	0.234
TEC	0.185
miRNA	0.170

lncRNA = long non-coding RNA; snoRNA = small nucleolar RNA; TEC = "To be experimentally confirmed"

It is interesting to note that, despite using 18,024 features, only around 1,000 ended up having an importance greater than 0 in each model (Figure 14).

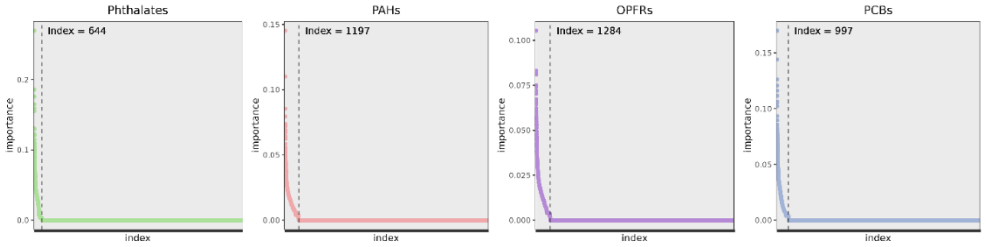


Figure 14. Importance per feature ordered by index for each random forest (RF) classification model. Index = 1 indicates the feature with the highest importance. The label "Index" reports the index of the first feature with importance = 0.

4. Discussion

In this work, we exposed previously described mouse embryonic stem cell (mESC)-derived thyroid organoids (50, 51) for 24 hours to five increasing doses (1-10-100 nM-1-10 uM) of endocrine disrupting chemicals (EDCs) belonging to the four classes of phthalates, polycyclic aromatic hydrocarbons (PAHs), organophosphate flame retardants (OPFRs) and polychlorinated biphenyls (PCBs), and to three doses (10-100 uM-1 mM) of methimazole (MMI) and sodium perchlorate (PER), compounds that interfere with the production of the thyroid hormone (TH) and which we termed "reference compounds" (MMI is a clinical drug while PER is an endocrine disrupting chemical with a known negative effect on the thyroid). At the end of the exposure time, we analyzed the transcriptome and proteome with RNA sequencing (RNA-Seq) and liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS), respectively, to study how the treatments would affect gene and protein expression. Due to the limited amounts of organoids that can be obtained in a single differentiation batch and the high amount of input cells required for untargeted proteomics, we could only expose the samples dedicated to proteomic analysis to the highest dose (10 uM or 1 uM). For transcriptomic libraries, we used the Combo-Seq kit, and we were thus able to construct combined poly(A) and small RNA libraries. The low doses we selected for the EDCs aimed at reflecting the serum levels measured for some of these compounds in human biomonitoring studies, which study the exposure to several types of chemicals in the general population (52-56).

We used maSigPro to analyze whether gene expression would be affected in a dose dependent way by EDC treatment (57). The tool performs two sequential regression steps: first it selects genes with a non-flat profile (which indicates a lack of change in expression over the series), and then it creates the best regression model for each gene that can fit a series (e.g. time or dose). The regression model is built upon a polynomial equation. Therefore, it is crucial to choose the degree of the polynomial based on the characteristics of the data series under analysis. The higher the degree, the more changes of direction in expression are allowed in the modelled curve. Our data had 6 or 4 levels (0, 1 nM, 10 nM, 100 nM, 1 uM, 10 uM for EDCs, or 0, 10 uM, 100 uM, 1 mM for the reference compounds) and we opted for a degree of 3. In this way, we would be able to observe non-monotonic dose-responses, which have been described for several EDCs and summarized in an excellent review by Vandenberg and colleagues (2012), including some of the compounds we selected (58). It must be noted that the tool used limits the shapes of the curve that can be fitted, excluding for example exponential or sigmoid curves, which can be the case for EDC effects (58). A potential alternative tool for the analysis of dose-response curves is BMDExpress2 (59), which was developed for regulatory purposes by a collaboration of US and Canadian agencies for the identification of benchmark doses for toxicogenomic data and will fit the data to several curves (polynomial, sigmoid, exponential). However, more doses (at least 7-8) are required by the tool to perform proper modeling.

In our results, we could identify the point of departure (PoD) of many genes after treatment with phthalates. For PCB126, we observed that the AhR targets *Aldh3a*, *Cyp1a1*, *Cyp1b1* and *Tiparp* were induced from 1 uM. Indeed, among the selected PCBs, PCB126 is the only coplanar congener and possesses AhR-agonistic activity (45). Similarly, we observed an effect on the AhR targets gene expression for the strongest AhR agonists among the PAHs (BKF and DAHA) (60), but not for BAP and BAA. It is possible, then, that for this model, the PoD for BKF and DAHA lies below 1 nM, and for BAP and BAA above 10 uM. Regarding the reference compounds, we observed an overlap of 20 dysregulated genes only between MMI and PER in Experiment 3, showing the influence of the organoids differentiation batch. Notably, 13 of these were mitochondrial tRNA genes. However, only 1 and 2 genes overlapped between either MMI or PER and any of the EDC-treated samples in Experiment 2 and 3, respectively, but with a different profile. Thus, it appears that the actions of MMI and PER are distinct from those exhibited by the other compounds tested in the same experiment. Interestingly, none of the thyroid markers or genes important for the synthesis of the TH were affected according to the dose-series analysis, save for *Tshr* by PER in Experiment 3.

Dose analysis of miRNAs expression gave interesting results, and we observed some miRNAs being induced (mmu-miR-142a-3p by BKF) or repressed (mmu-miR-1249-3p by BAP, mmu-miR-582-5p by BADP, mmu-miR-215-5p by DMMP) by treatment irrespectively of the dose, others being steadily induced by treatment (mmu-miR-449-5p by PCB118) or at a certain dose (mmu-miR-335-3p by DnOP).

Interestingly, the protein Hmbg1, target of mmu-miR-142a-3p, was downregulated after BKF treatment. HMGB1 is a multifunctional protein: in the nucleus, it acts as DNA chaperone binding both to DNA and histones and has a role in regulating gene expression. Extracellular HMGB1 acts as a cytokine to promote inflammation and can be released by immune cells, stressed, or dying cells. Cytosolic HMGB1 can promote autophagy in response to cellular stress (61). Due to the proteomic sample preparation protocol, the main source of the Hmgb1 we measured would be mostly intracellular, though it was not possible to distinguish between the nuclear and cytosolic content. Mmu-miR-142-3p was induced in the other PAHs treatments, but it was not detected in some samples (3 to 5 depending on the compound) and was thus not picked up by maSigPro. This could be a consequence of the Combo-Seq protocol used for library prep: indeed, the chosen RNA input (from 5 ng to 50 ng) affects the mRNA/miRNA ratio, with high RNA amounts leading to more reads mapping to mRNAs than to miRNAs. For this work, we opted for 20 ng, and the normalized expression of mmu-miR-142a-3p is rather low compared to other miRNAs, in the range of 50-100 reads per sample. It is then possible that its absence of detection in some samples was a consequence of the technique used, and this could hold true for other miRNAs with a similar expression pattern. When we performed differential expression analysis comparing the 10 uM dose of all PAHs together to the

control, mmu-miR-142a-3p was found upregulated (we selected just this dose to perform a better comparison with the results of differential protein expression). Mmu-miR-142-3p has been described to promote inflammation (62-64) but also to act as tumor suppressor and negatively regulate HMGB1 in *in vitro* cancer models (65-69).

To demonstrate whether, with the data available to us, we would be able to identify alterations in features expression (genes, miRNAs or proteins) characteristic of a certain EDC class, we tried two different approaches: first, we grouped the expression data of all the samples exposed to 10 uM of individual EDC classes and compared it to the solvent control, and second we tried to apply to the same samples (10 uM or DMSO control) a random forest (RF) classification algorithm using the integrated gene, miRNA and protein expression. With the first approach, we identified both genes and miRNAs for PAHs, OPFRs and PCBs, while only two genes for phthalates. In the case of PCBs, the thyroid markers *Dio1*, *Tpo* and *Tshr* were upregulated, suggesting how PCBs could directly interfere with the TH production in the thyroid.

With the second approach, using RF, we built four models (one per EDC class vs the rest of the samples) whose accuracy resulted to be 1 (PAHs and OPFRs) or 0.997 (phthalates and PCBs). The consistent cumulative changes in features expression and the pattern of these changes allow the machine learning (ML) algorithm to progressively build a consensus tree able to eventually discern whether a sample belongs to the EDC class it has been trained on or not. A limitation of our approach, which is intrinsic to ML algorithms, is that any dataset we want to evaluate our models on needs to have the same features of the dataset the models were constructed on. This is not always the case for the data at hand, since different RNA-Seq library construction approaches or mass spectrometry techniques, bioinformatics tools, identification criteria and genome or proteome references can all affect which features are detected and at what levels. An alternative approach could be to select a subset of features based on based on their contribution of the models (Gini importance value). Indeed, for our models, around 1,000 features contributed. For example, a higher threshold for the expression level could be applied, which could select features more likely to be detected regardless of the techniques used for data acquisition.

Another limitation was the low number of proteins identified by untargeted proteomics. This technique aims at identifying and quantifying as many proteins as possible, however, several factors in data acquisition, for example variations in liquid chromatography or peptide fragmentation, the stochasticity of the sampling process, the database used, can affect peptide detection and limit the number of detected proteins or cause missing data points (70-72). The stochasticity of the sampling could also cause least abundant peptides that elute at the same time of more abundant peptides to not be detected by the machine. This could explain, for

example, why we could detect only some of thyroid proteins or proteins important for the synthesis of the TH, and their detection was not consistent across experiments.

Both MMI and PER impact on the transcriptome exhibited inconsistencies across the two experiments using different organoids differentiation batches. Based on our results, they are not suitable compounds to use as "positive control" for short-term omics studies as the observed effects cannot be reliably associated with influencing the process of TH production when examining the genes affected.

In conclusion, we show how it is possible to analyze transcriptomics and proteomics data alone or integrated to gain biological insight in the effects of EDCs on *in vitro* thyroid organoids.

5. Supplementary Material

Supplementary table 1. LC-MS/MS worklists for the two runs (run 1 = samples from Experiment 1; run 2 = samples from Experiments 2 and 3). In each run, blanks, HeLa QC and follicles samples were analyzed. Samples were injected in order, from top to bottom.

Run 1	Datafile name	Sample
	2021_08_11_UCD_timsTOF_Cl4_Blank_01	Blank_01
	2021_08_11_UCD_timsTOF_Cl4_HeLa_tr01	HeLa_tr01
	2021_08_11_UCD_timsTOF_Cl4_HeLa_tr02	HeLa_tr02
	2021_08_11_UCD_timsTOF_Cl4_HeLa_tr03	HeLa_tr03
	2021_08_11_UCD_timsTOF_Cl4_Blank_02	Blank_02
	2021_08_11_UCD_timsTOF_Cl4_mU1	mU1
	2021_08_11_UCD_timsTOF_Cl4_mU2	mU2
	2021_08_11_UCD_timsTOF_Cl4_mU3	mU3
	2021_08_11_UCD_timsTOF_Cl4_mC1	mC1
	2021_08_11_UCD_timsTOF_Cl4_mC2	mC2
	2021_08_11_UCD_timsTOF_Cl4_mC3	mC3
	2021_08_11_UCD_timsTOF_Cl4_mDOP1	mDOP1
	2021_08_11_UCD_timsTOF_Cl4_mDOP2	mDOP2
	2021_08_11_UCD_timsTOF_Cl4_mDOP3	mDOP3
	2021_08_11_UCD_timsTOF_Cl4_mDINP1	mDINP1
	2021_08_11_UCD_timsTOF_Cl4_mDINP2	mDINP2
	2021_08_11_UCD_timsTOF_Cl4_mDINP3	mDINP3
	2021_08_11_UCD_timsTOF_Cl4_mDIDP1	mDIDP1
	2021_08_11_UCD_timsTOF_Cl4_mDIDP2	mDIDP2
	2021_08_11_UCD_timsTOF_Cl4_mDIDP3	mDIDP3
	2021_08_11_UCD_timsTOF_Cl4_mDEHP1	mDEHP1
	2021_08_11_UCD_timsTOF_Cl4_mDEHP2	mDEHP2
	2021_08_11_UCD_timsTOF_Cl4_mDEHP3	mDEHP3
	2021_08_11_UCD_timsTOF_Cl4_Blank_03	Blank_03
	2021_08_11_UCD_timsTOF_Cl4_HeLa_tr04	HeLa_tr04
	2021_08_11_UCD_timsTOF_Cl4_HeLa_tr05	HeLa_tr05
	2021_08_11_UCD_timsTOF_Cl4_HeLa_tr06	HeLa_tr06
	2021_08_11_UCD_timsTOF_Cl4_Blank_04	Blank_04
Run 2	**Datafile name**	**Sample**
	2022_04_11_UCD_timsTOF_Cl4_Blank_01	Blank_01
	2022_04_11_UCD_timsTOF_Cl4_HeLa_tr01	HeLa_tr01
	2022_04_11_UCD_timsTOF_Cl4_HeLa_tr02	HeLa_tr02
	2022_04_11_UCD_timsTOF_Cl4_HeLa_tr03	HeLa_tr03
	2022_04_11_UCD_timsTOF_Cl4_Blank_02	Blank_02
	2022_04_11_UCD_timsTOF_Cl4_m2U1	m2U1
	2022_04_11_UCD_timsTOF_Cl4_m2U3	m2U3
	2022_04_11_UCD_timsTOF_Cl4_m2C1_1	m2C1_1
	2022_04_11_UCD_timsTOF_Cl4_m2C2_1	m2C2_1
	2022_04_11_UCD_timsTOF_Cl4_m2C3_1	m2C3_1

	2022_04_11_UCD_timsTOF_Cl4_m2C4_1	m2C4_1
	2022_04_11_UCD_timsTOF_Cl4_m2TPP1	m2TPP1
	2022_04_11_UCD_timsTOF_Cl4_m2TPP2	m2TPP2
	2022_04_11_UCD_timsTOF_Cl4_m2TPP3	m2TPP3
	2022_04_11_UCD_timsTOF_Cl4_m2BADP1	m2BADP1
	2022_04_11_UCD_timsTOF_Cl4_m2BADP2	m2BADP2
	2022_04_11_UCD_timsTOF_Cl4_m2BADP3	m2BADP3
	2022_04_11_UCD_timsTOF_Cl4_m2DMMP1	m2DMMP1
	2022_04_11_UCD_timsTOF_Cl4_m2DMMP2	m2DMMP2
	2022_04_11_UCD_timsTOF_Cl4_m2DMMP3	m2DMMP3
	2022_04_11_UCD_timsTOF_Cl4_m2TDCPP1	m2TDCPP1
	2022_04_11_UCD_timsTOF_Cl4_m2TDCPP2	m2TDCPP2
	2022_04_11_UCD_timsTOF_Cl4_m2TDCPP3	m2TDCPP3
	2022_04_11_UCD_timsTOF_Cl4_m2C1_2	m2C1_2
	2022_04_11_UCD_timsTOF_Cl4_m2C2_2	m2C2_2
	2022_04_11_UCD_timsTOF_Cl4_m2C3_2	m2C3_2
	2022_04_11_UCD_timsTOF_Cl4_m2C4_2	m2C4_2
	2022_04_11_UCD_timsTOF_Cl4_m2BAA1	m2BAA1
	2022_04_11_UCD_timsTOF_Cl4_m2BAA2	m2BAA2
	2022_04_11_UCD_timsTOF_Cl4_m2BAA3	m2BAA3
	2022_04_11_UCD_timsTOF_Cl4_m2BAP1	m2BAP1
	2022_04_11_UCD_timsTOF_Cl4_m2BAP2	m2BAP2
	2022_04_11_UCD_timsTOF_Cl4_m2BAP3	m2BAP3
	2022_04_11_UCD_timsTOF_Cl4_m2BKF1	m2BKF1
	2022_04_11_UCD_timsTOF_Cl4_m2BKF2	m2BKF2
	2022_04_11_UCD_timsTOF_Cl4_m2BKF3	m2BKF3
	2022_04_11_UCD_timsTOF_Cl4_m2DAHA1	m2DAHA1
	2022_04_11_UCD_timsTOF_Cl4_m2DAHA2	m2DAHA2
	2022_04_11_UCD_timsTOF_Cl4_m2DAHA3	m2DAHA3
	2022_04_11_UCD_timsTOF_Cl4_m2Per1	m2Per1
	2022_04_11_UCD_timsTOF_Cl4_m2Per2	m2Per2
	2022_04_11_UCD_timsTOF_Cl4_m2Per3	m2Per3
	2022_04_11_UCD_timsTOF_Cl4_m2MMI1	m2MMI1
	2022_04_11_UCD_timsTOF_Cl4_m2MMI2	m2MMI2
	2022_04_11_UCD_timsTOF_Cl4_m2MMI3	m2MMI3
	2022_04_11_UCD_timsTOF_Cl4_m3U1	m3U1
	2022_04_11_UCD_timsTOF_Cl4_m3U2	m3U2
	2022_04_11_UCD_timsTOF_Cl4_m3U3	m3U3
	2022_04_11_UCD_timsTOF_Cl4_m3C1	m3C1
	2022_04_11_UCD_timsTOF_Cl4_m3C2	m3C2
	2022_04_11_UCD_timsTOF_Cl4_m3C3	m3C3
	2022_04_11_UCD_timsTOF_Cl4_m3-118-1	m3-118-1
	2022_04_11_UCD_timsTOF_Cl4_m3-118-2	m3-118-2
	2022_04_11_UCD_timsTOF_Cl4_m3-118-3	m3-118-3
	2022_04_11_UCD_timsTOF_Cl4_m3-126-1	m3-126-1
	2022_04_11_UCD_timsTOF_Cl4_m3-126-2	m3-126-2

	2022_04_11_UCD_timsTOF_Cl4_m3-126-3	m3-126-3
	2022_04_11_UCD_timsTOF_Cl4_m3-138-1	m3-138-1
	2022_04_11_UCD_timsTOF_Cl4_m3-138-2	m3-138-2
	2022_04_11_UCD_timsTOF_Cl4_m3-138-3	m3-138-3
	2022_04_11_UCD_timsTOF_Cl4_m3-153-1	m3-153-1
	2022_04_11_UCD_timsTOF_Cl4_m3-153-2	m3-153-2
	2022_04_11_UCD_timsTOF_Cl4_m3-153-3	m3-153-3
	2022_04_11_UCD_timsTOF_Cl4_m3Per1	m3Per1
	2022_04_11_UCD_timsTOF_Cl4_m3Per2	m3Per2
	2022_04_11_UCD_timsTOF_Cl4_m3Per3	m3Per3
	2022_04_11_UCD_timsTOF_Cl4_m3MMI1	m3MMI1
	2022_04_11_UCD_timsTOF_Cl4_m3MMI2	m3MMI2
	2022_04_11_UCD_timsTOF_Cl4_m3MMI3	m3MMI3
	2022_04_11_UCD_timsTOF_Cl4_Blank_03	Blank_03
	2022_04_11_UCD_timsTOF_Cl4_Blank_04	Blank_04
	2022_04_11_UCD_timsTOF_Cl4_HeLa_tr04	HeLa_tr04
	2022_04_11_UCD_timsTOF_Cl4_HeLa_tr05	HeLa_tr05
	2022_04_11_UCD_timsTOF_Cl4_HeLa_tr06	HeLa_tr06
	2022_04_11_UCD_timsTOF_Cl4_Blank_05	Blank_05

Supplementary table 2. Max Quant search parameters for peptide identification.

Parameter	Value	Parameter	Value	Parameter	Value
Version	1.6.17.0	Separate LFQ in parameter groups	FALSE	MS/MS tol. (ITMS)	0.5 Da
Include contaminants	TRUE	Require MS/MS for LFQ comparisons	TRUE	Top MS/MS peaks per Da interval. (ITMS)	8
PSM FDR	0.01	Calculate peak properties	FALSE	Da interval. (ITMS)	100
PSM FDR Crosslink	0.01	Main search max. combinations	200	MS/MS deisotoping (ITMS)	FALSE
Protein FDR	0.01	Advanced site intensities	TRUE	MS/MS deisotoping tolerance (ITMS)	0.15
Site FDR	0.01	Write msScans table	FALSE	MS/MS deisotoping tolerance unit (ITMS)	Da

Parameter	Value	Parameter	Value	Parameter	Value
Use Normalized Ratios For Occupancy	TRUE	Write msmsScans table	TRUE	MS/MS higher charges (ITMS)	TRUE
Min. peptide Length	7	Write ms3Scans table	TRUE	MS/MS water loss (ITMS)	TRUE
Min. score for unmodified peptides	0	Write allPeptides table	TRUE	MS/MS ammonia loss (ITMS)	TRUE
Min. score for modified peptides	40	Write mzRange table	TRUE	MS/MS dependent losses (ITMS)	TRUE
Min. delta score for unmodified peptides	0	Write DIA fragments table	FALSE	MS/MS recalibration (ITMS)	FALSE
Min. delta score for modified peptides	6	Write pasefMsmsScans table	TRUE	MS/MS tol. (TOF)	40 ppm
Min. unique peptides	0	Write accumulatedPasefMsmsScans table	FALSE	Top MS/MS peaks per Da interval. (TOF)	10
Min. razor peptides	1	Max. peptide mass [Da]	4600	Da interval. (TOF)	100
Min. peptides	1	Min. peptide length for unspecific search	8	MS/MS deisotoping (TOF)	TRUE
Use only unmodified peptides and	TRUE	Max. peptide length for unspecific search	25	MS/MS deisotoping tolerance (TOF)	0.01
Modifications included in protein quantification	Oxidation (M);Acetyl (Protein N-term)	Razor protein FDR	TRUE	MS/MS deisotoping tolerance unit (TOF)	Da
Peptides used for protein quantification	Razor	Disable MD5	FALSE	MS/MS higher charges (TOF)	TRUE
Discard unmodified counterpart peptides	TRUE	Max mods in site table	3	MS/MS water loss (TOF)	TRUE

Parameter	Value	Parameter	Value	Parameter	Value
Label min. ratio count	2	Match unidentified features	FALSE	MS/MS ammonia loss (TOF)	TRUE
Use delta score	FALSE	Epsilon score for mutations		MS/MS dependent losses (TOF)	TRUE
iBAQ	FALSE	Evaluate variant peptides separately	TRUE	MS/MS recalibration (TOF)	FALSE
iBAQ log fit	FALSE	Variation mode	None	MS/MS tol. (Unknown)	20 ppm
Match between runs	TRUE	MS/MS tol. (FTMS)	20 ppm	Top MS/MS peaks per Da interval. (Unknown)	12
Matching time window [min]	0.7	Top MS/MS peaks per Da interval. (FTMS)	12	Da interval. (Unknown)	100
Match ion mobility window [indices]	0.05	Da interval. (FTMS)	100	MS/MS deisotoping (Unknown)	TRUE
Alignment time window [min]	20	MS/MS deisotoping (FTMS)	TRUE	MS/MS deisotoping tolerance (Unknown)	7
Alignment ion mobility window [indices]	1	MS/MS deisotoping tolerance (FTMS)	7	MS/MS deisotoping tolerance unit (Unknown)	ppm
Find dependent peptides	FALSE	MS/MS deisotoping tolerance unit (FTMS)	ppm	MS/MS higher charges (Unknown)	TRUE
Fasta file	2021.08.13 - Mouse FASTA - uniprotproteome_UP000000589+reviewed_yes.fasta	MS/MS higher charges (FTMS)	TRUE	MS/MS water loss (Unknown)	TRUE
Decoy mode	revert	MS/MS water loss (FTMS)	TRUE	MS/MS ammonia loss (Unknown)	TRUE

Parameter	Value	Parameter	Value	Parameter	Value
Include contaminants	TRUE	MS/MS ammonia loss (FTMS)	TRUE	MS/MS dependent losses (Unknown)	TRUE
Advanced ratios	TRUE	MS/MS dependent losses (FTMS)	TRUE	MS/MS recalibration (Unknown)	FALSE
Second peptides	TRUE	MS/MS recalibration (FTMS)	FALSE	Site tables	Oxidation (M)Sites.txt
Stabilize large LFQ ratios	TRUE				

Supplementary table 3. List of genes following a dose-response curve after maSigPro analysis (FDR < 0.05). The number in each EDC column refers to the cluster the gene belongs to. If a cell is empty, the gene was not dysregulated after EDC treatment. Note that the clusters numbers across compounds do not implicate the same type of curve.

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
1600014C10Rik		3																	
2310039H08Rik																			5
2610020C07Rik																		9	
2610021A01Rik		7																	
2610035D17Rik													9						8
2610301B20Rik																			4
2900009J06Rik																6			
4833411C07Rik				3															
4930430F08Rik		7																	
4930581F22Rik																			4
9430037G07Rik																	8		
AI597479																			5
AW146154						7												1	
Aamdc																6			
Abcb9																		7	
Abcf3	1																		
Abhd1																6			
Abhd3																3			
Abhd4																2			
Abi2																7			6
Acaa2			4	9															
Actr3																			6
Adam9																4			
Adgra3																			3

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Ahrr							2												
Ak7																			7
Aldh3a1													2						
Alg8																	9		
Ankrd39																		6	
Anxa10																			7
Ap1s3																			9
Aplp2		4																	
Arhgap11a																7			
Arl6ip4																6			
Arpc5l										1									
Asf1b				4															
Atg9a																	2		
Aurkb													3						
B3galnt1																		3	
B830012L14Rik											9								
Basp1				4															
Bbs5																			1
Bet1l																			5
Bivm													7						
Blmh	1																		
Blzf1																			5
Bmt2																8			
Bphl																			4
Brf2		1																	
Brip1os		1																	
Bst2							9												
Btbd3																			8
C1qbp																			6
C2cd4b																3			
Cacng1																8			
Cadm1																			9
Calcrl						5					8								
Capn2																			6
Capns1		1																	
Casp2																3		5	
Cav2																	1		
Cbarp											6								
Cbx1				2															
Cbx4																		1	
Cbx5																5			
Ccdc25																		1	
Ccdc43																	1		
Ccdc74a													6						
Ccdc82																			4

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Ccdc84		1																	
Ccdc90b					3														
Ccnb1				2															
Ccnd2				1															
Ccpg1os																			5
Ccrl2							8												
Cd2bp2		8																	
Cdc14a	9																		
Cdk12															1				
Cdk13																7			
Cdv3																			6
Cdyl																5			
Celf2		2																	
Cers6																			8
Chd6																4			
Chil1																		8	
Cideb																			5
Cldn18										2									
Cldn7								2								3			
Cmc1																			5
Cmc4																			7
Cnbd2													5						4
Col4a1						3													
Col4a2																9			
Cox7c-ps1																	7		
Cp													1						
Cr1l																		2	
Creld1																8			
Csnk1g3																			6
Cstf2t																	5		
Ctsz												1							
Cxcl14		5																	
Cyb5b				7															
Cyp1a1							5	4					2						
Cyp1b1							3	3					2						
Cyria																5			
Dab1																7			
Dap				5															
Ddit3				5															
Ddr1			2																
Ddx6		8																	
Desi2																			6
Dhrs7b																3			
Dipk1a																			7
Dmxl2																			3

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Dnajb1			3																
Dnajb11																			3
Dnajc8		3		9															
Dnase1l1	4																		
Dop1a																			2
Dpf2		1																	
Dpm1																			1
Dusp2							4												
Echdc3																	9		
Eda2r		3																	
Eif2ak3																			9
Eif5a2																4			
Elavl1																5			
Eno1																			4
Enox2																			7
Epb41l4a																			3
Epb41l5																	5		
Epn2						1													
Exoc6				8															
Fam149b																			6
Fam20a															2				
Fam50a																	2		
Fanci							7									9			
Fbxo17																	9		
Fbxo44																			7
Fbxo9																		1	
Fbxw2												4							
Fetub																	6		
Fgfr1											6								
Fkbp1a																9			
Flvcr1																	3		
Fmo5													4						
G0s2									1										
G430095 P16Rik																			4
Galk2																6			
Gar1																	2		
Gas2l1					4														
Gata6																4			
Gbp7																	5		
Gem																	3		
Gfod2																	2		
Glb1l2									7										
Glod4																			5
Glrx5																			3
Gm10323		3																	

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Gm10941																	5		
Gm11635									4										
Gm11688																			4
Gm14648																			5
Gm15191																	7		4
Gm15446																	1		
Gm15501																			5
Gm20765																4			
Gm21312																4			
Gm22513																			8
Gm25099																	8		
Gm25395				3															
Gm26532																		3	
Gm28661			9																
Gm29170																			4
Gm32885				4															
Gm4221																			4
Gm42420																			9
Gm42826										3									
Gm43201																	6		
Gm45871																			1
Gm5141																		1	
Gm6069																	7		4
Gm7876			7																
Gm7908				3															
Gm7972																	1		
Gm8318		2	8																
Gm8355																			4
Gmpr																			2
Gnpat																			7
Golph3																	4		
Gorasp2		4																	
Gpm6b																	6		
Gpr107																			1
Gpr137b				6															
Gpr137b-ps																	7		
Gprc5a														2					
Gpsm1																	2		
Grem2																4			
Gripap1																	2		
Gsn																	7		3
Gspt1																			6
Gstm1				5															
H3f3b									3										
Haghl																6			

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Haus4																3			
Higd2a	7																		
Hnrnpdl																	1		
Hnrnpm																	9		
Hpcal4																		1	
Hprt																			4
Hs3st1																4			
Hsp90b1																			7
Hspbap1				2															
Htr1d																			9
Iars																7			
Igfbp6																2			
Ikbke																	2		
Il34																	9		2
Ildr1																			8
Ilf3																			9
Inca1																			5
Ing5			4																
Inpp5f																4			
Ipo8																			6
Isg20																			7
Itgae											1								
Itgav											5								
Itsn2																	1		8
Josd1																6			
Kank1		1																	
Kat7		1																	
Kif1b																			6
Kif23				2															
Klf6			1																
Klhl11									8										
Klhl28																	1		
Krit1																	1		
Larp7				2															
Lbp		6																	
Lbr				2															
Lgi2																		7	
Lipt1	5																		
Lman2l				3												1			
Lmbrd1																	1		
Lnpk																	1		
Lonrf3									2										
Lor				1															
Lrch4																7			
Lrp12																			6

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Luc7l				5															
Malsu1																			1
Man1a	2																		
Man1b1																	6		
Map1b				5							7								
Mapkapk2																1			
Marchf7																	4		
Mark4	8																		
Marveld2																		4	
Med31																			7
Mettl8															7				
Mfsd8																		5	
Mgme1		8																	
Mgp									5										
Mid1				6															
Minpp1																			3
Mir1249																			1
Mir125a																			1
Mir1291																	7		5
Mir135a-1																			7
Mir141																			1
Mir181b-2																			1
Mir1981																			1
Mir224																			1
Mir27a																			7
Mir30a																			7
Mir324																			1
Mir340												6							1
Mir351																			7
Mirlet7d																			1
Mirt1		9																	
Mmp17															5				
Mospd2																	1		
Mrm1																		4	
Mroh1																	2		
Mrps18b											2								
Mrps22				8															
Mrps34		5																	
Msantd3																			2
Mtm1																	4		
Mtmr10																			3
Mto1											2								
Mtres1				7															
Mxi1			3																
Mymx																		6	

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Myo5c												3							
Naa15																			3
Naa20		3																	
Nat9				6															
Nbeal1																	5		
Ncbp2											2								
Nckap1																			3
Ncoa4				4															
Ndc80				2															
Ndufv3																			7
Nectin3															3				
Net1			3																
Nfyc				2															
Nif3l1											4								
Nkd1									6										
Nkiras1						2													
Nktr																			1
Nog																4			
Nop53		5																	
Nop56																	8		8
Noxo1					1														
Npc2																			1
Nqo1							1	1											
Nrd1																			6
Nsd1	6																		
Nt5c2																			2
Nubpl																			5
Nucks1																			6
Olfr920					6														
Osbpl5																			6
Pank4				3															
Pard3b																	9		
Parm1		6																	
Parpbp																			2
Pbld1																			2
Pcsk1n				3															
Pcx																2			
Pcyt2																			2
Pde3a																7			
Pde4d				5															
Pdik1l						4													
Pdlim1																			7
Pdlim5																			6
Pgp																2			
Pi4k2a				3															

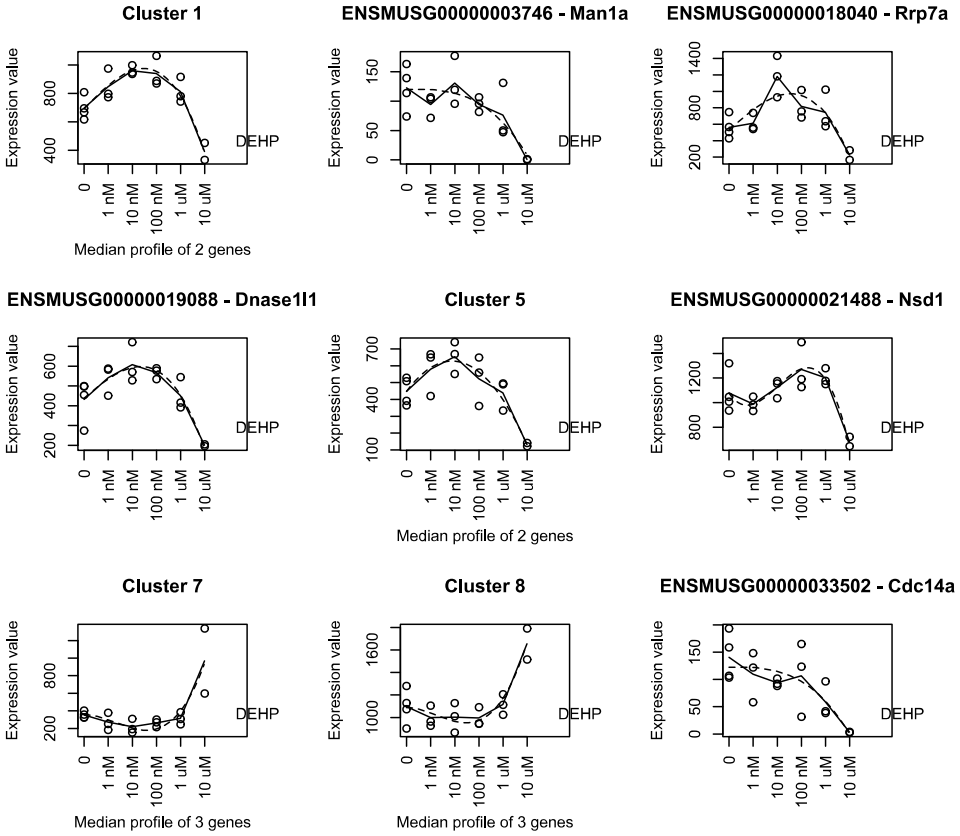
	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Pigc																		1	
Pigg																		5	
Pitpna																			6
Pkm	8																		
Pnrc1			5																
Poli																			5
Polk				6															
Polr3b																			2
Ppp1r18		4																	
Ppp1r2																			9
Ppp1r3c																			7
Prdm9																			4
Prdx4																			5
Prmt2																1			
Prpf8		5																	
Prpsap1		5																	
Psd3																	7		
Psmd4																3			
Psme3																			6
Ptpn23				3															
Ptprn																	6	4	
Qtrt2																			3
R3hdm2					2						3								
Rab4a		5																	
Rab5c		7																	
Rab7b																			4
Ramp1																			2
Ranbp2																	1		
Rassf4																			7
Rbm18															4				
Rbm5																6			
Recql5																	2		
Reep6		1																	
Rhcg																6			
Rhoa																			3
Rhoc																			4
Rmdn3																2			
Rnasek				3															
Rnf19b																			6
Rpa3											2								
Rpl7l1			6																
Rps4l																3			
Rps6kb1																	4		
Rps6kl1									4										
Rrp7a	3																		

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Rusf1																2			
Sall2																			3
Scamp2		1																	
Schip1																		6	
Scyl1																	2		
Sdc2																	5		
Sdf2l1																	3		
Secisbp2l	7																		
Sel1l		6		6															
Selenow		5																	
Serhl																			5
Serpinb9b	5																		
Serpinh1																			6
Setd5			5																
Sh3bgrl																	4		
Sike1																	5		
Slc20a2																			3
Slc22a23																			6
Slc24a3																4			
Slc25a3		5																	
Slc25a34				3															
Slc25a39				3															
Slc30a4																	3		
Slc34a2																4			
Slc38a1				3															
Slc39a7				3															
Slc3a2				5															
Slc41a1																7			
Slc44a1																			6
Slc48a1				5															
Slc6a2	7																		
Slc7a7				2															
Smap1																9			
Smg9																2			
Smim1																			1
Smim40																4			
Smpdl3b																	9		
Snapc3																4			
Snhg5																	8		8
Snn									4										
Snx1		8																	
Snx13																	1		
Snx5																			1
Spcs3																	5		
Spop												5				3			

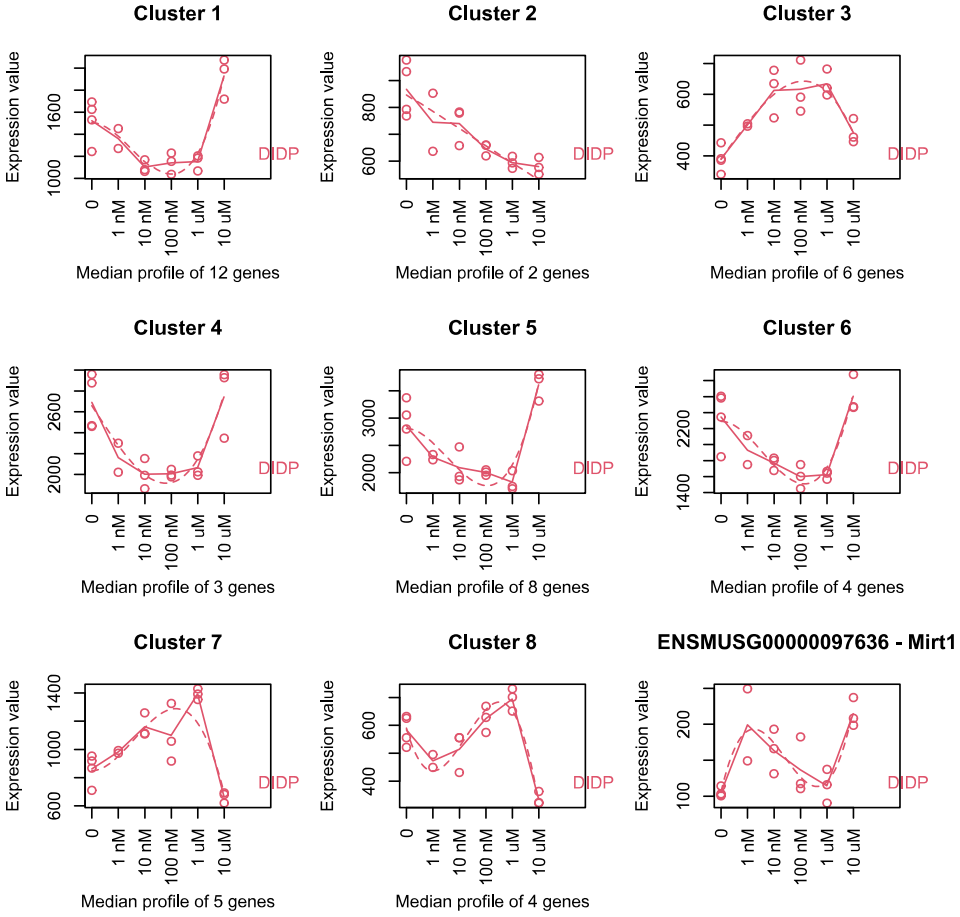
	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Sprr2g													8						
Sptbn2																		7	
Srp72																			1
Srsf3				2															
Srxn1				5															
Syngr2																6			
Sys1				7															
Syt2																	6		
Tbc1d23																			1
Tbl2			3																
Tek																5			
Tgfbi		1																	
Tgfbr2																			6
Thnsl1																	1		
Tiparp							6						2						
Tmbim1				5															
Tmem117																	6		
Tmem199									9										
Tmem35b																			5
Tmem38b																	8		8
Tmem79																			3
Tnfaip1																			3
Tnni3																4			
Tomm5											2								
Tor3a																			2
Trak1															6				
Trim45														1					
Trmt10c		7																	
Trmt44																	9		
Tshr																			8
Tspan7		6																	
Ttc3				5															
Tubb2b																4			
Tvp23b																			1
Uba2				2															
Ubap1	8																		
Ufl1																			4
Ufsp2																			5
Usp33																	1		
Washc4																	4		
Wdr37																		5	
Xpo7																			9
Xrcc6		7																	
Yars		1																	
Yrdc												2				1			

	Phthalates				PAHs				OPFRs			PCBs				Reference			
Gene name	DEHP	DIDP	DINP	DnOP	BAA	BAP	BKF	DAHA	BADP	DMMP	TPP	PCB118	PCB126	PCB138	PCB153	MMI Exp2	MMI Exp3	PER Exp2	PER Exp3
Ywhae																			3
Zbtb20				6															
Zc3h13																1			
Zcchc10																		3	
Zcchc24					5														
Zdhhc15																			5
Zfp236											3								
Zfp317																	1		
Zfp418																7			
Zfp53																	1		
Zfp715																	4		
Zfp839											3								
Zfp867																			5
Zfp87		3																	
Zfp930																	1		
Zfp934																	1		
Zfp943																			1
Zfp946																	1		
Zfp955a																			5
Zfp958																			5
Zfp971																			2
Zfyve21																			5
Zswim4		1																	
Zzz3				2															
mt-Co2																			4
mt-Nd6					7	6													
mt-Ta																	7		
mt-Tc																	7		5
mt-Td																	7		5
mt-Te																	7		5
mt-Tf																	7		5
mt-Tg																	7		5
mt-Th																	7		5
mt-Ti																	7		5
mt-Tm																	7		5
mt-Tn																	7		5
mt-Tq																	7		5
mt-Tr																			5
mt-Ts2																	5		5
mt-Tv																	7		5
mt-Tw																	7		5
ENSMUSG 00002075017																	3		
ENSMUSG 00002075686																	7		

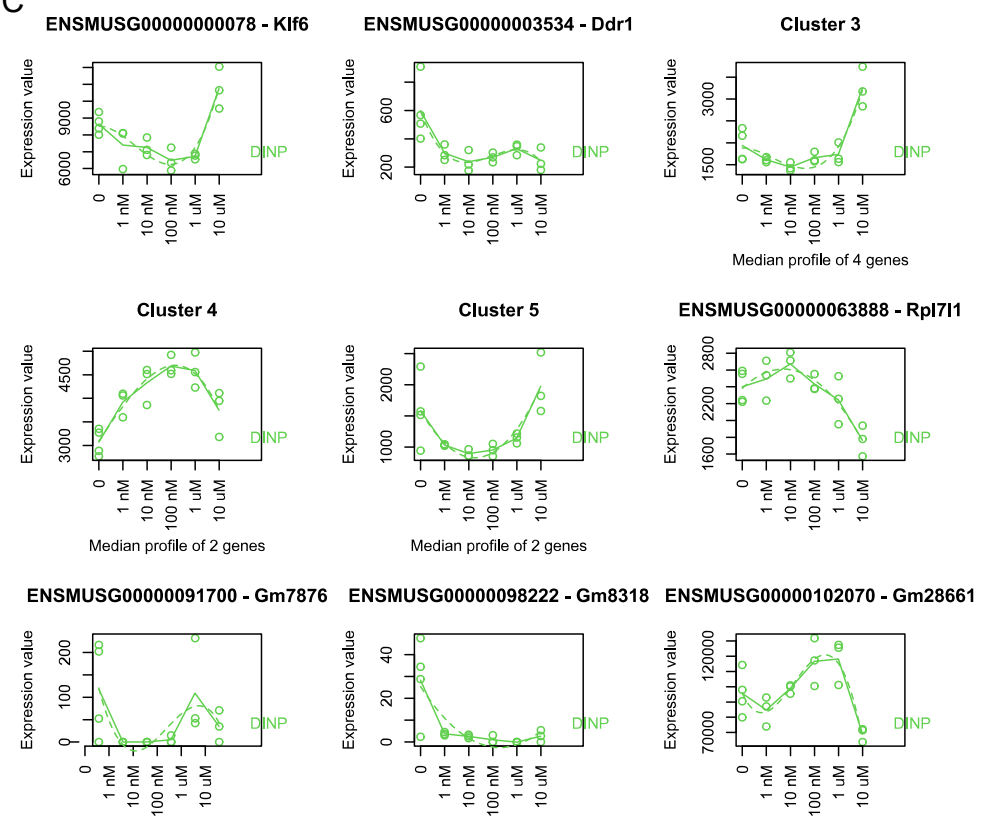
A


Cluster 1
ENSMUSG00000003746 - Man1a
ENSMUSG00000018040 - Rrp7a
ENSMUSG00000019088 - Dnase1l1
Cluster 5
ENSMUSG00000021488 - Nsd1
Cluster 7
Cluster 8
ENSMUSG00000033502 - Cdc14a
Expression value
DEHP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 2 genes
Median profile of 3 genes

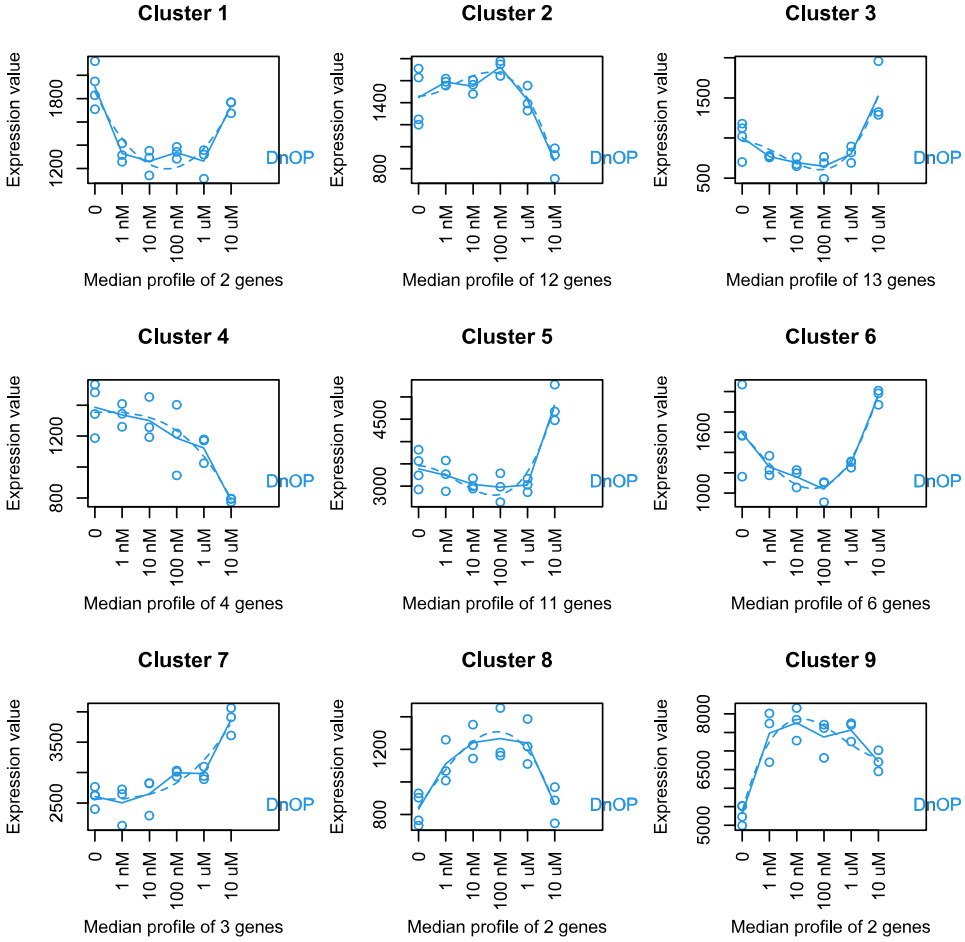
B


Cluster 1
Expression value
1000
1600
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 12 genes
Cluster 2
Expression value
600
800
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 2 genes
Cluster 3
Expression value
400
600
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 6 genes
Cluster 4
Expression value
2000
2600
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 3 genes
Cluster 5
Expression value
2000
3000
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 8 genes
Cluster 6
Expression value
1400
2200
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 4 genes
Cluster 7
Expression value
600
1000
1400
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 5 genes
Cluster 8
Expression value
400
600
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 4 genes
ENSMUSG00000097636 - Mirt1
Expression value
100
200
DIDP
0
1 nM
10 nM
100 nM
1 uM
10 uM

C

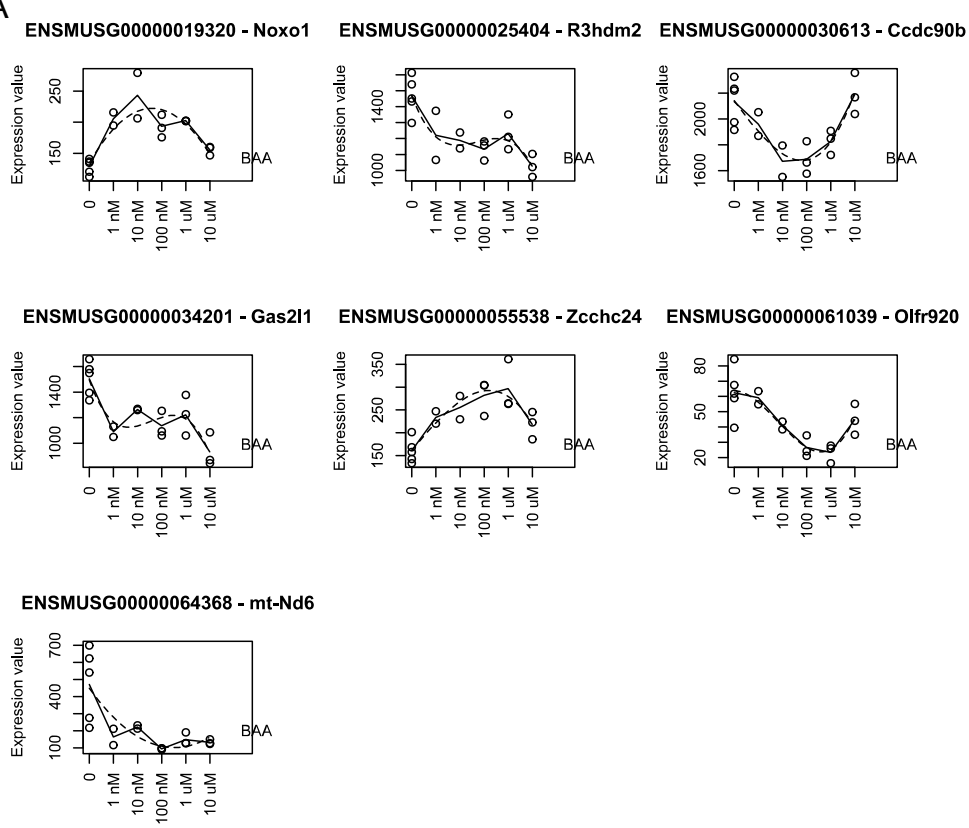

ENSMUSG00000000078 - Klf6
ENSMUSG00000003534 - Ddr1
Cluster 3
Cluster 4
Cluster 5
ENSMUSG00000063888 - Rpl7l1
ENSMUSG00000091700 - Gm7876
ENSMUSG00000098222 - Gm8318
ENSMUSG00000102070 - Gm28661
Expression value
DINP
0
1 nM
10 nM
100 nM
1 uM
10 uM
Median profile of 4 genes
Median profile of 2 genes
Median profile of 2 genes

D

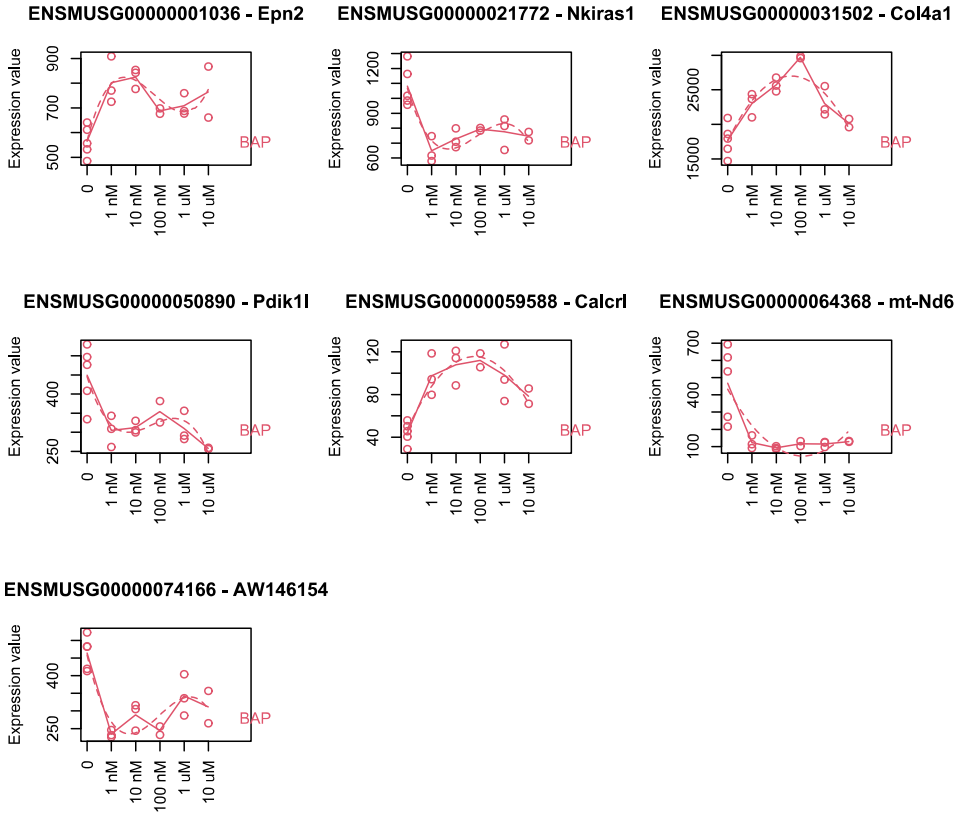


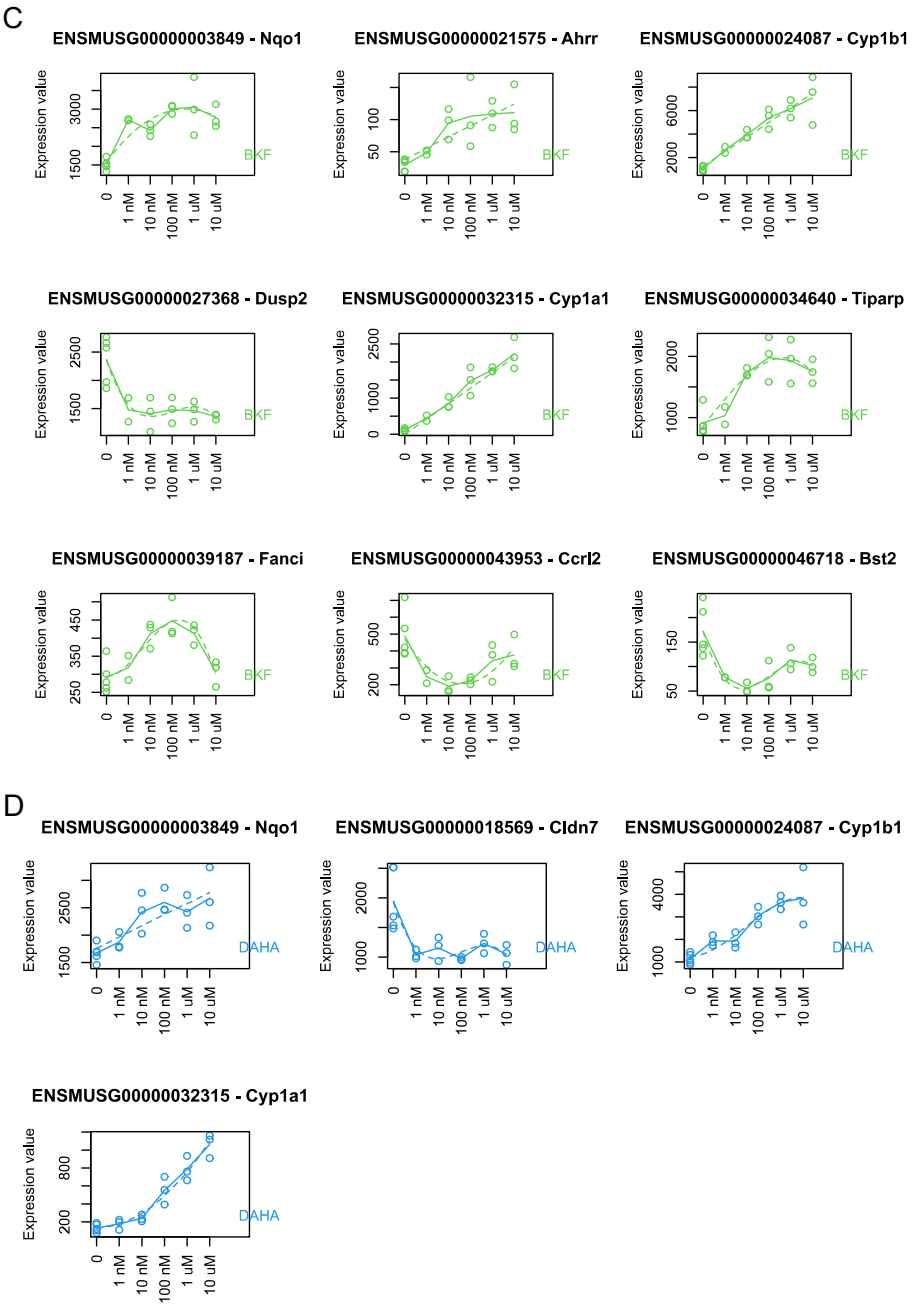
Supplementary figure 1. Results of dose series analysis of gene expression on phthalates (A = DEHP; B = DIDP; C = DINP; D = DnOP). Each plot shows the expression values of each replicate (dots) and fit (dotted line) of genes for which FDR < 0.05. The Ensembl gene ID and gene name are reported at the top of each panel. If a cluster comprises more genes, only the cluster name is reported (the genes are in Supplementary table 3). The x-axis reports the dose range used (nM = nanomoles per liter, uM = micromoles per liter).

A

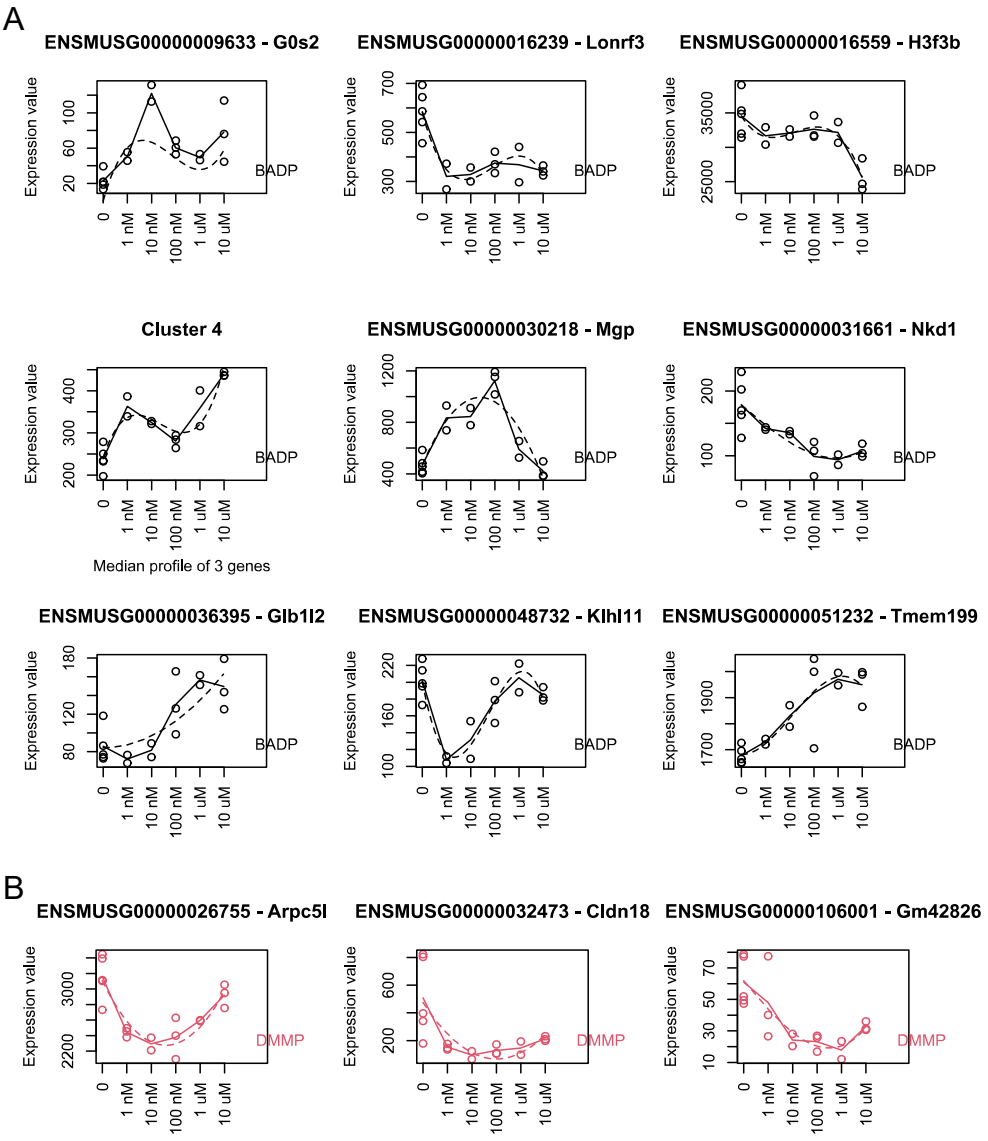

ENSMUSG00000019320 - Noxo1
ENSMUSG00000025404 - R3hdm2
ENSMUSG00000030613 - Ccdc90b
ENSMUSG00000034201 - Gas2l1
ENSMUSG00000055538 - Zcchc24
ENSMUSG00000061039 - Olfr920
ENSMUSG00000064368 - mt-Nd6
Expression value
BAA
0
1 nM
10 nM
100 nM
1 uM
10 uM

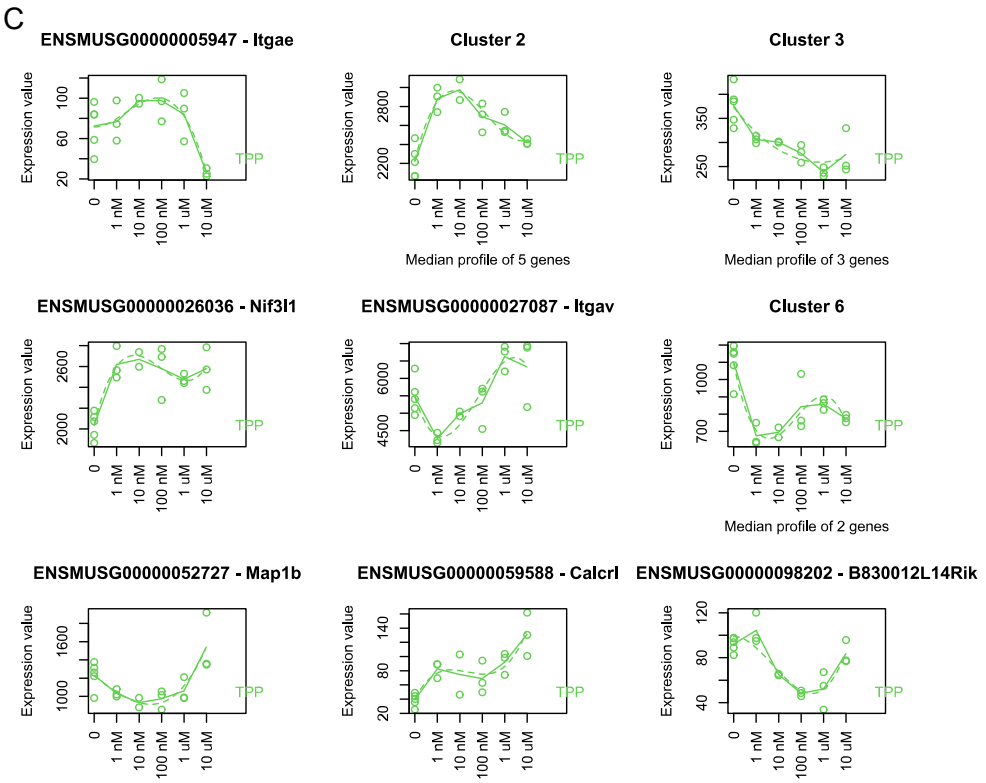
B


ENSMUSG00000001036 - Epn2
ENSMUSG00000021772 - Nkiras1
ENSMUSG00000031502 - Col4a1
ENSMUSG00000050890 - Pdik1l
ENSMUSG00000059588 - Calcrl
ENSMUSG00000064368 - mt-Nd6
ENSMUSG00000074166 - AW146154
Expression value
BAP
0
1 nM
10 nM
100 nM
1 uM
10 uM

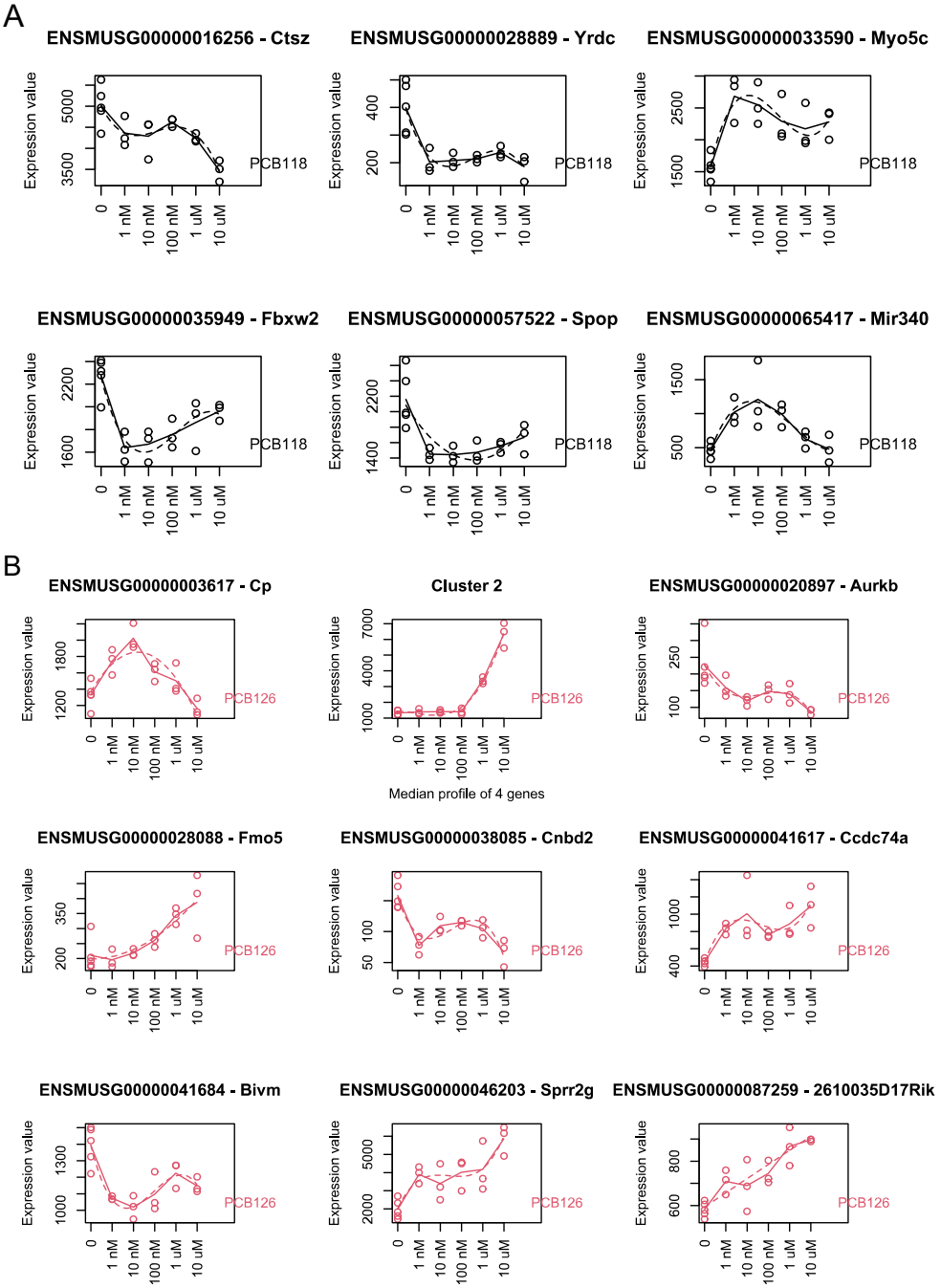


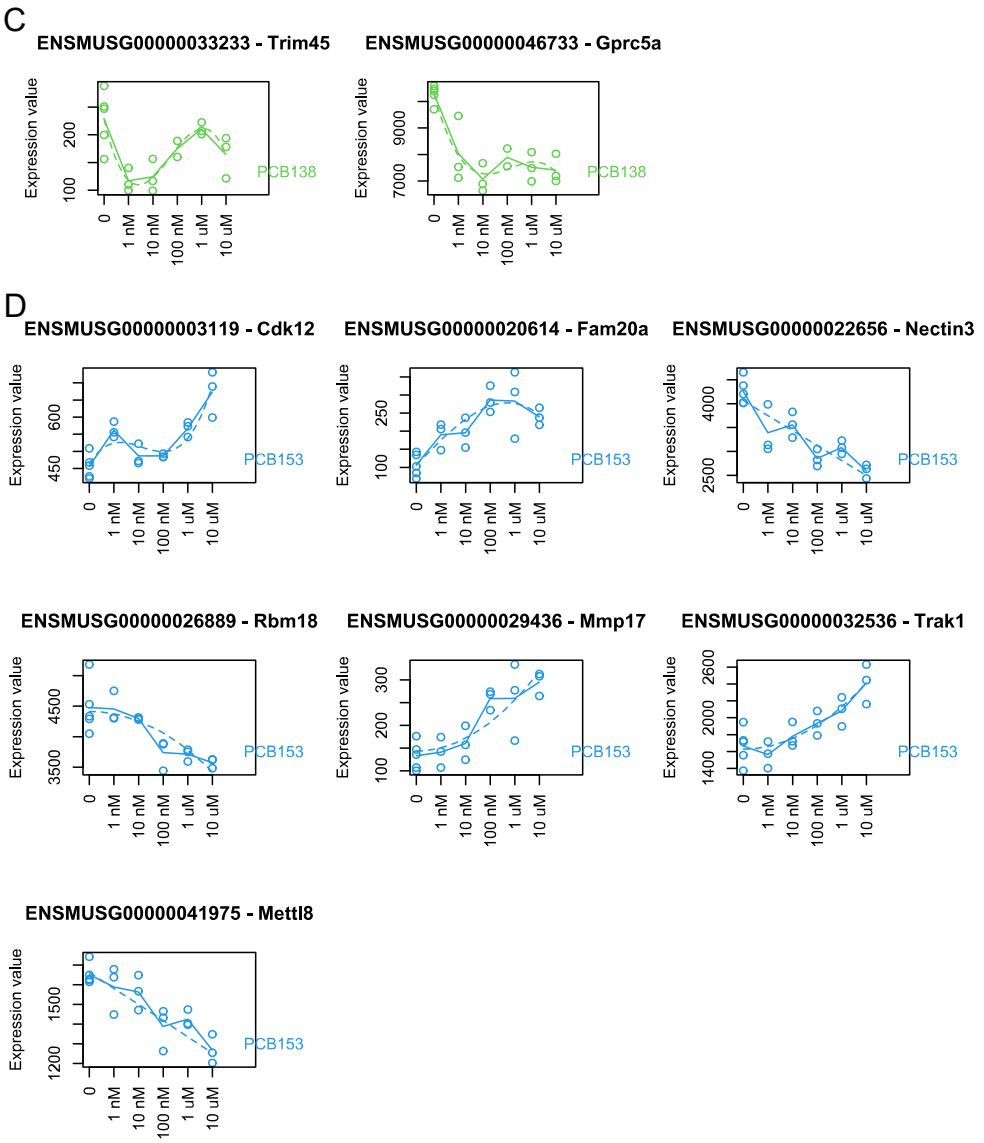
Supplementary figure 2. Results of dose series analysis of gene expression on PAHs (A = BAA; B = BAP, C = BKF; D = DAHA). Each plot shows the expression values of each replicate (dots) and fit (dotted line) of genes for which FDR < 0.05. The Ensembl gene ID and gene name are reported at the top of each panel. The x-axis reports the dose range used (nM = nanomoles per liter, uM = micromoles per liter).


A
ENSMUSG00000009633 - G0s2
ENSMUSG00000016239 - Lonrf3
ENSMUSG00000016559 - H3f3b
Cluster 4
ENSMUSG00000030218 - Mgp
ENSMUSG00000031661 - Nkd1
Median profile of 3 genes
ENSMUSG00000036395 - Glb1l2
ENSMUSG00000048732 - Klhl11
ENSMUSG00000051232 - Tmem199
B
ENSMUSG00000026755 - Arpc5l
ENSMUSG00000032473 - Cldn18
ENSMUSG00000106001 - Gm42826
Expression value
BADP
DMMP
0
1 nM
10 nM
100 nM
1 uM
10 uM

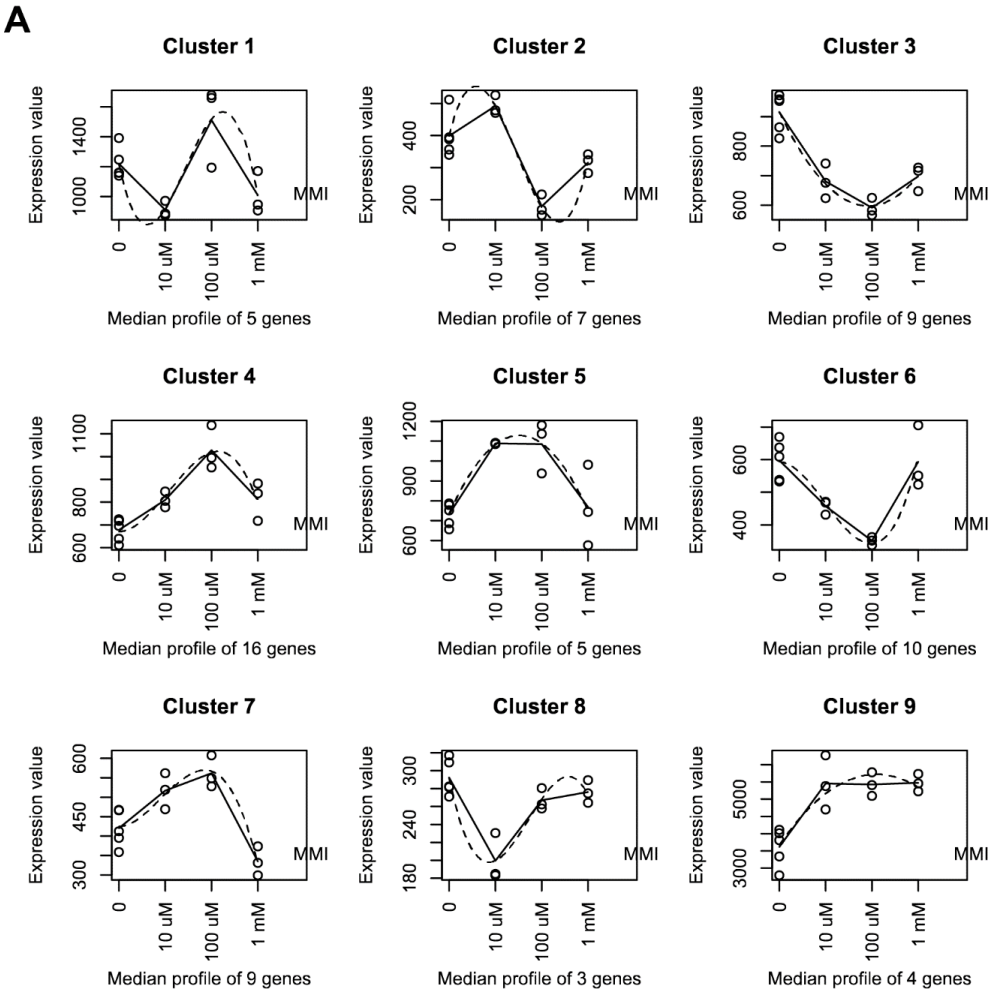


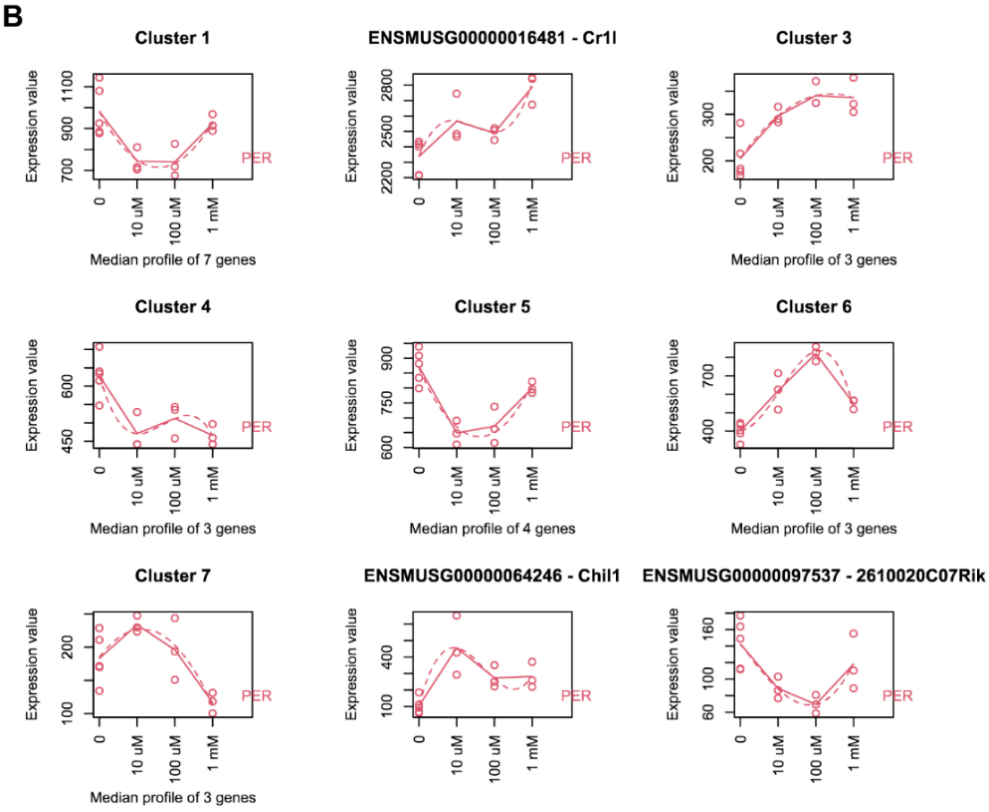
Supplementary figure 3. Results of dose series analysis of gene expression on OPFRs (A = BADP; B = DMMP; C = TPP). Each plot shows the expression values of each replicate (dots) and fit (dotted line) of genes for which FDR < 0.05. The Ensembl gene ID and gene name are reported at the top of each panel. If a cluster comprises more genes, only the cluster name is reported (the genes are in Supplementary table 2). The x-axis reports the dose range used (nM = nanomoles per liter, uM = micromoles per liter).


A
ENSMUSG00000016256 - Ctsz
ENSMUSG00000028889 - Yrdc
ENSMUSG00000033590 - Myo5c
ENSMUSG00000035949 - Fbxw2
ENSMUSG00000057522 - Spop
ENSMUSG00000065417 - Mir340
Expression value
PCB118
0
1 nM
10 nM
100 nM
1 uM
10 uM
B
ENSMUSG00000003617 - Cp
Cluster 2
ENSMUSG00000020897 - Aurkb
ENSMUSG00000028088 - Fmo5
ENSMUSG00000038085 - Cnbd2
ENSMUSG00000041617 - Ccdc74a
ENSMUSG00000041684 - Bivm
ENSMUSG00000046203 - Sprr2g
ENSMUSG00000087259 - 2610035D17Rik
PCB126
Median profile of 4 genes

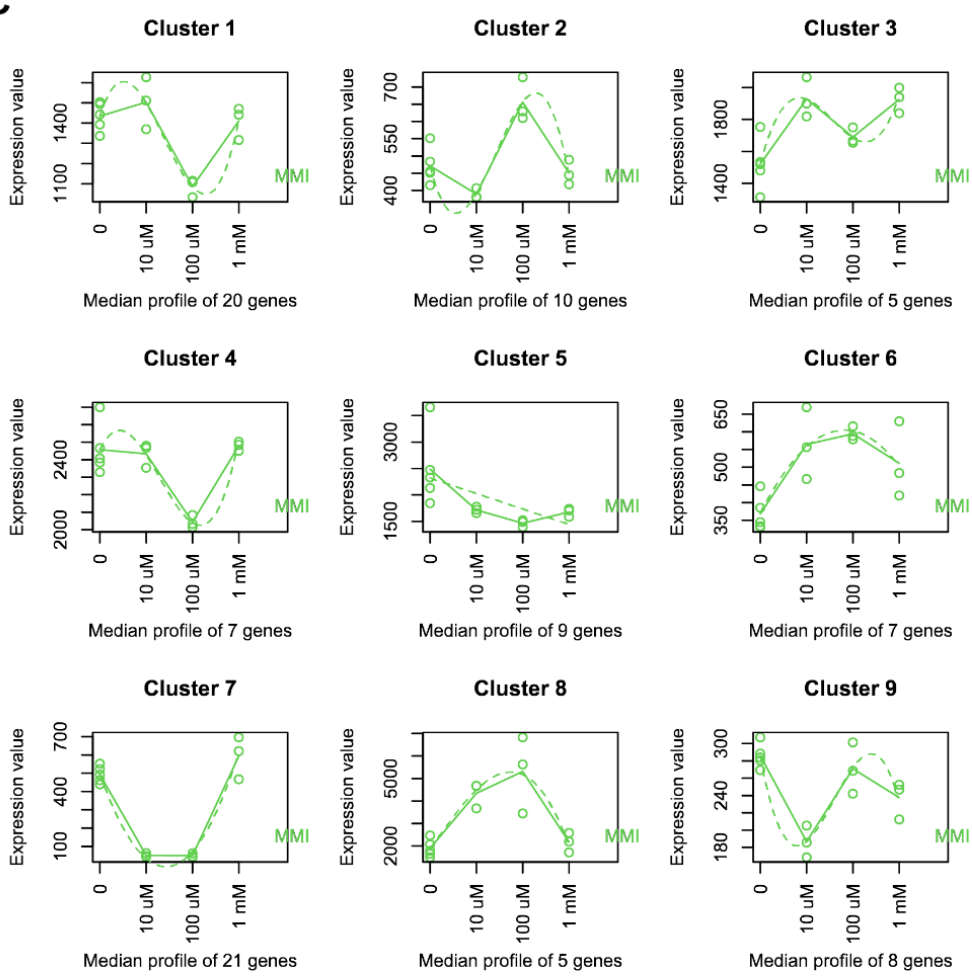


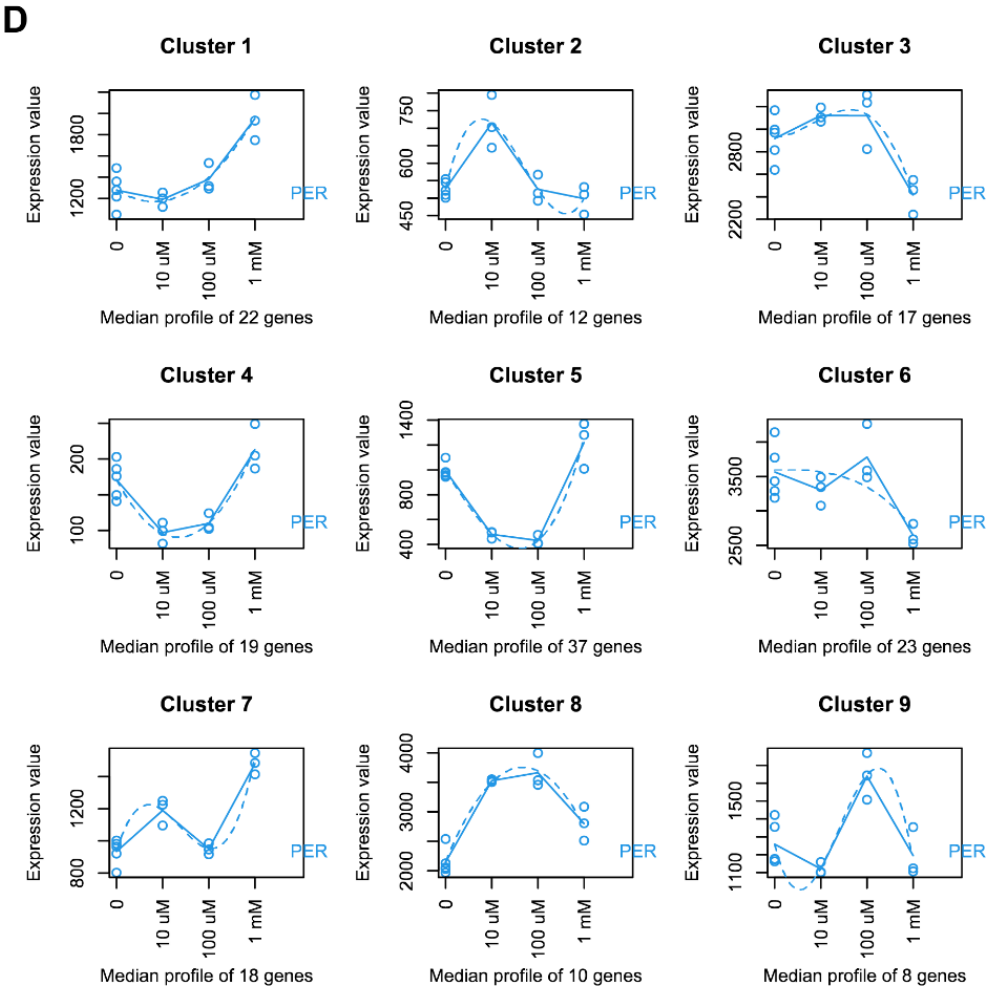
Supplementary figure 4. Results of dose series analysis of gene expression on PCBs (A = PCB118; B = PCB126; C = PCB138; D = PCB153). Each plot shows the expression values of each replicate (dots) and fit (dotted line) of genes for which FDR < 0.05. The Ensembl gene ID and gene name are reported at the top of each panel. If a cluster comprises more genes, only the cluster name is reported (the genes are in Supplementary table 3). The x-axis reports the dose range used (nM = nanomoles per liter, uM = micromoles per liter).


A
Cluster 1
Expression value
1000
1400
MMI
0
10 uM
100 uM
1 mM
Median profile of 5 genes
Cluster 2
Expression value
200
400
MMI
0
10 uM
100 uM
1 mM
Median profile of 7 genes
Cluster 3
Expression value
600
800
MMI
0
10 uM
100 uM
1 mM
Median profile of 9 genes
Cluster 4
Expression value
600
800
1100
MMI
0
10 uM
100 uM
1 mM
Median profile of 16 genes
Cluster 5
Expression value
600
900
1200
MMI
0
10 uM
100 uM
1 mM
Median profile of 5 genes
Cluster 6
Expression value
400
600
MMI
0
10 uM
100 uM
1 mM
Median profile of 10 genes
Cluster 7
Expression value
300
450
600
MMI
0
10 uM
100 uM
1 mM
Median profile of 9 genes
Cluster 8
Expression value
180
240
300
MMI
0
10 uM
100 uM
1 mM
Median profile of 3 genes
Cluster 9
Expression value
3000
5000
MMI
0
10 uM
100 uM
1 mM
Median profile of 4 genes


B
Cluster 1
Expression value
700 900 1100
PER
0 10 uM 100 uM 1 mM
Median profile of 7 genes
ENSMUSG00000016481 - Cr1l
Expression value
2200 2500 2800
PER
0 10 uM 100 uM 1 mM
Cluster 3
Expression value
200 300
PER
0 10 uM 100 uM 1 mM
Median profile of 3 genes
Cluster 4
Expression value
450 600
PER
0 10 uM 100 uM 1 mM
Median profile of 3 genes
Cluster 5
Expression value
600 750 900
PER
0 10 uM 100 uM 1 mM
Median profile of 4 genes
Cluster 6
Expression value
400 700
PER
0 10 uM 100 uM 1 mM
Median profile of 3 genes
Cluster 7
Expression value
100 200
PER
0 10 uM 100 uM 1 mM
Median profile of 3 genes
ENSMUSG00000064246 - Chil1
Expression value
100 400
PER
0 10 uM 100 uM 1 mM
ENSMUSG00000097537 - 2610020C07Rik
Expression value
60 100 160
PER
0 10 uM 100 uM 1 mM

C


Cluster 1
Expression value
1100
1400
MMI
0
10 uM
100 uM
1 mM
Median profile of 20 genes
Cluster 2
Expression value
400
550
700
MMI
0
10 uM
100 uM
1 mM
Median profile of 10 genes
Cluster 3
Expression value
1400
1800
MMI
0
10 uM
100 uM
1 mM
Median profile of 5 genes
Cluster 4
Expression value
2000
2400
MMI
0
10 uM
100 uM
1 mM
Median profile of 7 genes
Cluster 5
Expression value
1500
3000
MMI
0
10 uM
100 uM
1 mM
Median profile of 9 genes
Cluster 6
Expression value
350
500
650
MMI
0
10 uM
100 uM
1 mM
Median profile of 7 genes
Cluster 7
Expression value
100
400
700
MMI
0
10 uM
100 uM
1 mM
Median profile of 21 genes
Cluster 8
Expression value
2000
5000
MMI
0
10 uM
100 uM
1 mM
Median profile of 5 genes
Cluster 9
Expression value
180
240
300
MMI
0
10 uM
100 uM
1 mM
Median profile of 8 genes



Supplementary figure 5. Results of dose series analysis of gene expression on thyroid hormone synthesis inhibiting compounds methimazole (MMI) and sodium perchlorate (PER) (A = MMI Experiment 2; B = PER Experiment 2; C = MMI Experiment 3; D = PER Experiment 3). Each plot shows the expression values of each replicate (dots) and fit (dotted line) of genes for which FDR < 0.05. If a cluster is composed of a single gene, the Ensembl gene ID and gene name are reported at the top of the panel. If a cluster comprises more genes, only the cluster name is reported (the genes are reported in Supplementary figure 3). The x-axis reports the dose range used (uM = micromoles per liter; mM = millimoles per liter).

6. References

1. Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, et al. EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. Endocr Rev. 2015;36(6):E1-E150, doi:https://doi.org/10.1210/er.2015-1010.
2. Hiller-Sturmhöfel S, Bartke A. The endocrine system: an overview. Alcohol Health Res World. 1998;22(3):153-64.
3. Crantz FR, Silva JE, Larsen PR. An analysis of the sources and quantity of 3,5,3'-triiodothyronine specifically bound to nuclear receptors in rat cerebral cortex and cerebellum. Endocrinology. 1982;110(2):367-75, doi:https://doi.org/10.1210/endo-110-2-367.
4. Schweizer U, Kohrle J. Function of thyroid hormone transporters in the central nervous system. Biochim Biophys Acta. 2013;1830(7):3965-73, doi:https://doi.org/10.1016/j.bbagen.2012.07.015.
5. Bianco AC, Salvatore D, Gereben Bz, Berry MJ, Larsen PR. Biochemistry, Cellular and Molecular Biology, and Physiological Roles of the Iodothyronine Selenodeiodinases. Endocrine Reviews. 2002;23(1):38-89, doi:https://doi.org/10.1210/edrv.23.1.0455.
6. Dentice M, Marsili A, Zavacki A, Larsen PR, Salvatore D. The deiodinases and the control of intracellular thyroid hormone signaling during cellular differentiation. Biochim Biophys Acta. 2013;1830(7):3937-45, doi:https://doi.org/10.1016/j.bbagen.2012.05.007.
7. Larsen PR, Zavacki AM. The role of the iodothyronine deiodinases in the physiology and pathophysiology of thyroid hormone action. Eur Thyroid J. 2012;1(4):232-42, doi:https://doi.org/10.1159/000343922.
8. Sabatino L, Vassalle C, Del Seppia C, Iervasi G. Deiodinases and the Three Types of Thyroid Hormone Deiodination Reactions. Endocrinol Metab (Seoul). 2021;36(5):952-64, doi:https://doi.org/10.3803/EnM.2021.1198.
9. Luster M, Duntas LH, Wartofsky L. The Thyroid and Its Diseases: Springer; 2019.
10. U.S EPA. Phthalates Action Plan2012. Available from: https://www.epa.gov/sites/default/files/2015-09/documents/phthalates_actionplan_revised_2012-03-14.pdf.
11. Babich MA. Overview of phthalates toxicity. US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/phthalover.pdf.
12. Wittassek M, Koch HM, Angerer J, Bruning T. Assessing exposure to phthalates - the human biomonitoring approach. Mol Nutr Food Res. 2011;55(1):7-31, doi:https://doi.org/10.1002/mnfr.201000121.
13. Versar Inc. Review of Exposure Data and Assessments for Select Dialkyl Ortho-PhthalatesFebruary 2010. Available from: https://www.cpsc.gov/s3fs-public/pthalexp.pdf.
14. Hyunok Choi, Roy Harrison, Hannu Komulainen, Saborit JMD. Polycyclic aromatic hydrocarbons. 2010. Geneva: World Health Organization. WHO Guidelines for Indoor Air Quality: Selected Pollutants. Available from: https://www.ncbi.nlm.nih.gov/books/NBK138709/.

15. World Health Organization. Regional Office for Europe. Air quality guidelines for Europe, 2nd ed.: WHO Regional Publications, European Series; 91; 2000. Available from: https://apps.who.int/iris/handle/10665/107335.
16. Agency for Toxic Substances and Disease Registry (ATSDR) (US). Toxicological Profile for Polychlorinated Biphenyls (PCBs). Department of Health and Human Services, Public Health Service; 2000. Report No.: NTIS/02928747_a.
17. Ghisari M, Bonefeld-Jorgensen EC. Effects of plasticizers and their mixtures on estrogen receptor and thyroid hormone functions. Toxicol Lett. 2009;189(1):67-77, doi:https://doi.org/10.1016/j.toxlet.2009.05.004.
18. Bereketoglu C, Pradhan A. Plasticizers: negative impacts on the thyroid hormone system. Environ Sci Pollut Res Int. 2022;29(26):38912-27, doi:https://doi.org/10.1007/s11356-022-19594-0.
19. Huang P-C, Kuo P-L, Guo Y-L, Liao P-C, Lee C-C. Associations between urinary phthalate monoesters and thyroid hormones in pregnant women. Human Reproduction. 2007;22(10):2715-22, doi:https://doi.org/10.1093/humrep/dem205.
20. Teles M, Oliveira M, Pacheco M, Santos MA. Endocrine and metabolic changes in Anguilla anguilla L. following exposure to β-naphthoflavone—a microsomal enzyme inducer. Environment International. 2005;31(1):99-104, doi:https://doi.org/10.1016/j.envint.2004.07.003.
21. Johnson S, McKillop D, Miller J, Smith IK. The Effects on Rat Thyroid Function of an Hepatic Microsomal Enzyme Inducer. Human & Experimental Toxicology. 1993;12(2):153-8, doi:https://doi.org/10.1177/096032719301200021.
22. Song M, Kim Y-J, Park Y-K, Ryu J-C. Changes in thyroid peroxidase activity in response to various chemicals. Journal of Environmental Monitoring. 2012;14(8):2121-6, doi:10.1039/C2EM30106G.
23. Agency for Toxic Substances and Disease Registry (ATSDR) (US). 3.6 ENDOCRINE DISRUPTION. 2000. In: Toxicological Profile for Polychlorinated Biphenyls (PCBs) [Internet]. Atlanta (GA). Available from: https://www.ncbi.nlm.nih.gov/books/NBK587429/.
24. Van Birgelen AP, Smit EA, Kampen IM, Groeneveld CN, Fase KM, Van der Kolk J, et al. Subchronic effects of 2,3,7,8-TCDD or PCBs on thyroid hormone metabolism: use in risk assessment. Eur J Pharmacol. 1995;293(1):77-85, doi:https://doi.org/10.1016/0926-6917(95)90021-7.
25. Boas M, Feldt-Rasmussen U, Main KM. Thyroid effects of endocrine disrupting chemicals. Molecular and Cellular Endocrinology. 2012;355(2):240-8, doi:https://doi.org/10.1016/j.mce.2011.09.005.
26. Nazzari M, Hauser D, van Herwijnen M, Romitti M, Carvalho DJ, Kip AM, Caiment F. CODA: a combo-Seq data analysis workflow. Brief Bioinform. 2022;24(1):1-14, doi:https://doi.org/10.1093/bib/bbac582.
27. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1): 10-2, doi:https://doi.org/10.14806/ej.17.1.200.

28. PerkinElmer Inc. NEXTFLEX® Combo-Seq Analysis Guidelines 2020 [Available from: https://perkinelmer-appliedgenomics.com/wp-content/uploads/2020/06/NOVA-5139-AG_v01_NEXTFLEX-Combo-seq-Analysis-Guideline.pdf.
29. Li B. rsem-prepare-reference documentation page [Available from: https://deweylab.github.io/RSEM/rsem-prepare-reference.html.
30. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323, doi:https://doi.org/10.1186/1471-2105-12-323.
31. Patil AH, Halushka MK. miRge3.0: a comprehensive microRNA and tRF sequencing analysis pipeline. NAR Genom Bioinform. 2021;3(3):lqab068, doi:https://doi.org/10.1093/nargab/lqab068.
32. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
33. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184-91, doi:https://doi.org/10.1038/nprot.2009.97
34. CEFIC C4 team. Omics Data Analysis Framework for Regulatory application (R-ODAF) 2021 [Available from: https://github.com/R-ODAF/Main.
35. Verheijen MC, Meier MJ, Asensio JO, Gant TW, Tong W, Yauk CL, Caiment F. R-ODAF: Omics data analysis framework for regulatory application. Regul Toxicol Pharmacol. 2022;131:105143, doi:https://10.1016/j.yrtph.2022.105143.
36. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014;32(9):896-902, doi:https://doi.org/10.1038/nbt.2931.
37. Conesa A, Nueda MJ. maSigPro: Significant Gene Expression Profile Differences in Time Course Gene Expression Data. R package version 1.70.0. 1.70.0 ed2022.
38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550, doi:https://doi.org/10.1186/s13059-014-0550-8.
39. Conesa A, Nueda MJ. maSigPro User's Guide4 September 2017. Available from: https://www.bioconductor.org/packages/release/bioc/vignettes/maSigPro/inst/doc/maSigProUsersGuide.pdf.
40. Huang HY, Lin YC, Cui S, Huang Y, Tang Y, Xu J, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2022;50(D1):D222-d30, doi:10.1093/nar/gkab1079.
41. Nazzari M, Romitti M, Hauser D, Carvalho DJ, Giselbrecht S, Moroni L, et al. Investigation of the effects of phthalates on in vitro thyroid models with RNA-Seq and ATAC-Seq. Frontiers in Endocrinology. 2023;14, doi:https://doi.org/10.3389/fendo.2023.1200211.
42. Mesnier A, Champion S, Louis L, Sauzet C, May P, Portugal H, et al. The Transcriptional Effects of PCB118 and PCB153 on the Liver, Adipose Tissue, Muscle and Colon of Mice: Highlighting of Glut4 and Lipin1 as Main Target Genes for PCB Induced Metabolic Disorders. PLoS One. 2015;10(6):e0128847, doi:10.1371/journal.pone.0128847.

43. Nebert DW, Roe AL, Dieter MZ, Solis WA, Yang Y, Dalton TP. Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis. Biochemical Pharmacology. 2000;59(1):65-85, doi:https://doi.org/10.1016/S0006-2952(99)00310-X.
44. Huang M, Zhang L, Mesaros C, Zhang S, Blaha MA, Blair IA, Penning TM. Metabolism of a representative oxygenated polycyclic aromatic hydrocarbon (PAH) phenanthrene-9,10-quinone in human hepatoma (HepG2) cells. Chem Res Toxicol. 2014;27(5):852-63, doi:https://doi.org/10.1021/tx500031p.
45. McFarland VA, Clarke JU. Environmental occurrence, abundance, and potential toxicity of polychlorinated biphenyl congeners: considerations for a congener-specific analysis. Environmental Health Perspectives. 1989;81:225-39, doi:https://doi.org/10.1289/ehp.8981225.
46. Laurberg P. Remission of Graves' disease during anti-thyroid drug therapy. Time to reconsider the mechanism? Eur J Endocrinol. 2006;155(6):783-6, doi:https://doi.org/10.1530/eje.1.02295.
47. Dohan O, Portulano C, Basquin C, Reyna-Neyra A, Amzel LM, Carrasco N. The Na+/I symporter (NIS) mediates electroneutral active transport of the environmental pollutant perchlorate. Proc Natl Acad Sci U S A. 2007;104(51):20250-5, doi:https://doi.org/10.1073/pnas.0707207104.
48. ECHA ECA. Substance Infocard - Sodium perchlorate 2023 [updated 06-Sep-2023.
49. Correia de Sousa M, Gjorgjieva M, Dolicka D, Sobolewski C, Foti M. Deciphering miRNAs' Action through miRNA Editing. Int J Mol Sci. 2019;20(24), doi:https://doi.org/10.3390/ijms20246249.
50. Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao XH, Dumitrescu AM, et al. Generation of functional thyroid from embryonic stem cells. Nature. 2012;491(7422):66-71, doi:https://doi.org/10.1038/nature11525.
51. Romitti M, Eski SE, Fonseca BF, Gillotay P, Singh SP, Costagliola S. Single-Cell Trajectory Inference Guided Enhancement of Thyroid Maturation In Vitro Using TGF-Beta Inhibition. Frontiers in Endocrinology. 2021;12, doi:https://doi.org/10.3389/fendo.2021.657195.
52. Koch HM, Angerer J. Chapter 3A. Phthalates: Biomarkers and Human Biomonitoring. Biomarkers and Human Biomonitoring. Issues in Toxicology. Cambridge: Royal Society of Chemistry; 2011. p. 179-233.
53. Yang Z, Guo C, Li Q, Zhong Y, Ma S, Zhou J, et al. Human health risks estimations from polycyclic aromatic hydrocarbons in serum and their hydroxylated metabolites in paired urine samples. Environ Pollut. 2021;290:117975, doi:10.1016/j.envpol.2021.117975.
54. (CDC) CfDCaP. Fourth National Report on Human Exposure to Environmental Chemicals, Updated Tables. January 2019;1, doi:https://doi.org/10.15620/cdc75822.
55. Haines DA, Saravanabhavan G, Werry K, Khoury C. An overview of human biomonitoring of environmental chemicals in the Canadian Health Measures Survey: 2007-2019. Int J Hyg Environ Health. 2017;220(2 Pt A):13-28, doi:https://doi.org/10.1016/j.ijheh.2016.08.002.

56. Berntsen HF, Berg V, Thomsen C, Ropstad E, Zimmer KE. The design of an environmentally relevant mixture of persistent organic pollutants for use in in vivo and in vitro studies. J Toxicol Environ Health A. 2017;80(16-18):1002-16, doi:http://dx.doi.org/10.1080/15287394.2017.1354439.
57. Nueda MJ, Tarazona S, Conesa A. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics. 2014;30(18):2598-602, doi:https://doi.org/10.1093/bioinformatics/btu333.
58. Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs DR, Jr., Lee DH, et al. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. Endocr Rev. 2012;33(3):378-455, doi:https://doi.org/10.1210/er.2011-1050.
59. Phillips JR, Svoboda DL, Tandon A, Patel S, Sedykh A, Mav D, et al. BMDExpress 2: enhanced transcriptomic dose-response analysis workflow. Bioinformatics. 2019;35(10):1780-2, doi:https://doi.org/10.1093/bioinformatics/bty878.
60. Simeckova P, Pencikova K, Kovac O, Slavik J, Parenicova M, Vondracek J, Machala M. In vitro profiling of toxic effects of environmental polycyclic aromatic hydrocarbons on nuclear receptor signaling, disruption of endogenous metabolism and induction of cellular stress. Sci Total Environ. 2022;815:151967, doi:10.1016/j.scitotenv.2021.151967.
61. Tang D, Kang R, Zeh HJ, Lotze MT. The multifunctional protein HMGB1: 50 years of discovery. Nat Rev Immunol. 2023, doi:https://doi.org/10.1038/s41577-023-00894-6.
62. Sun Y, Varambally S, Maher CA, Cao Q, Chockley P, Toubai T, et al. Targeting of microRNA-142-3p in dendritic cells regulates endotoxin-induced mortality. Blood. 2011;117(23):6172-83, doi:https://doi.org/10.1182/blood-2010-12-325647.
63. Peng Y, Song L, Zhao M, Harmelink C, Debenedittis P, Cui X, et al. Critical roles of miRNA-mediated regulation of TGFbeta signalling during mouse cardiogenesis. Cardiovasc Res. 2014;103(2):258-67, doi:https://doi.org/10.1093/cvr/cvu126.
64. Mandolesi G, De Vito F, Musella A, Gentile A, Bullitta S, Fresegna D, et al. miR-142-3p Is a Key Regulator of IL-1beta-Dependent Synaptopathy in Neuroinflammation. J Neurosci. 2017;37(3):546-61, doi:https://doi.org/10.1523/JNEUROSCI.0851-16.2016.
65. Liang L, Fu J, Wang S, Cen H, Zhang L, Mandukhail SR, et al. MiR-142-3p enhances chemosensitivity of breast cancer cells and inhibits autophagy by targeting HMGB1. Acta Pharm Sin B. 2020;10(6):1036-46, doi:https://doi.org/10.1016/j.apsb.2019.11.009.
66. Chen Y, Zhou X, Qiao J, Bao A. MiR-142-3p Overexpression Increases Chemo-Sensitivity of NSCLC by Inhibiting HMGB1-Mediated Autophagy. Cell Physiol Biochem. 2017;41(4):1370-82, doi:https://doi.org/10.1159/000467896.
67. Zhang Y, Liu Y, Xu X. Upregulation of miR-142-3p Improves Drug Sensitivity of Acute Myelogenous Leukemia through Reducing P-Glycoprotein and Repressing Autophagy by Targeting HMGB1. Transl Oncol. 2017;10(3):410-8, doi:https://doi.org/10.1016/j.tranon.2017.03.003.
68. Fu Y, Sun LQ, Huang Y, Quan J, Hu X, Tang D, et al. miR-142-3p Inhibits the Metastasis of Hepatocellular Carcinoma Cells by Regulating HMGB1 Gene Expression. Curr Mol Med. 2018;18(3):135-41, doi:https://doi.org/10.2174/1566524018666180907161124.

69. Jiang D, Wang H, Li Z, Li Z, Chen X, Cai H. MiR-142 inhibits the development of cervical cancer by targeting HMGB1. Oncotarget. 2017;8(3):4001-7, doi:https://doi.org/10.18632/oncotarget.13136.
70. Orsburn BC. Evaluation of the Sensitivity of Proteomics Methods Using the Absolute Copy Number of Proteins in a Single Cell as a Metric. Proteomes. 2021;9(3), doi:https://doi.org/10.3390/proteomes9030034.
71. Schubert OT, Rost HL, Collins BC, Rosenberger G, Aebersold R. Quantitative proteomics: challenges and opportunities in basic and applied research. Nat Protoc. 2017;12(7):1289-94, doi:https://doi.org/10.1038/nprot.2017.040.
72. Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJ, Bunk DM, et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. J Proteome Res. 2010;9(2):761-76, doi:https://doi.org/10.1021/pr9006365.

Chapter 5

Impact of Endocrine Disrupting Chemicals and Sex Hormones on Human ESC-Derived Thyroid Follicles Using Single Cell Transcriptomics

Marta Nazzari, Mírian Romitti, Anna M Kip, Rick Kamps, Sabine Costagliola, Twan van den Beucken, Lorenzo Moroni, Florian Caiment. **Impact of Endocrine Disrupting Chemicals and Sex Hormones on Human ESC-Derived Thyroid Follicles Using Single Cell Transcriptomics**. 2023
(preprint in *Environment International*; under review)

Abstract

Endocrine disruptors are compounds able to interfere with the endocrine system and constitute an important environmental concern. Indeed, detrimental effects on thyroid physiology and functioning have been described. Differences exist in the susceptibility of human sexes to the incidence of thyroid disorders, like autoimmune diseases or cancer. To study how different hormonal environments impact the thyroid response to endocrine disruptors, we exposed human embryonic stem cell-derived thyroid organoids to physiological concentrations of sex hormones resembling the serum levels of human females post-ovulation or males of reproductive age for three days. Afterwards, we added 10 µM benzo[a]pyrene or PCB153 for 24 hours and analyzed the transcriptome changes via single-cell RNA sequencing with differential gene expression and gene ontology analysis. The sex hormones receptors genes *AR*, *ESR1*, *ESR2* and *PGR* were expressed at low levels. Among the thyroid markers, only *TG* resulted downregulated by benzo[a]pyrene or benzo[a]pyrene with the "male" hormones mix. Both hormone mixtures and benzo[a]pyrene alone upregulated ribosomal genes and genes involved in oxidative phosphorylation, while their combination decreased the expression compared to benzo[a]pyrene alone. The "male" mix and benzo[a]pyrene, alone or in combination, upregulated genes involved in lipid transport and metabolism (*APOA1*, *APOC3*, *APOA4*, *FABP1*, *FABP2*, *FABP6*). The combination of "male" hormones and benzo[a]pyrene induced also genes involved in inflammation and NFkB targets. Benzo[a]pyrene upregulated *CYP1A1*, *CYP1B1* and *NQO1* irrespective of the hormonal context. The induction was stronger in the "female" mix. Benzo[a]pyrene alone upregulated genes involved in cell cycle regulation, response to reactive oxygen species and apoptosis. PCB153 had a modest effect in presence of "male" hormones, while we did not observe any changes with the "female" mix. This work shows how single cell transcriptomics can be applied to selectively study the *in vitro* effects of endocrine disrupters and their interaction with different hormonal contexts.

1. Introduction

The hormonal status of the individuals contributes to the development of thyroid diseases in humans. The incidence of thyroid cancer is higher in females than males (1), with the female-to-male ratio peaking at around 15 years of age (when female puberty insets) and decreasing to ~2 at around 50-55 years (2, 3). In a rat model, treatment with estradiol was shown to promote thyroid tumorigenesis (4). Autoimmune diseases of the thyroid are also more frequent in women (5).

Endocrine-disrupting chemicals (EDCs) is a term referring to environmental pollutants that can interfere with the normal functioning of the endocrine system (6). Among the various existing EDCs classes, in this work we focused on the polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyl (PCBs), selecting benzo[a]pyrene and PCB153 as representative for PAHs and PCBs, respectively. For both PAHs and PCBs, the main route of human exposure in the general population is inhalation and ingestion of contaminated food and water (7, 8). PAHs are organic compounds that derive from the incomplete combustion of organic material. They are able to bind the aryl hydrocarbon receptor (AHR), whose activation leads to the induction of cytochrome P450 enzymes, which can metabolize PAHs into toxic and carcinogenic compounds (7). PCBs are synthetic organic compounds whose production was banned in the 1970s due to recognized toxicity. However, owing to their remarkable persistence, they still contaminate the soil and atmosphere. PCBs interfere with the thyroid and thyroid hormone (TH) activity at multiple levels, by impairing the TH synthesis in the thyroid itself and peripheral tissues, decreasing TH transport though the blood and increasing TH clearance (9).

Since EDCs, by definition, are able to disrupt hormonal processes, it is not surprising that sex-specific effects have been observed. Perinatal and developmental exposure to endocrine disrupting chemicals (EDCs) impact differently females and males in *in vivo* experiments analyzing cognitive functions and behavior (10, 11), brain development (12) and intestinal inflammation (13), indicating a system-wide influence of EDCs.

The application of single cell technologies to the field of toxicogenomics and environmental toxicology is still at an early stage, with Liu et al. (2022) (14) reporting how, in the 2017-2022 period of the 2917 articles employing high-throughput sequencing, only 85 concerned single cell RNA sequencing (scRNA-Seq). Some of the areas of focus are the effects of toxic compounds on embryonic development, cell differentiation *in vitro* and *in vivo*, *in vivo* response of tissues composed on heterogeneous cell populations, the composition of complex microorganism populations.

To understand how different sex hormone contexts can affect the response of thyrocytes to EDCs, we preformed single cell RNA sequencing (scRNA-Seq) on human and mouse embryonic stem cell-derived (ESC) thyroid follicles (15-17) after treatment with a mixture of the sex hormones beta-estradiol (E2), 5-alpha-dihydrotestosterone (DHT) and progesterone (PG)

that resemble the serum levels of the female (luteal phase) and male human sexes of reproductive age. We also treated the follicles with the endocrine disruptors BAP and PCB153 in presence of either combination of sex hormones. To gain insight into the effects of the treatments, we performed differential gene expression analysis and gene ontology (GO) analysis.

2. Materials and Methods

2.1 Chemicals Information

The following chemicals were used for the exposure of differentiated thyroid follicles: beta-estradiol (E2) (CAS 50-28-2; E1024-1G, Sigma-Aldrich); 5-alpha-dihydrotestosterone (DHT) solution, 1.0 mg/mL in methanol (CAS 521-18-6; D-073-1ML, Sigma-Aldrich); progesterone (PG), 99%, powder (CAS 57-83-0; P8783-1G, Sigma-Aldrich); PCB153 (CAS 35065-27-1; DRE-C20015300, LGS Standards); benzo[a]pyrene (BAP) (CAS 50-32-8; B-1760, Sigma-Aldrich). All chemicals except DHT were dissolved in 100% DMSO (CAS 67-68-5; 1029521000, Merck Millipore) (stock concentrations: E2 200 mg/mL, 40 mg/mL PG, PCB153 50 mM, BAP 50 mM). The chemicals were aliquoted in single-use vials and stored at -80 °C.

2.2 Mouse Thyroid Follicles Differentiation

Thyroid follicles were differentiated from the A2Lox.Cre_TRE-Nkx2-1/Pax8_Tg-EGFP mouse embryonic stem cell (ESC) line and enriched as previously described (15, 17). For more information see Supplementary Material.

2.3 Human Thyroid Follicles Differentiation

Human thyroid follicles were differentiated from the modified HES3 human embryonic stem cell line following the protocol described by Romitti et al. (2022) (16). Briefly, HES3 cells were infected with lentivirus containing the sequences of the doxycycline (Dox; Sigma-Aldrich)-inducible system, TRE_NXK2.1/PAX8_Ubc_rtTA_Neo. Modified cells were then initially cultured as Embryonic Bodies (EBs) and then embedded in Matrigel. Differentiation into endoderm was induced by Activin A (GFH6, Cell GS) incubation, followed by NKX2.1 and PAX8 overexpression promoted by Dox treatment. Sequential treatment with cAMP, hrTSH, dexamethasone and TGF-beta inhibitors promoted thyroid differentiation and maturation. By day 58, organized and functional thyroid follicles were observed.

2.4 Mouse Follicles Exposure to Hormones and Viability Testing

2.4.1 Exposure to Sex Hormones

Enriched mouse thyroid follicles were extracted from Matrigel as described in the Supplementary Methods and plated in fresh Matrigel droplets at a density of 1,750 follicles in

20 µL Matrigel. After 24 hours, the medium was changed with fresh serum-free mouse differentiation medium (for composition see Supplementary Methods) without hormones (untreated control) or supplemented with "female" (250 pg/mL (0.9 nM) E2, 0.1 ng/mL (0.34 nM) DHT, 10 ng/mL (32 nM) PG) or "male" sex hormones mixtures (20 pg/mL (0.07 nM) E2, 0.6 ng/mL (2 nM) DHT, 0.5 ng/mL (1.6 nM) PG) (n = 2 for each condition). The final [DMSO] was < 0.001%). Cells were cultured for 24, 48, 72 or 96 hours at 5% CO_2, 37 C and 95% humidity. After 48 hours, the media was refreshed.

2.4.2 Viability Testing with FACS

At the end of the incubation period, Matrigel was digested as before, and follicles dissociated into a single cell suspension by incubating the samples with TrypLE Express (12604021, ThermoFisher Scientific) for 10 minutes. Afterwards, samples were washed twice with PBS, resuspended in FACS buffer (PBS, 2% BSA, 10 mM HEPES) with a 1:500 dilution of Draq7 (25191, Biolegend) and filtered through a 30 µm filter for FACS analysis. Samples were run on the BD Accuri™ C6 Plus Flow Cytometer (BD Biosciences) and viable cells were identified as being Draq7 negative (Draq7-). Data analysis was performed on the Floreada web app (v1.0) (https://floreada.io/).

2.5 Human Follicles Exposure to Chemicals

Similarly to mouse follicles, differentiated human ESC-derived thyroid follicles were extracted from Matrigel and enriched as described in the Supplementary Methods. They were plated in fresh Matrigel droplets at a density of 1,500 follicles in 20 µL Matrigel and cultured in serum-free differentiation human medium (for composition see Supplementary Methods) for 3 days. After 3 days, the medium was changed with fresh serum-free human medium, without hormones or with "female" or "male" sex hormones mixtures (at the same concentrations described above) and cells were cultured for 3 additional days. The media were further changed after 24 hours (7 days post follicles reseeding), this time adding also 10 µM BAP or PCB153 or DMSO only as solvent control (final [DMSO] in well = 0.5%) and cells were incubated for 24 hours (Figure 1). The following 8 conditions were tested: Untreated (i.e. no sex hormones added) + Ctrl (DMSO only) (labelled "Untr Ctrl"); Untreated + 10 µM BAP ("Untr BAP"); Female (i.e. "female" hormones added) + Ctrl ("Female Ctrl"); Female + 10 µM BAP ("Female BAP"); Female + 10 µM PCB153 ("Female PCB153"); Male (i.e. "male" hormones added) + Ctrl ("Male Ctrl"); Male + 10 µM BAP ("Male BAP"); Male + 10 µM PCB153 ("Male PCB153").

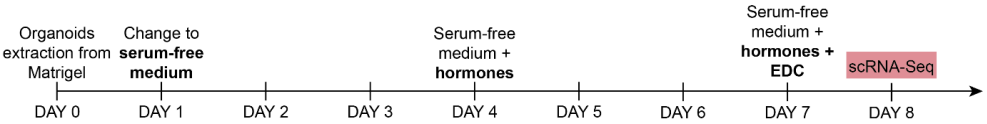


Figure 1. Graphical representations of the exposure regimens of human follicles.

2.6 Single Cell RNA-Seq Libraries Preparation

At the end of the 24 hours incubation period, we performed Matrigel digestion of the human follicles culture and prepared single-cell suspensions as described for the mouse follicles. Cells were manually counted using Trypan blue and resuspended in PBS + 0.04% BSA at a concentration of 1,000 cells/µL to recover ~5,000 cells per sample as indicated by the 10X Genomics guidelines (18). Cells were used to prepare single cell RNA-Seq libraries using the Chromium™ Next GEM Single Cell 3' Kit v3.1 (PN-1000269, 10X Genomics) and Dual Index Kit TT Set A, 96 rxns (PN-1000215, 10X Genomics) following the manufacturer's instructions.

Prepared libraries were quantified on the Qubit 2.0 Fluorometer (ThermoFisher) and quality control performed on a BioAnalyzer 2100 expert (Agilent). Sequencing was performed on a NovaSeq 6000 (Illumina) using a SP Illumina flowcell 100 cycles (v1.5) (Illumina) in paired-end mode. The sequencing data used in this manuscript has been deposited in BioStudies under the accession E-MTAB-13502.

2.7 Single Cell RNA-Seq Data Analysis

Fastq files were generated from Illumina BCL files using the Cell Ranger software (v7.1.0) (10X Genomics) tool mkfastq. We then used Cell Ranger count with the options --include-introns true --expect-cells 5000 to align and quantify reads. We used the Gencode reference genome GRCh38 version 32 (Ensembl 98, 2019-09-05).

Filtered features matrices output by the Cell Ranger software were used for downstream analyses using Seurat (v4.3.0.9012) (19). We filtered the samples using the following parameters: 1,700 < nFeatures < 10,000, nCount > 800 and percent of reads mapping to the mitochondrial genome < 12.5%. Data was log-normalized and scaled using the Seurat function ScaleData using all genes as selected features. We performed principal component analysis (PCA) using the 2,000 most highly variable features and selected the top 15 principal components for K-nearest neighbor (KNN) graph construction. We clustered the cells with the original Louvain algorithm (default) and a resolution of 0.5. Non-linear dimensional reduction was performed with the Uniform Manifold Approximation and Projection (UMAP) algorithm using 16 clusters, determined using the Seurat function FindClusters. To annotate clusters, we used ScType (20) and the full version of the ScType database file available at https://github.com/IanevskiAleksandr/sc-type. We selected "Thyroid" as the tissue type and

modified the cell types to include the ones previously described in Romitti et al. (2022) (16) (Table 1).

Table 1. Gene annotations used for cell type identification.

Cell type	Genes
Immature Thyrocytes	PAX8, FOXE1, NKX2-1, HHEX, TSHR, TG
Mature Thyrocytes	PAX8, FOXE1, NKX2-1, HHEX, TSHR, TG, TPO
Thyroid progenitors	PAX8, FOXE1, NKX2-1, HHEX
Goblet cells	AGR2, AQP3, ATOH1, BACE2, CDON, CDX2, CLCA1, CREB3L1, FCGBP, GALNT12, GUCA2A, ITLN1, KLK1, KRT20, KRT7, LRRC26, MANF, MUC13, MUC2, MUC4, MUC5B, NLRP6, PDIA5, PHGR1, PLA2G10, REP15, SLC9A8, SPDEF, SPINK4, TFF3, TPSG1, ZG16
Embryonic stem cells	ABCG2, ANO6, BCL3, BNIP3, CD24, CD59, CD9, CDH1, CDK8, CTNNB1, DNMT3B, DPPA2, DPPA3, DPPA4, EPCAM, ESRRB, FBXO15, FUT4, FZD1, GAL, GDF3, GJB1, GJB4, GJC1, HES1, HHEX, HMGA2, HOXB5, IL6ST, ITGA4, ITGA6, ITGB1, KCNIP3, KIT, KITLG, KLF4, L1TD1, LEF1, LIFR, LMNA, MYC, NACC1, NANOG, NR6A1, PCGF2, PECAM1, PITX2, PIWIL1, PIWIL2, PIWIL4, PML, PODXL, POU5F1, PRDM5, PROM1, PUM2, SALL4, SLC46A2, SMAD1, SMAD2, SMAD3, SMAD4, SMAD5, SMAD9, SOX15, SOX2, STAT3, SUMO2, TAF8, TDGF1, TEX19, THY1, TRIM28, TRIM6, ZFP42, ZFX, ZIC1
Airway progenitor cells	ABI3BP, AQP3, DAPL1, GSTM2, HPGD, ICAM1, KRT14, KRT15, KRT5, PHLDA3, RPS18, SDC1
Epithelial cells	CD24, CEACAM1, ST6GAL1, ITGB4, IL1R1, PROM1, CDH1, KRT1, KRT7, MUC1, ICAM1, KRT14, KRT5, ITGAL, CD2, KLK3, ITGA5, ITGA4, ITGA2, KRT3, KRT16, SCNN1A, KRT15, ITGA1, KRT2, SCNN1B, SCNN1D, SCNN1G, IFI16, BOK, NKD1, FZD6, DKK3, NRP2, SFRP5, RAI14, DEFB1, KLK1, AGR2, APOA1, GPA33, ANPEP, CRYBA1, BMI1, BRCA1, MUC16, CEACAM5, CTSE, SCGB1A1, EXO1, FOXA1, GABRP, GGT1, SFN, KRT13, LTF, SLC46A2, KLK10, P2RX7, CDKN2A, TP63, CDH3, PSCA, AGER, ZFP42, SPRR1B, SI, TTR, TM4SF20, TSTD1, SYCN, HBEGF, PIGR, MUC13, SELENBP1, ELF3, TSPAN1, GUCA2A, PHGR1, LYPD8, LGALS4, GATA2, SEC23B, TSPAN8, DLX5, DGAT2, ITPR2, THRSP, PLA2G4A, SLC25A48, PGR, FERMT1, EHF, PLEKHS1, CDKL1, MECOM, MSX1, RNF128, ANLN, CKAP2, HMMR, KIF15, CKAP2L, KIF20B, HIRIP3, INCENP, KIF23, PRC1, ECT2, CXCL10, CXCL8, CCL20, CXCL17, PRG4, ALOX15, F5, EMILIN2, SPTSSB, FMO5, IVL, VSIG2, AQP3, PAQR5, EPCAM, CLDN1, OCLN, MUC5AC
Cardiovascular cells	ACTA2, CNN1, CALD1, DES
Fibroblasts	IL1R1, FAP, FLI1, CELA1, LOX, PDGFRB, P4HA1, UCP2, CCR2, ITGAL, FGR, HCK, TNFRSF1B, PRKCD, ENO3, ABI3, TREML4, PIP4K2A, CD300E,

Cell type	Genes
	SERPINB10, CTHRC1, TBX18, COL15A1, GJB2, IL34, EDN3, SLC6A13, VTN, ITIH5, LUM, DPT, POSTN, PENK, MMP14, COL6A2, FABP4, ASPN, ANGPTL2, EFEMP1, SCARA5, IGFBP3, COPZ2, DPEP1, ADAMTS5, COL5A1, CD248, PI16, PAMR1, TNXB, MMP2, COL14A1, CLEC3B, IGFBP6, COL5A2, FBN1, MFAP5, FKBP10, PALLD, WIF1, SNHG18, CDH11, PTCH1, ARAP1, FBLN2, IGF1, PRRX1, FKBP7, OAF, COL6A3, CTSK, DKK1, C1S, RARRES2, GREM1, SPON2, TCF21, PCSK6, COL8A1, ENTPD2, CXCL8, CXCL3, IL6, CYP1B1, COL13A1, ADAMTS10, CCL11, ADAM33, COL4A3, COL4A4, LAMA2, ACKR3, CD55, FBLN7, FIBIN, THBS2, NOV, PTX3, MMP3, LRRK1, HGF, FRZB, COL12A1, COL7A1, MEOX1, PRG4, PKD2, CCL19, NNMT, FOXF1, HAS1, CTGF, ERCC1, WISP1, TWIST2, RIPK3, DDR2, ELN, FN1, HHIP, FMO2, COL1A2, COL3A1, VIM, FSTL1, GSN, SPARC, S100A4, NT5E, COL1A1, MGP, NOX4, THY1, CD40, SERPINH1, CD44, PDGFRA, EN1, DCN, CEBPB, EGR1, FOSB, FOSL2, HIF1A, KLF2, KLF4, KLF6, KLF9, NFAT5, NFATC1, NFKB1, NR4A1, NR4A2, PBX1, RUNX1, STAT3, TCF4, ZEB2, LAMC1, MEDAG, LAMB1, DKK3, TBX20, MDK, GSTM5, NGF, VEGFA, FGF2, P4HTM, CKAP4, INMT, CXCL14

To calculate the percentage of cells expressing a particular gene, we used the R package scCustomize (21) with the default threshold of 0. To identify differentially expressed genes between the clusters, we used the Seurat functions FindAllMarkers or FindMarkers with the default parameters (logfc.threshold = 0.25, test.use = "wilcox", min.pct = 0.1, min.cells.feature = 3, min.cells.group = 3, base = 2) and set an FDR threshold of 0.05.

2.7.1 Gene Ontology Analysis

Gene ontology (GO) analysis was performed with R package clusterProfiler (22) using the GO database version 2.1 (2023-01-01) (23, 24). The genes related to the GO term “Lipid transport” GO:0006869 were downloaded from AmiGO (25).

The code used for the analysis is accessible in a notebook at https://marta-nazzari.github.io/scrnaseq-follicles-hormones/.

3. Results

3.1 Viability Testing of Sex Hormones Combinations

To assess the impact of hormones levels on our *in vitro* thyroid model, we selected two sex hormones combinations that closely reflect the serum levels concentrations of the sex hormones progesterone (PG), beta-estradiol (E2) and dihydrotestosterone (DHT) of the human male and female sexes at reproductive age and post puberty. We will refer to these two combinations as the "male" and "female" condition throughout this manuscript. As the concentrations of PG and E2 in the female fluctuate with the menstrual cycle, we selected concentrations representative of the luteal phase (post ovulation) (26).

Due to the long protocol required to differentiate human follicles and the limited number of follicles that can be generated in a differentiation batch, we tested the cytotoxicity of the selected sex hormones mixtures in a similar model derived from mouse embryonic stem cells (mESC) (15, 17). We reasoned that, since the hormones structures and signaling pathways are conserved across species, we would still be able to observe any sign of cytotoxicity. To this end, we incubated the mouse embryonic stem cell (mESC)-derived thyroid follicles with media supplemented with "male" or "female" sex hormones mixtures for 1, 2, 3 or 4 days and tested the viability via flow cytometry. No cytotoxic effects on the cells were detected at these concentrations, with viability never going below ~ 96% also for thyroid cells, which represent a part of the whole population and express GFP+-thyroglobulin (*Tg*) (Figure 2, Table **2**, Figure 3).

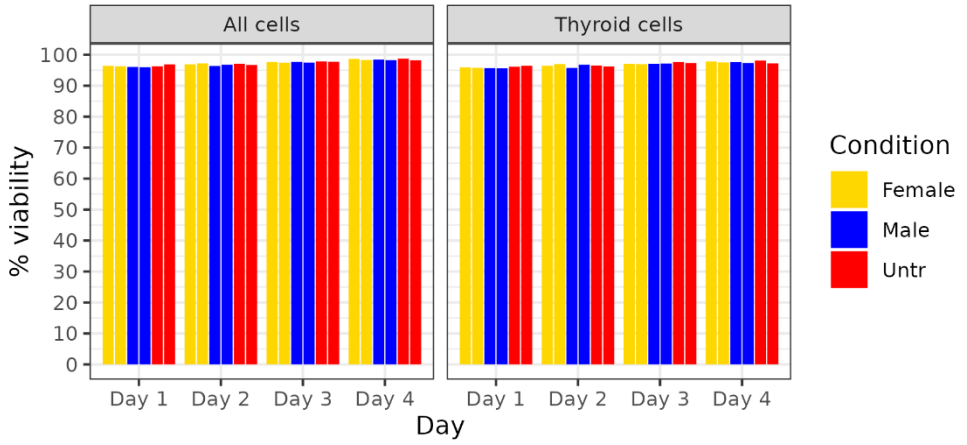


Figure 2. Viability of mouse embryonic stem cell (mESC)-derived thyroid follicles measured by flow cytometry using the viability dye Draq7. The viability corresponds to the percentage of Draq7- cells identified in the population. The thyroid cells are the GFP+ subpopulation. Each condition was performed in duplicate, and each bar represents a sample.

Table 2. Viability of embryonic stem cell (mESC)-derived thyroid follicles measured by flow cytometry using the viability dye Draq7, which stains only dead cells. This table reports the individual values plotted in Figure 2.

	All cells				Thyroid cells			
Sample	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 1**	**Day 2**	**Day 3**	**Day 4**
Untreated 1	96.23	97.07	97.85	98.71	96.11	96.53	97.62	98.09
Untreated 2	96.90	96.67	97.71	98.16	96.46	96.16	97.32	97.19
Female 1	96.44	96.91	97.63	98.64	95.95	96.46	97.05	97.83
Female 2	96.23	97.19	97.44	98.30	95.76	96.99	96.96	97.52
Male 1	96.06	96.38	97.66	98.46	95.65	95.72	97.06	97.61
Male 2	96.00	96.75	97.47	98.21	95.61	96.75	97.14	97.33

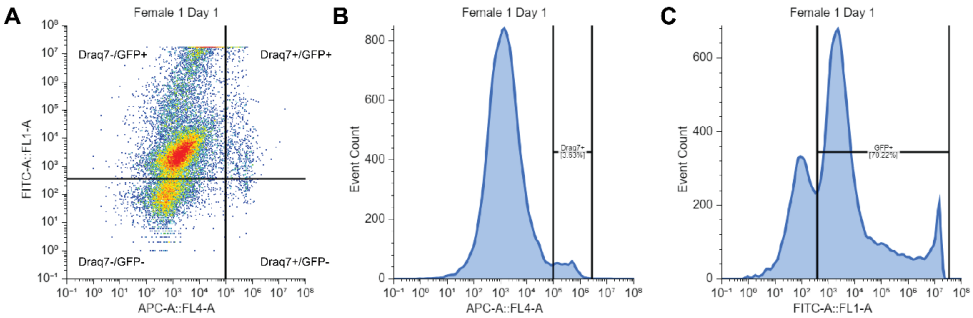


Figure 3. Flow cytometry plots showing the gate strategy for the selection of GFP+ and Draq7+ populations. The sample "Female replicate 1 Day 1" is reported here as representative. (A) Gating strategy for Draq7+ and GFP+ cells. (B) histogram of Draq7+ cells. (C) Histogram of GFP+ cells (includes both Draq7+ and Draq7- cells).

Since we confirmed that the hormones concentrations were not cytotoxic for the mouse cells, we proceeded to perform the exposures with the hormones and endocrine disrupting chemicals (EDCs) on the human model. To this end, we treated the follicles for 3 days with the "female" or "male" hormones mixtures or media only. After this period, we added 10 µM benzo[a]pyrene (BAP) or 10 µM PCB153 for 24 hours and performed single cell (sc) RNA-Seq. No toxicity was observed during this experiment based on trypan blue staining.

3.2 ScRNA-Seq Quality Control and Filtering

The median number of sequenced cells was 3,471 (min = 1,865; max = 5,387), the libraries sizes ranged from 19.8 million bases (Mb) to 27.3 Mb (median = 22.3 Mb) and were of high quality, with a median 70.4% of bases having a quality score (QS) of 30 or more (i.e. a

probability of a sequencing error of 0.001) (min = 70.3%; max = 70.8%). The mean QS was between 30.65 and 30.75 (Figure 4).

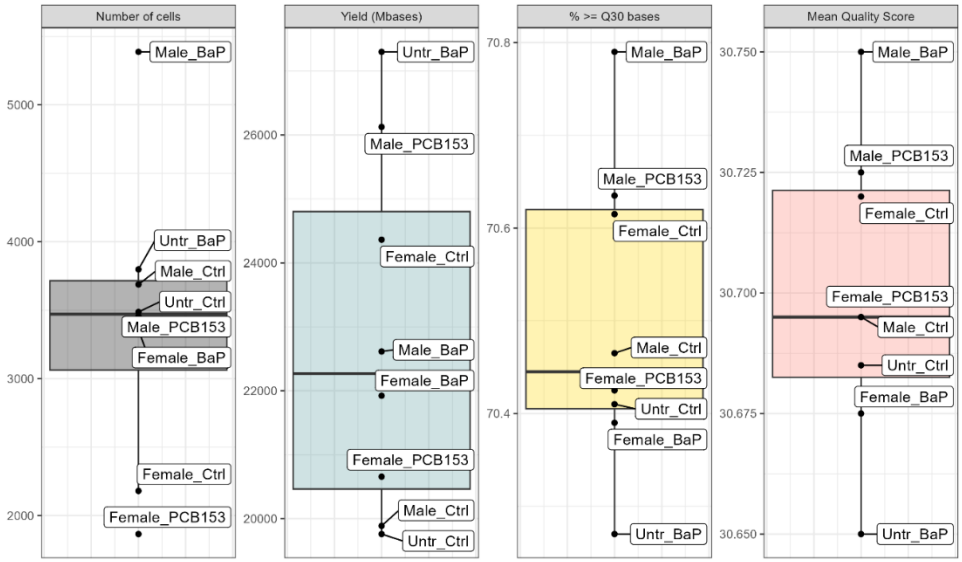


Figure 4. Single cell (sc) RNA-Seq dataset metrics. (A) Number of cells per sample. (B) Libraries sizes. (C) Percentage of sequenced reads with quality score >= 30. (D) Mean quality score. Each dot represents a sample.

We filtered the samples based on the distribution of the number of features (i.e. genes per cell), read count (i.e. mapping reads per cell) and percentage of reads mapping to the mitochondrial genome (Figure 5A) retaining between 1,120 and 3,767 cells (Figure 5B).

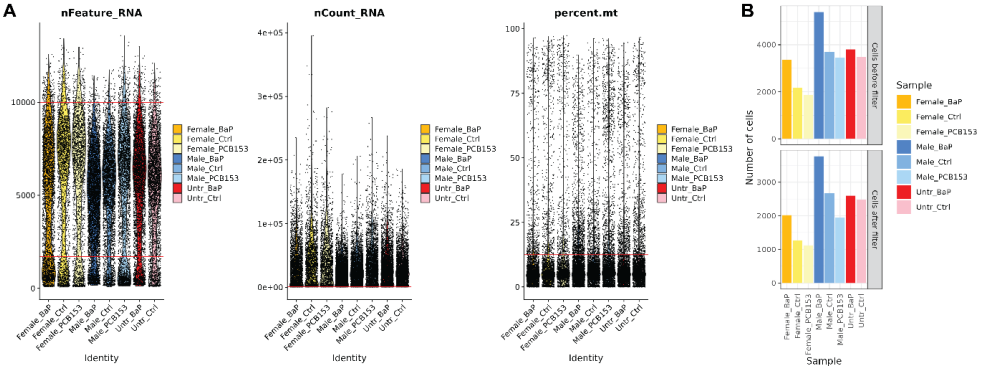


Figure 5. (A) Distribution of nFeature (i.e., number of genes with mapped reads per cell), nCount (i.e., mapping reads per cell) and percent.mt (i.e., percentage of reads mapping to the mitochondrial genome) of the scRNA-Seq samples. The red lines represent the threshold used for filtering (1,700 < nFeature < 10,000;

nCount > 800; percent.mt < 12.5). Each dot is one cell. (B) Number of cells per sample before and after filtering.

3.3 Cluster Annotation and Expression of Thyroid Genes

Cluster annotation identified six cell types: airway progenitor cells, embryonic stem cells, epithelial cells, fibroblasts, goblet cells and mature thyrocytes (Figure 6A). Mature thyrocytes composed between 5.9% (Female PCB153) and 11% (Male BAP) of the whole population (Figure 6B).

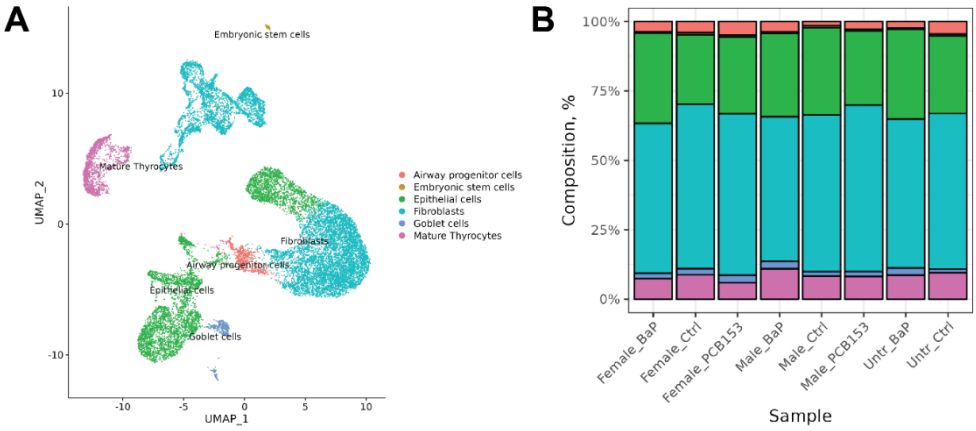


Figure 6. Cell annotation and samples composition. (A) UMAP plot with cells colored for each annotated cell type. (B) Cell composition per sample expressed as percentage of the total of each annotated cell type.

We inspected the expression of the main thyroid markers and other genes important for the synthesis of the thyroid hormone (*DIO1*, *DIO2*, *DUOX1*, *DUOX2*, *DUOXA1*, *DUOXA2*, *FOXE1*, *HHEX*, *IYD*, *NKX2-1*, *PAX8*, *SLC5A5*, *SLC16A2*, *SLC26A7*, *TG*, *TPO*) (Figure 7). Most genes were expressed in the "Mature Thyrocytes" cluster, with some genes being expressed exclusively in this cluster (*DIO1*, *DIO2*, *SLC26A7*, *FOXE1*, *IYD*, *PAX8*, *TG*, *TPO*), while others were expressed by subpopulations of other clusters as well (*DIO1*, *DUOX2*, *DUOXA1*, *DUOXA2*, *SLC16A2*, *HHEX*, *NKX2-1*). Interestingly, *SLC5A5* (coding for the Sodium-Iodide Symporter NIS) did not seem to be expressed in our samples.

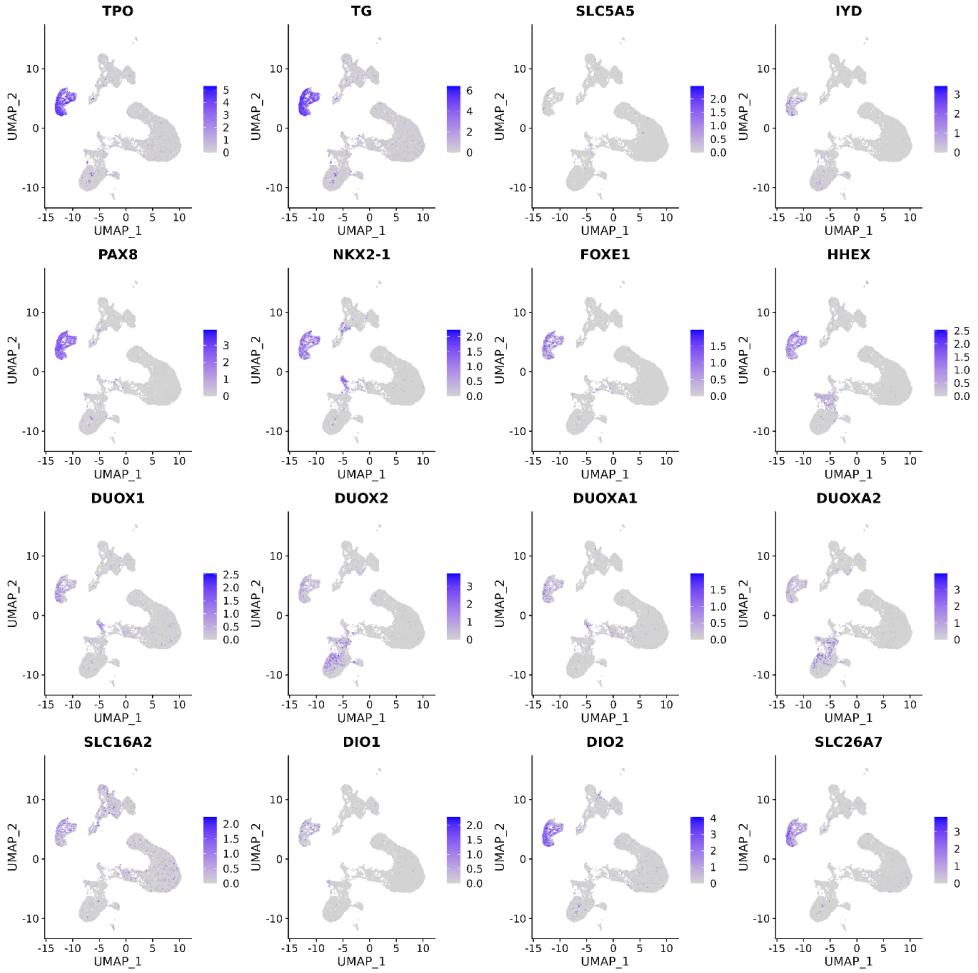


Figure 7. UMAP plots showing the expression of thyroid marker genes and genes important for the synthesis of the thyroid hormone. Each plot includes the cells from all samples.

3.4 Expression of Sex Hormones Receptors

Since the sex hormones we selected exert their effects when binding their receptors, we investigated the expression of the corresponding genes *AR*, *ESR1*, *ESR2* and *PGR*. All these genes showed a very low expression across all cell types (Figure 8A-C), including the "Mature Thyrocytes" cluster (Figure 8D). The percentage of cells expressing the genes in the "Mature Thyrocytes" cluster was for 12.5% *AR*, 1.9% for *ESR1*, 20.4% for *ESR2* and 0.3% for *PGR* (Figure 8E), with the expression levels and percentages of expressing cells being comparable across samples (Figure 8F).

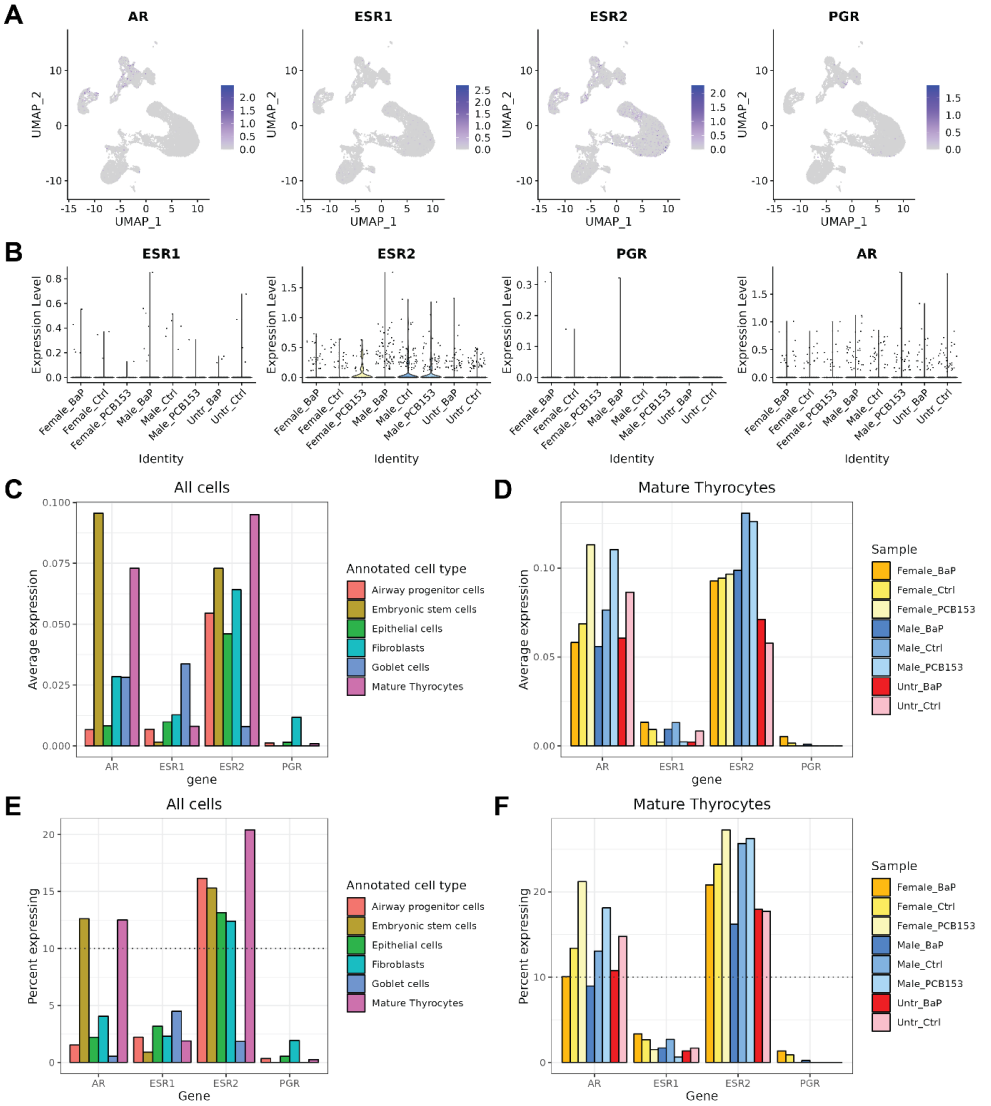


Figure 8. Expression of sex hormones receptors *AR*, *ESR1*, *ESR2* and *PGR*. (A) UMAP plots showing the expression in the samples. Each plot includes the cells from all samples. (B) Violin plot showing the level of expression in the "Mature Thyrocytes" annotated cluster divided by sample. (C) Average expression in the different annotated cell types. The values are calculated including all eight samples. (D) Average expression in the "Mature Thyrocytes" cluster. (E) Percent of expressing cells divided by annotated cell type. The values are calculated including all eight samples. (F) Percent of "Mature Thyrocytes" expressing cells divided by sample.

3.5 Differential Expression Analysis

In this study, we decided to focus our analysis on the responses of thyrocytes cells, therefore we performed differential expression analysis on the "Mature Thyrocytes" cluster. We analyzed several pairwise comparisons, reported in Table 3, which allowed us to investigate different aspects of the effects caused by hormones and/or EDC treatment. The number of differentially expressed genes (DEGs, FDR < 0.05) was between 0 (response to PCB153 in presence of female hormones) and 101 (response to BAP in absence of hormones mixtures) (Figure 9). The complete list of genes that passed the filtering criteria for differential expression analysis and of DEGs is available in the Supplementary Material.

Table 3. Pairwise comparisons included in the differential expression analysis. Since we focused on the thyrocytes response, we performed the analysis on the "Mature Thyrocytes" cluster only. The information we wanted to obtain from the comparison and the number of resulting differentially expressed genes (DEGs) are reported. (Untr = no hormones added, Ctrl = DMSO solvent control.)

Comparison	Aim	Number of DEGs
Male Ctrl vs Untr Ctrl	Response to "male" hormones	30
Female Ctrl vs Untr Ctrl	Response to "female" hormones	15
Male BAP vs Untr BAP	Influence of "male" hormones on response to BAP	19
Female BAP vs Untr BAP	Influence of "female" hormones on response to BAP	25
Untr BAP vs Untr Ctrl	Response to BAP only	101
Male BAP vs Male Ctrl	Response to BAP in environment with "male" hormones	66
Female BAP vs Female Ctrl	Response to BAP in environment with "female" hormones	21
Male PCB153 vs Male Ctrl	Response to PCB153 in environment with "male" hormones	6
Female PCB153 vs Female Ctrl	Response to PCB153 in environment with "female" hormones	0

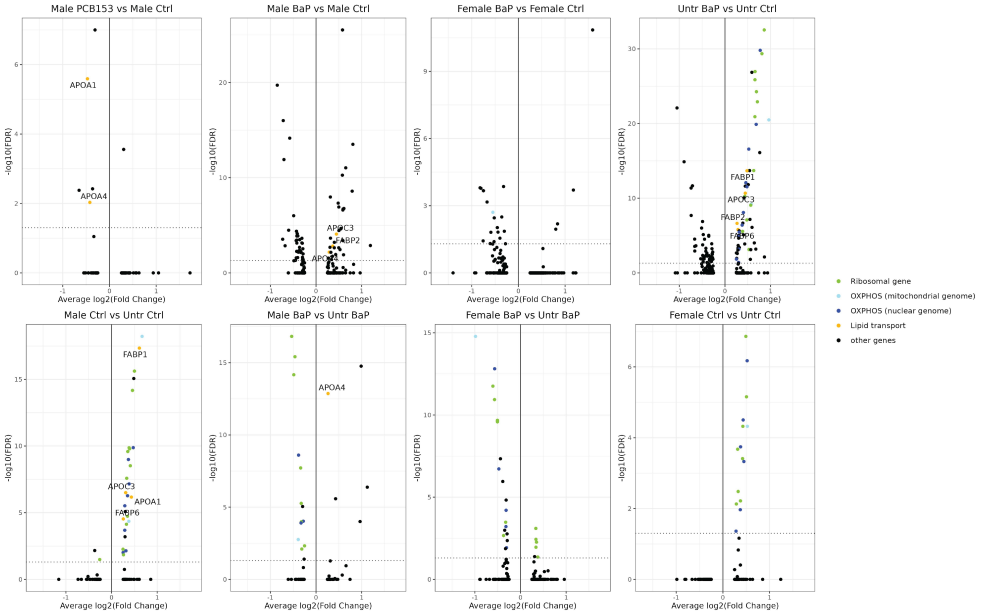


Figure 9. Volcano plots of the genes in each comparison reported in Table 1. The name of each comparison is reported on top of each plot. For every comparison, the genes that pass the criteria were included in the analyses. Every dot represents a gene. The False Discovery Rate (FDR) threshold of 0.05 is indicated as a dotted line on the y axis. Ribosomal genes are indicated in green. Genes belonging to the OXPHOS pathway coded by the mitochondrial genome and nuclear genome are colored in light blue and blue, respectively, and were retrieved from the KEGG pathway hsa00190 "Oxidative phosphorylation" (KEGG release 107). Genes belonging to the GO term "Lipid transport" (GO:0006869) are colored yellow and the gene name is reported. The other genes, which do not belong to any of the previous three categories, are labelled "other genes".

3.5.1 Effects on Thyroid Markers

We then investigated how the level of expression of thyroid markers and other genes important for thyroid hormone synthesis were affected by each treatment.

Only *TG* expression was affected, by BAP alone and in the context of "male" sex hormones (Figure 10), although it seemed to also be downregulated in with "female" hormones (but did not reach statistical significance) (Figure 11).

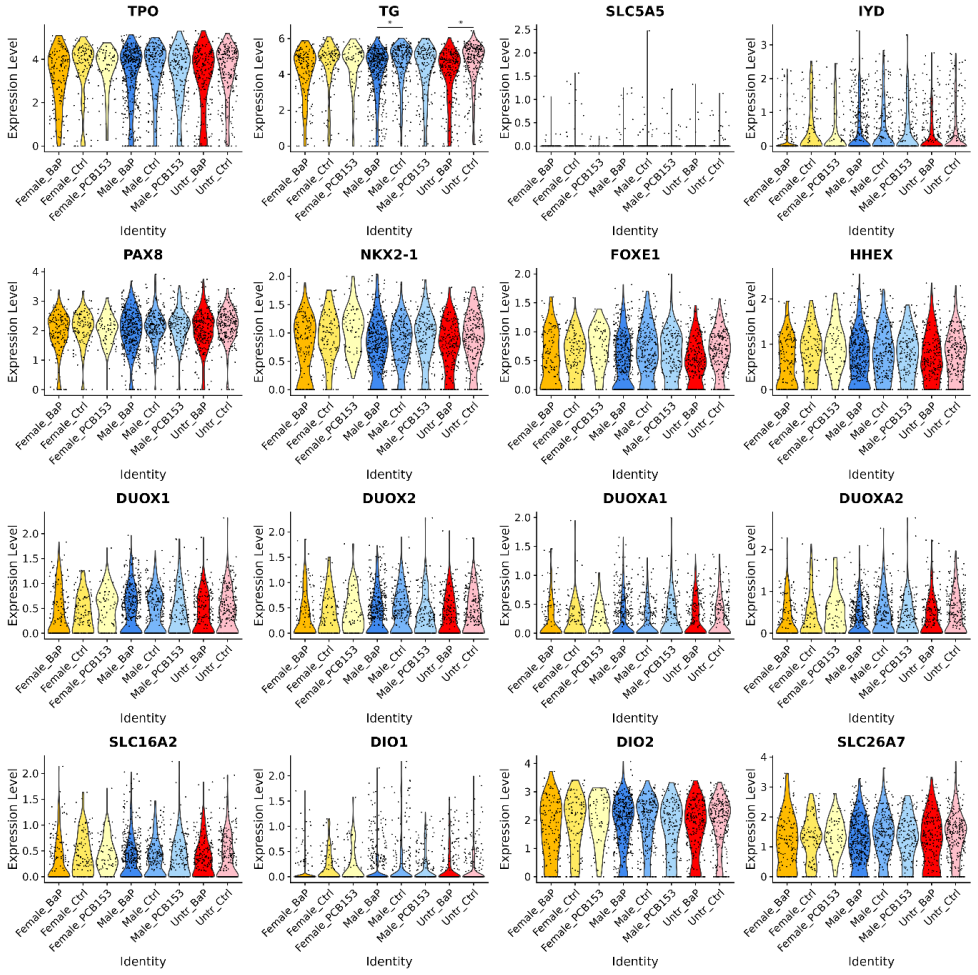


Figure 10. Expression of thyroid markers and genes important for the synthesis of the thyroid hormone. Violin plots reporting the expression levels in the "Mature Thyrocytes" cluster divided by sample. A horizontal bar with an asterisk on top of the violin indicates that the gene is differentially expressed in that comparison (FDR = 0.05, Wilcoxon Rank Sum test).

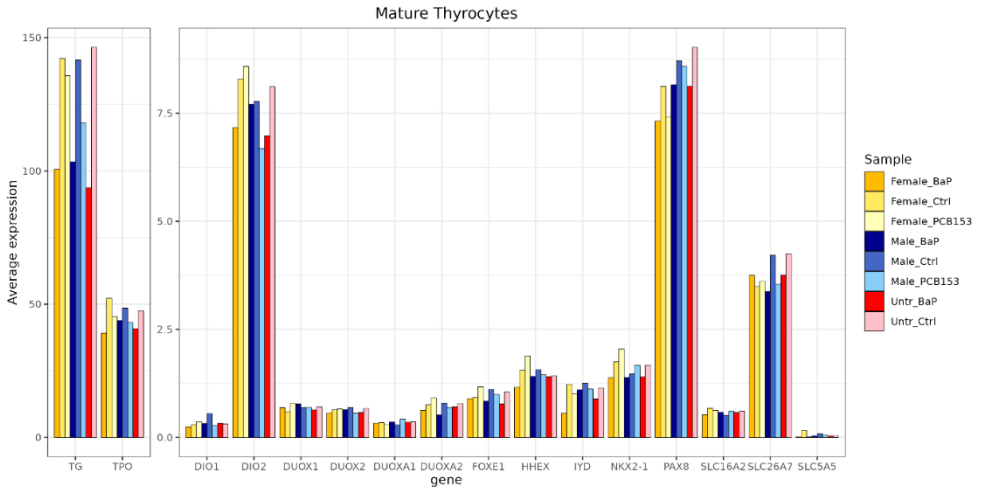


Figure 11. Average expression of thyroid markers and genes important for the synthesis of the thyroid hormone in the "Mature Thyrocytes" cluster divided by sample.

3.5.2 Effects of Treatment with Sex Hormones

To gain insight on the effect of the treatment of the thyrocytes with physiological concentrations of sex hormones, we focused on the samples exposed to hormones only ("Female Ctrl vs Untr Ctrl" and "Male Ctrl vs Untr Ctrl" comparisons).

Treatment with both mixtures of sex hormones induced an upregulation of ribosomal genes and genes coding for elements of the oxidative phosphorylation (OXPHOS) pathway. Additionally, "male" sex hormones upregulated genes coding for proteins involved in lipid transport and metabolism (*APOA1*, *APOC3*, *FABP1*, *FABP6*) (Figure 12).

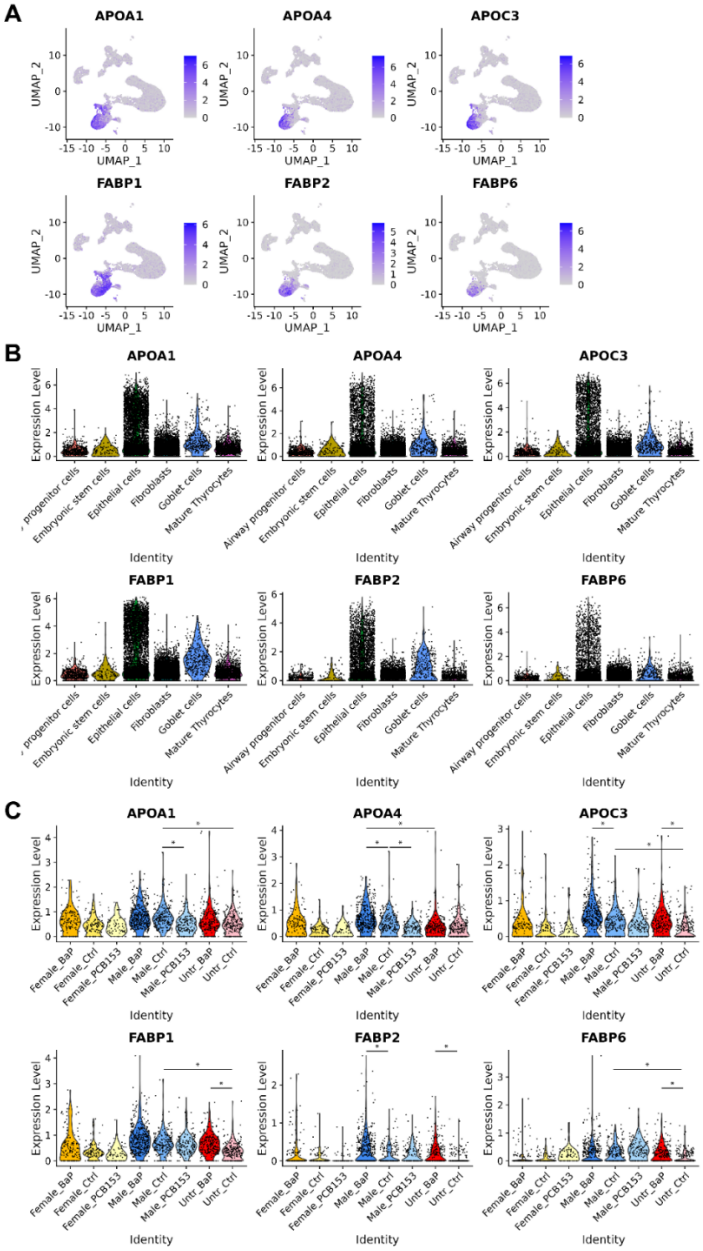


Figure 12. Expression of genes belonging to the GO term "Lipid transport (GO:0006869) that are differentially expressed in at least one of the comparisons in Table 1. (A) UMAP plots showing gene expression in the different cell populations. Each plot includes the cells from all samples. (B) Violin plot showing the level of gene expression in the various annotated cell types. (C) Violin plot showing the level of gene expression in the mature thyrocytes subpopulation. A horizontal bar with an asterisk on top of the violin indicates that the gene is differentially expressed in that comparison (FDR = 0.05, Wilcoxon Rank Sum test).

Enrichment analysis of DEGs on Gene Ontology Biological Process (GO BP) showed enrichment (q-value < 0.01) for 21 terms, related to the three ontologies of OXPHOS, ribosome assembly and translation and lipid transport (Figure 13A-B). Interestingly, OXPHOS and ribosome assembly and translation were also enriched in response to "female" hormones (Figure 13C-D).

Overall, treatment with mixtures of sex hormones upregulated the expression of ribosomal genes and genes involved in the OXPHOS pathway. Additionally, the "male" sex hormones mixture upregulated genes involved in lipid transport.

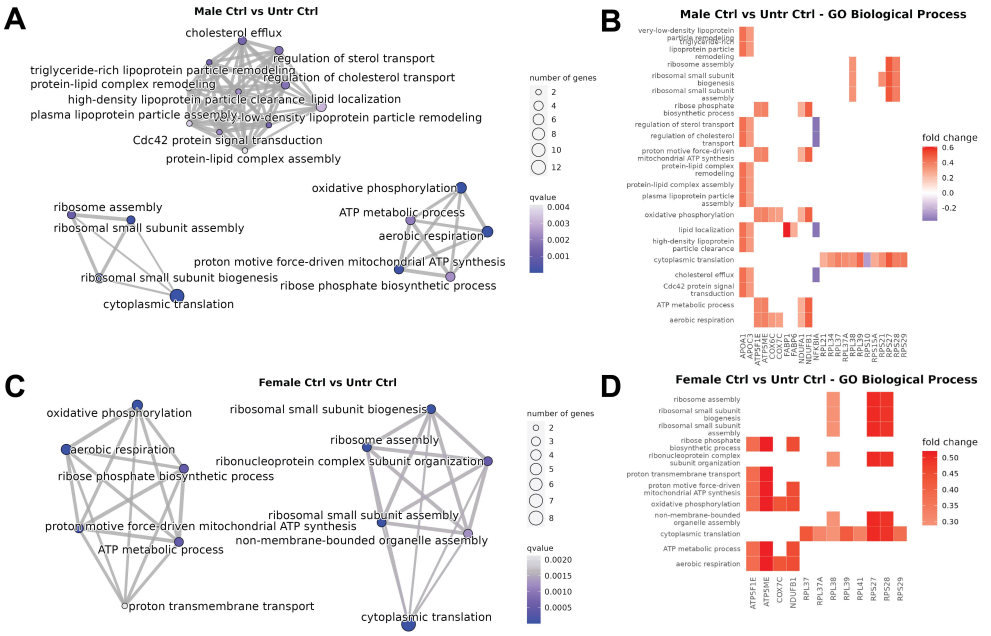


Figure 13. Results of Gene Ontology enrichment analysis. (A, C) Emap plots and (B, D) heatplots of the results of the Gene Ontology – Biological Process (GO BP) enrichment analysis (q-value < 0.01) on the differentially expressed genes (DEGs) in the (A, B) "Male Ctrl vs Untr Ctrl" and (C, D) "Female Ctrl vs Untr Ctrl" comparisons. For emap plots, the labels are GO BP terms, the size of each circle is proportional to the number of DEGs that belong to the GO BP term and a darker color corresponds to a lower q-value. The heatplots report the same pathways as the emap plots but display the DEGs involved and their fold change with respect to the control.

3.5.3 Effects on AHR Target Genes

As BAP is an agonist of the aryl hydrocarbon receptor (AHR), we investigated whether the presence of sex hormones would affect the expression of AHR target genes in response to BAP.

In the mature thyrocytes population, *AHR* and *AHRR* showed a low expression compared to the other annotated cell types (Figure 14A), with a mean 37.8% and 19.3% of mature thyrocytes expressing *AHR* and *AHRR*, respectively (Figure 14B). *CYP1A1* and *CYP1B1* were expressed by only a thyrocytes subpopulation and were clearly induced by BAP treatment, while *TIPARP*, *NQO1* and *ALDH3A2* had a more homogeneous expression. *ALDH3A1*, *ALDH3B1*, *ALDH3B2* and *CYP1A2* were expressed by very few mature thyrocytes.

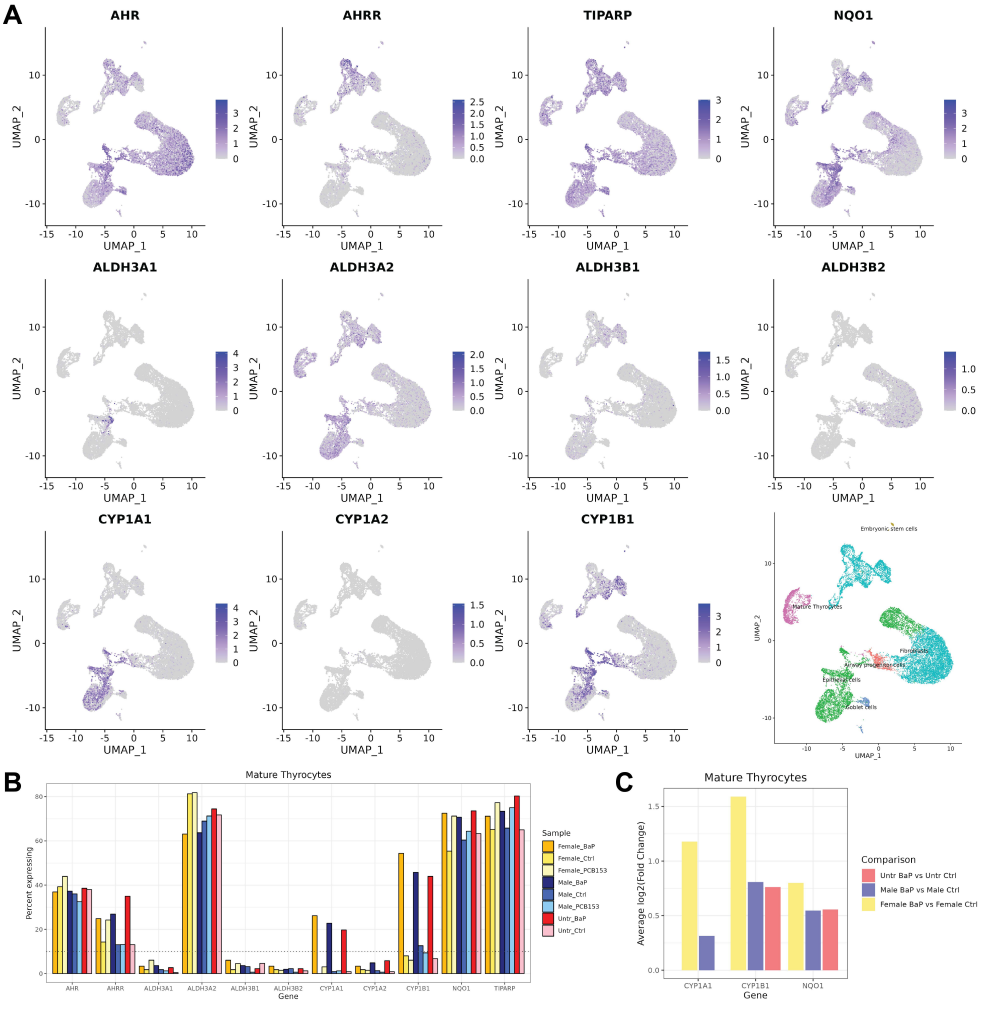


Figure 14. Expression of *AHR* and some of the genes induced by AHR activation upon BAP binding. (A) UMAP plots showing the expression in the different cell populations. Each plot includes the cells from all samples. The cell type annotation is also reported for easier interpretation. (B) Percent of mature thyrocytes per sample expressing the genes. The threshold of 10% of minimum fraction of expressing cells used for differential expression analysis is reported as a dotted line. (C) Barplot of average log2(fold change) of

CYP1A1, *CYP1B1* and *NQO1* in the "Untr BAP vs Untr Ctrl", "Male BAP vs Male Ctrl" and "Female BAP vs Female Ctrl" conditions in the mature thyrocytes population.

CYP1B1 and *NQO1* were upregulated by BAP in presence or absence of sex hormones (Figure 15). The level of induction was higher in the "female" sex hormones condition compared to the "male" sex hormones or BAP only conditions. *CYP1A1* differential expression passed the set filters only in the presence of "female" or "male" sex hormones, and the level of induction was stronger in the presence of "female" hormones (Figure 14C, Figure 15). *CYP1A2* appeared induced but was excluded by the filters applied to the differential expression analysis. We did not detect any GST among the DEGs.

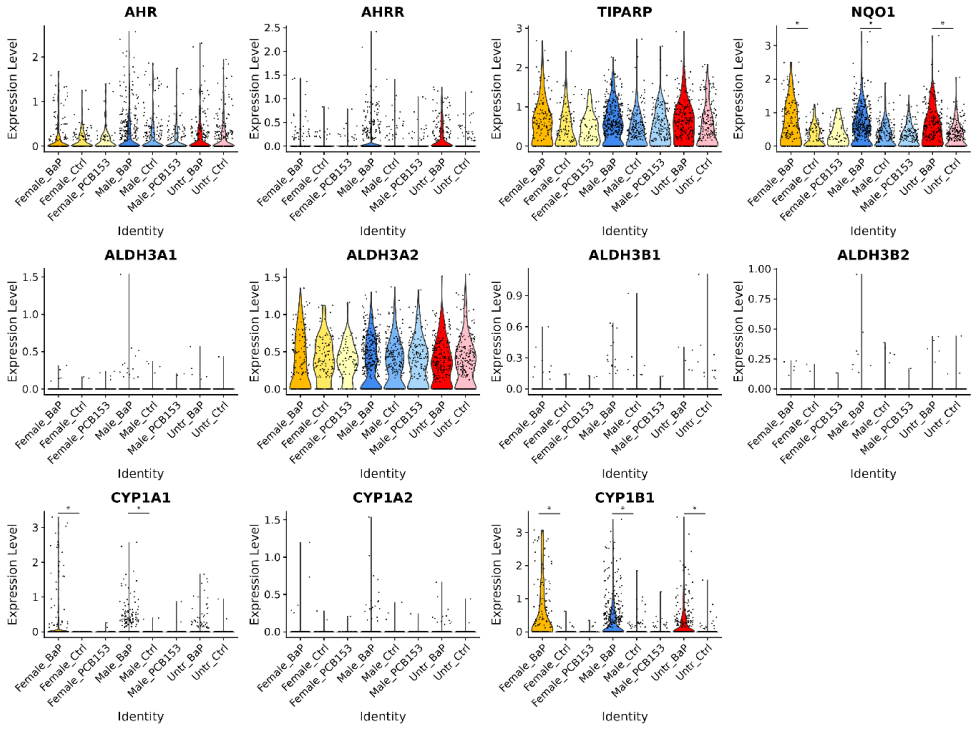


Figure 15. Violin plot showing the level of expression of *AHR* and some of the genes induced by AHR activation upon BAP binding in the "Mature Thyrocytes" annotated cluster divided by sample. A horizontal bar with an asterisk on top of the violin indicates that the gene is differentially expressed in that comparison (FDR = 0.05, Wilcoxon Rank Sum test).

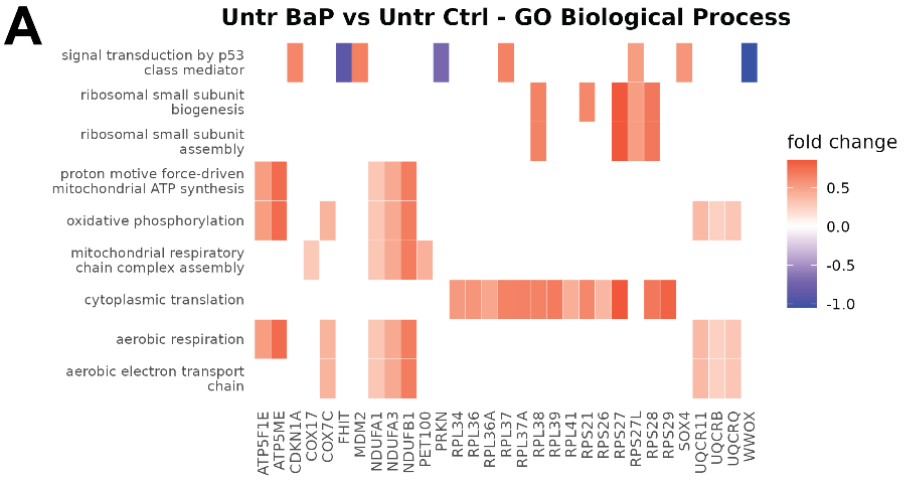
3.5.4 Other Effects of BAP Treatment

3.5.4.1 Effects of BAP Alone

To understand the effects of thyrocytes to BAP treatment, we focused on the response of the sample treated with BAP only ("Untr BAP vs Untr Ctrl" comparison).

In total, we identified 101 DEGs (43 downregulated, 58 upregulated). Similar to our analysis on the sex hormones effect, we observed an upregulation of ribosomal and OXPHOS genes, as well as the upregulation of genes involved in lipid transport and metabolism (*APOC3*, *FABP1*, *FABP2*, *FABP6*) (Figure 12C). *TTR* was also upregulated. To gain further insight into the response to BAP treatment, we performed GO BP enrichment analysis. As expected, we observed enrichment for terms relating to oxidative phosphorylation, ribosomal assembly and translation, but also for the term "Signal transduction by p53 class mediator" (GO:0072331) (Figure 16A). The DEGs involved in this pathway were *CDKN1A*, *MDM2*, *RPL37*, *RPS27L* and *SOX4* (upregulated), *FHIT*, *PRKN* and *WWOX* (downregulated). Additionally, we observed dysregulation of genes involved in the response to oxidative stress: *NFE2L2* and *ROMO1* (upregulated) and *MSRA* (downregulated).

GO BP enrichment analysis of only the upregulated genes showed enrichment for terms related to oxidative phosphorylation, ribosomal assembly and translation, DNA damage response and signaling at the G1 cell cycle phase and response to reactive oxygen species (ROS) (Figure 16B). GO BP analysis for only the downregulated genes did not result in any enrichment.



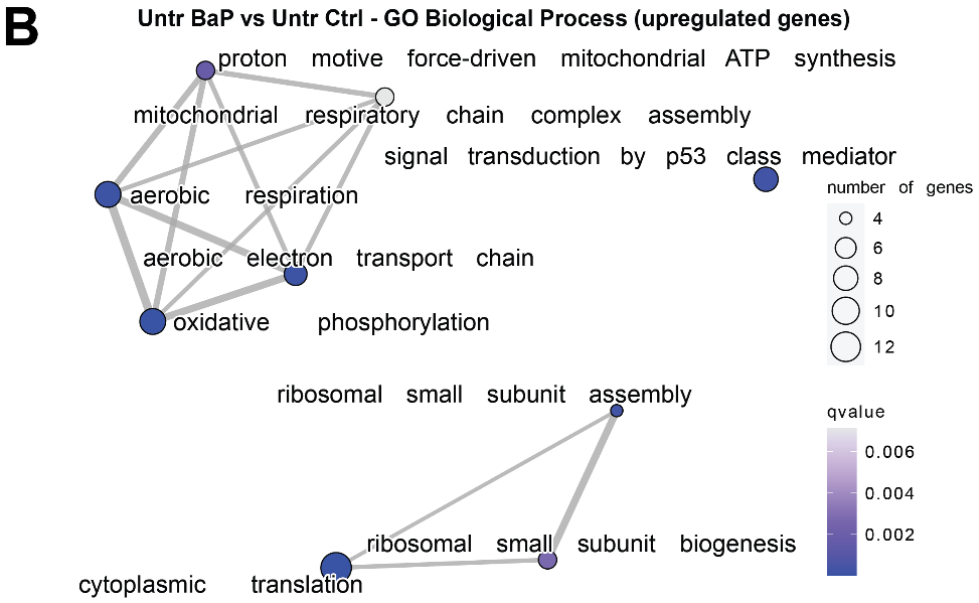


Figure 16. Results of Gene Ontology enrichment analysis. (A) Heatplot of the results of the enrichment analysis for the Gene Ontology – Biological Pathway (GO – BP) database in the "Untr BAP vs Untr Ctrl" comparison (q-value < 0.01). (B) Emap plot of the results of the enrichment analysis for the GO – BP database on the upregulated DEGs in the "Untr BAP vs Untr Ctrl" comparison (q-value < 0.01).

3.5.4.2 Influence of Sex Hormones Treatment to BAP Response

We then investigated how the treatment with sex hormones affected the response to BAP ("Male BAP vs Untr BAP" and "Female BAP vs Untr BAP" comparisons).

In both comparisons, we observed a downregulation of ribosomal and OXPHOS genes compared to the sample treated with BAP only (Figure 9). Interestingly, the response to BAP in a "male" sex hormones environment upregulated genes with a positive role in inflammation (*SA100A9*, *SA100A8*, *SAA1* and *SLPI*), while the presence of "female" sex hormones dysregulated genes involved in transcription (*POL2RL*) and translation (*NACA*, *SEC62*). Of note, the percentage of reads mapping to ribosomal genes did not differ among samples (Kruskal-Wallis test with Dunn post-hoc test) (Figure 17, Table 4).

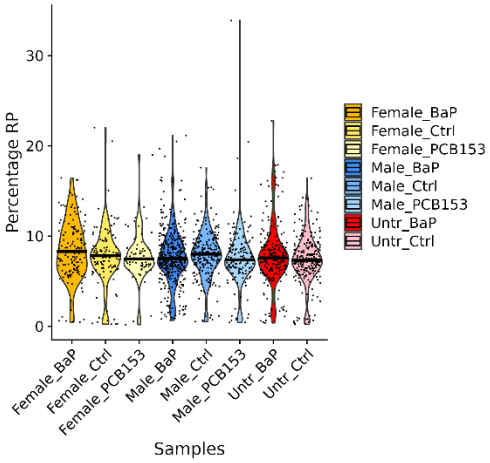


Figure 17. Violin plot showing the distribution of the percentage of reads per cell mapping to ribosomal genes in the "Mature Thyrocytes" population across the samples. The median per sample is reported as a bar.

Table 4. Fraction of reads mapping to ribosomal genes per cell in the "Mature Thyrocytes" population across samples. The minimum, median and maximum values are reported.

Sample	Min	Median	Max
Untr Ctrl	0.22	7.33	16.42
Untr BAP	0.38	7.57	22.78
Male PCB153	0.43	7.37	33.92
Male Ctrl	0.52	8.02	17.59
Male BAP	0.68	7.52	21.15
Female PCB153	0.19	7.47	19.00
Female Ctrl	0.25	7.84	22.02
Female BAP	0.46	8.31	16.45

3.5.4.3 Differences in Response to BAP in Presence or Absence of Sex Hormones

To see how the presence of sex hormones in the media would affect the response to BAP, we focused on the differences in response to BAP in presence or absence of sex hormones (comparisons "Male BAP vs Male Ctrl", "Female BAP vs Female Ctrl" and "Untr BAP vs Untr Ctrl").

In samples treated with both sex hormones and BAP we did not observe the upregulation of ribosomal and OXPHOS genes that we had in the sample treated with BAP alone (Figure 9).

We observed 31 and 5 unique DEGs in the samples treated with BAP in presence of "male" or "female" hormones, respectively (Figure 18A). GO BP analysis on the 31 unique DEGs for the sample treated with "male" hormones showed enrichment for terms related to the regulation of inflammation, mostly involving the upregulated genes *NFKB1*, *NFKBIA*, *S100A9* and *SAA1* (Figure 18B), which are NFkB target genes. Additionally, other upregulated DEGs that are also NFkB targets were *CXCL2*, *CXCL3*, *IER2*, *IER3*, *IL1R1*, *IL32*, *MAPK6* and *NFKBIZ*. The GO BP analysis on the genes dysregulated by BAP + "female" hormones did not show any enrichment.

In summary, treatment with BAP alone affected the expression of pro- and anti-apoptotic genes, genes involved in cell cycle regulation and response to ROS, as expected being BAP a carcinogenic compound. Addition of sex hormones to BAP treatment decreased the expression of some ribosomal and OXPHOS genes compared to BAP alone, increased the expression of inflammation genes ("male" hormones) or genes involved in transcription/translation ("female" hormones).

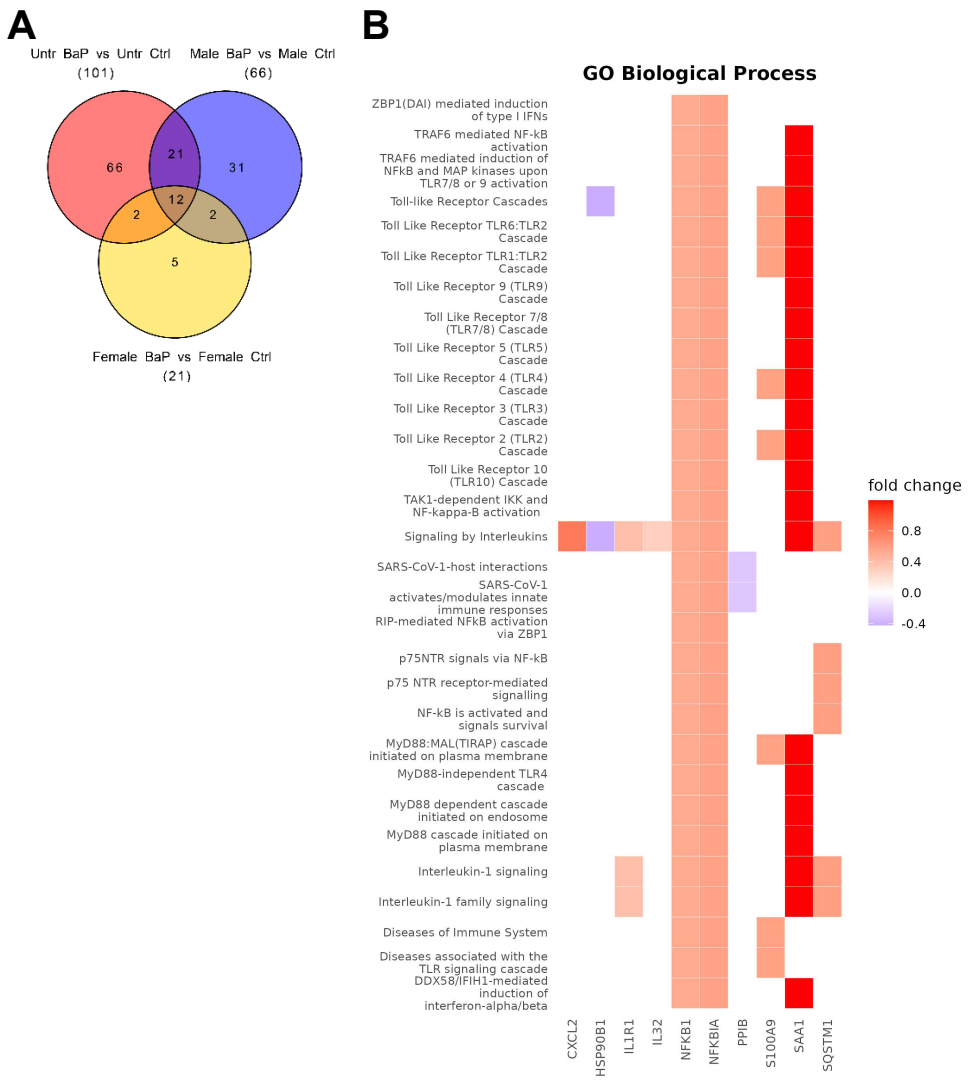


Figure 18. Comparison of differentially expressed genes (DEGs) in samples treated with BAP compared to the control in presence or absence of sex hormones. (A) Venn diagram showing the DEGs overlap among the conditions "Male BAP vs Male Ctrl", "Female BAP vs Female Ctrl" and "Untr BAP vs Untr Ctrl". The total number of DEGs is reported in parenthesis below the comparison name. (B) Gene Ontology – Biological Pathway analysis on the 31 unique DEGs in the "Male BAP vs Male Ctrl" comparison.

3.5.5 Effects of PCB153 Treatment

Treatment with "female" hormones and PCB153 did not result in any gene being differentially expressed, while treatment with "male" hormones and PCB153 yielded 6 DEGs (downregulated: *AFP*, *APOA1*, *APOA4*, *TTR*, *S100A9*; upregulated: *BAX*).

4. Discussion

To investigate whether different hormonal environments can change the response to endocrine disrupting chemicals (EDCs) in the thyroid, we exposed human embryonic stem cell (hESC)-derived thyroid follicles to benzo[a]pyrene (BAP) or the polychlorinated biphenyl PCB153 in presence of a sex hormones mixture of estrogen (E2), dihydrotestosterone (DHT) and progesterone (PG) that would resemble that of the human females and males of reproductive age. Since the levels of sex hormones in the female fluctuate with the menstrual cycle, we selected the concentrations observed during the luteal phase (between ovulation and menstruation) (26). We treated the follicles for 3 days with hormones only to "condition" our cells to the new hormonal environment, and then also to the selected EDCs for 24 hours. We then analyzed the changes in transcriptome with single cell RNA sequencing (scRNA-Seq) to focus on the response of the thyroid population, since, due to the ESC-derived nature of our model, other cell types are also present and have not been characterized (16), we did not focus on them for this analysis.

The expression of the sex hormones receptors *AR*, *ESR1*, *ESR2* and *PGR* in the mature thyrocytes population were very low. It must be noted that we did not measure the level of protein expression, and the levels of transcript and protein do not always correlate (27). According to the Human Protein Atlas consensus dataset, all these receptors are expressed in the thyroid, albeit at low levels (28). Survey of the literature on the estrogen receptors (ERs) expression in the thyroid shows conflicting results. Several reports exist of positive ERs expression in the thyroid (29, 30). Other works showed a higher expression of *ESR1* than *ESR2* (31) or only expression of *ESR2* or its protein product ER-beta (32, 33). The discrepancies could arise from methodological differences, such as different tissue preparations, antibodies used for immunohistochemistry or criteria for data evaluation (34). AR (35, 36) and PGR (30, 37-39) gene and protein expression in the normal human thyroid has also been described.

When we treated the follicles with sex hormones only or with BAP, we observed an increase in expression of ribosomal genes and gene coding for elements of the mitochondrial respiratory chain. This could indicate an increase in the overall requirement for cellular energy demand and rate of transcription or a compensatory mechanism due to compromised mitochondrial functioning (40). The combination of BAP and sex hormones instead downregulated the expression of these genes compared to BAP alone.

We also observed that the treatment with BAP and/or "male" sex hormones upregulated some genes whose products are involved in lipid transport and metabolism. These genes were *APOA1* and *APOA4* (upregulated by "male" hormones and by BAP in presence of "male" hormones), *APOC3*, *FABP1* and *FABP6* (upregulated by "male" hormones or BAP) and *FABP2* (upregulated by BAP). It must be noted that the highest expression of these genes was observed in a subpopulation of epithelial cells. APOA1 and APOC3 are components of chylomicrons, High Density Lipoproteins and Very Low Density Lipoproteins (APOC3 only), APOA4 of

chylomicrons, and are reported to be expressed in the liver and small intestine (41, 42) and play a role in cholesterol and fatty acid transport. *FABP1* is expressed at high levels in the liver, while *FABP2* and *FABP6* in the small intestine. FABP1 transports long-chain fatty acids (LCFA) linked to beta-oxidation and can interact with PPAR-alpha nuclear receptors. FABP2 seems to be involved in LCFA partitioning into triglycerides (rather than phospholipids). FABP6 has a role in chylomicron biogenesis (43). We identified these genes as being part of the Gene Ontology (GO) term "Lipid transport", which comprises 349 proteins at the date of our analysis. It is interesting how the expression of this specific group of genes consistently experiences alterations, especially given that no single control was employed in the comparative analyses where we detected differential gene expression, thereby eliminating the potential for sample bias. To our knowledge, the dysregulation of this set of genes by sex hormones and/or BAP in an *in vitro* thyroid model has not been described before.

Interestingly, in the samples treated with male hormones only and with PCB153 and male hormones compared to their control (comparisons "Male Ctrl vs Untr Ctrl" and "Male PCB153 vs Male Ctrl"), we observed a concomitant upregulation and downregulation of *TTR* and *APOA1*, respectively (*TTR* was also upregulated in the sample treated only with BAP compared to control but *APOA1* was not). TTR has been described to associate with APOA1 in the plasma and is thus present within circulating High Density Lipoproteins. Both are amyloidogenic proteins and their function is connected: TTR is a non-canonical protease able to cleave APOA1 and increase its amyloidogenicity *in vitro* (44). Elevated ROS and oxidative modifications can lead to TTR aggregation and amyloid formation (45). Indeed, TTR can cause senile systemic amyloidosis (46). On the other hand, both APOA1 and TTR can have anti-inflammatory and antioxidant properties (45). While overexpression of both genes is observed in malignancies of the thyroid and other organs, it may be a consequence of cell dysregulation caused by the tumor microenvironment (47). Thus, TTR and APOA1 proteins can interact, and it has been described in the brain, serum and amyloid plaques, but not within the thyroid.

BAP is a known agonist of the aryl hydrocarbon receptor (AHR), a transcription factor mostly located in the cytosol in its inactivated state. Upon ligand binding, it moves into the nucleus and dimerizes with its binding partner the AHR nuclear translocator (ARNT) (48). In the nucleus, the complex activates the expression of genes with xenobiotic metabolizing activity such as the aldehyde dehydrogenase 3 family (*ALDH3*), cytochrome P450 members *CYP1A1*, *CYP1A2* and *CYP1B1*, glutathione-S-transferases (GSTs) and the NAD(P)H quinone dehydrogenase 1 (*NQO1*) (49). Activated AHR also induces the expression of its repressor, the aryl hydrocarbon receptor repressor (*AHRR*) (48) and the TCDD inducible poly (ADP-ribose) polymerase (*TIPARP*) (50). Among all treatments tested in this study (EDCs or sex hormones), BAP showed the highest level of DEGs. CYP1A1 and CYP1B1 are the main BAP metabolizers. Irrespective of the presence of sex hormones, we observed a strong induction of both *CYP1B1*

and *CYP1A1* (although in the response to BAP only, the level of *CYP1A1* induction did not reach a high enough fold change to pass the threshold in the mature thyrocytes population). Additionally, we observed a stronger induction when "female" sex hormones were added, and this was true also for *NQO1*, the other consistent BAP-induced DEG we observed in our dataset. DHT has been shown to facilitate the formation of an AHR-AR complex and repress *CYP1B1* induction by AHR agonists (51, 52). On the opposite side, there is evidence for E2 ability to induce *CYP1B1* induction via ER-alpha activation (53). Angus et al (1999), showed how CYP1B1 is expressed at basal levels and can be induced by AHR agonists, while CYP1A1 can be detected only with treatment. This induction is strongly correlated with ER status (ER- or ER+) and does not depend on AHR levels (54). Oppositely, Spink et al (2003) showed how E2 treatment of the breast cancer MCF-7 cell line decreased the expression of *CYP1A1* and increased that of *CYP1B1* (55). BAP metabolites are the compounds with biological activity: they induce oxidative stress and form DNA adducts that cause DNA damage and trigger apoptosis (56). In both the sample treated with BAP only (comparison "Untr BAP vs Untr Ctrl") and with BAP + "male" hormones (comparison "Male BAP vs Male Ctrl") we saw upregulation of genes involved in cell cycle arrest (*CDKN1A* upregulation) and indirect cell cycle promotion (*MDM2* upregulation in Untr BAP), downregulation of mitophagy (*PRKN*, the main positive regulator) and pro- (*BAX* upregulation in Male BAP) and anti-apoptotic signals (*FHIT* and *WWOX* downregulation). *CDKN1A* expression is transcriptionally induced by activated p53 in response to stress stimuli such as DNA damage (57), while *MDM2* is a negative regulator of p53 and is induced upon p53 activation in a negative feedback loop to restore homeostasis (58). The activation of p53 leads to growth arrest and apoptosis (58). Additionally, genes involved in response to oxidative stress were dysregulated in the samples treated with BAP only (*NFE2L2*, *MSRA*, *ROMO1*) or with the addition of "male" hormones (*NFE2L2*). *NFE2L2* and *SOX4* and are two transcription factors both upregulated by treatment with BAP alone and previously described to be induced by treatment with the BAP metabolite benzo[a]pyrene diol epoxide (BPDE) (59). The absence of a clear induction or suppression of apoptosis at the transcriptomic level could be due to the reduced time of exposure (24 hours). Indeed, stress signals in a cell can trigger several different response mechanisms, and their integration ultimately tips the balance between cell survival or death.

Treatment with BAP in presence of "male" sex hormones ("Male BAP vs Male Ctrl") upregulated some NFkB target genes with a positive role in inflammation (*CXCL2, CXCL3, IER2, IER3, IL1R1, IL32, MAPK6, NFKB1, NFKBIA, NFKBIZ, S100A9, SAA1*) (60, 61). When analyzing how "male" hormones influence the response to BAP ("Male BAP vs Untr BAP") was also saw an upregulation of genes involved in inflammation, although fewer (*S100A8*, *S100A9*, *SAA1*, *SLPI*), showing how BAP had a stronger effect on the cells than hormones alone.

PCBs have been described as having weak estrogenic activity (potency of 0.000001 times that of E2) (62, 63). In line with this we observed very weak effects with PCB153 treatment in the context of "male" sex hormones, while we did not observe any effect in presence of "female" sex hormones. PCB153 is a di-*ortho*-substituted congener with 6 chlorine substitutions and is non-planar. As such, it is not considered a AHR agonist. Increase in serum T4 levels and reduction in size of the thyroid follicles were observed following low doses of PCB153 in rats (64, 65). Additionally, estrogenic activity of PCB153 is mediated by its metabolites, generated by the activity of CYP2B6 in human (which is the only functional member of the human CYP450 family 2 subfamily B) (64). However, *CYP2B6* was expressed at very low levels in our samples, and we did not observe any changes after treatment (Figure 19). It is possible that the doses or the time of exposure we selected were not sufficient to observe significant changes in the thyrocytes population, and/or that PCB153 is not metabolized sufficiently in our *in vitro* model.

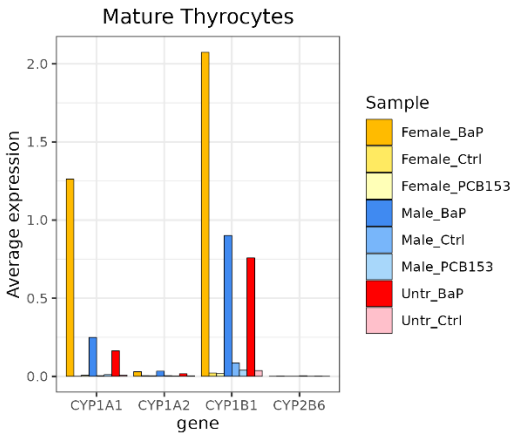


Figure 19. Average expression of CYP1 and CYP2B families in the "Mature Thyrocytes" cluster divided by sample.

To our knowledge, the effect on thyroid organoids of hormones mixtures resembling the physiological concentrations in human males and females has not been analyzed before. Bertoni et al. (2015) studied the effects on primary normal human thyrocytes of 10 nM PG and observed an upregulation of *NIS*, *TG* and *KI-67* (a marker of cell growth) (37). Santin and Furlanetto (2012) reviewed studies focusing on E2 effects on *in vitro* thyroid models, showing how there is evidence to support the hypothesis that E2 stimulates thyroid cells proliferation and function (66). This perspective adds an interesting dimension to our observation that genes associated with apoptosis did not exhibit upregulation and cell cycle arrest when combining "female" sex hormones and BAP.

In this work, we made use of single cell transcriptomics to specifically focus on the subpopulation of interest, overcoming the limitations encountered by bulk sample analysis.

Indeed, leveraging the level of resolution provided by scRNA-Seq provides the unique possibility of investigating the effects of endocrine disruption in a heterogeneous cell population such as organoids or tissues.

In conclusion, we described how different hormonal contexts affect the thyrocytes response to EDCs using thyroid organoids by analyzing the transcriptome changes via scRNA-Seq. We provide an example of how advanced *in vitro* and genomic technologies can be used to investigate toxicological responses to EDCs to reduce *in vivo* animal testing.

5. Supplementary Material

Supplementary table 1. List of differentially expressed genes. The two terms for the comparison are reported above the gene names. "pct.1" and "pct.2" refer to the percentage of cells expressing the gene in the first and second term of the comparison, respectively.

Male Ctrl vs Untr Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
MT-ND3	1.64E-23	0.669994	0.986	0.979	6.02E-19
FABP1	1.24E-22	0.610211	0.923	0.726	4.54E-18
RPS27	6.60E-21	0.502146	0.977	0.962	2.41E-16
TTR	2.37E-20	0.487502	0.815	0.506	8.66E-16
RPL39	1.85E-19	0.459062	0.968	0.958	6.76E-15
NDUFB1	3.62E-15	0.476527	0.941	0.907	1.33E-10
RPL37	3.68E-15	0.385228	0.968	0.97	1.35E-10
RPS28	4.36E-15	0.395923	0.977	0.966	1.59E-10
RPL37A	7.03E-15	0.356004	0.973	0.975	2.57E-10
ATP5F1E	2.80E-14	0.370022	0.977	0.937	1.03E-09
RPS29	8.25E-14	0.412182	0.977	0.937	3.02E-09
RPL38	7.17E-13	0.33256	0.968	0.945	2.62E-08
ATP5ME	1.89E-12	0.384656	0.955	0.903	6.92E-08
APOC3	8.56E-12	0.310157	0.739	0.468	3.13E-07
ATP5MD	1.46E-11	0.350896	0.968	0.932	5.33E-07
APOA1	1.85E-11	0.435084	0.959	0.802	6.78E-07
ATP5MPL	8.27E-11	0.29077	0.973	0.945	3.03E-06
PHGR1	2.26E-10	0.300471	0.635	0.409	8.29E-06
RPS21	5.01E-10	0.349522	0.973	0.958	1.83E-05
FABP6	8.00E-10	0.25826	0.59	0.346	2.93E-05
MT-ND4L	1.20E-09	0.378077	0.91	0.84	4.41E-05
RPL34	2.01E-09	0.327149	0.986	0.954	7.34E-05
COX7C	5.65E-09	0.28995	0.977	0.954	0.000207
SLIRP	1.74E-08	0.295484	0.932	0.861	0.000636
RPL21	1.51E-07	0.251509	0.995	0.949	0.005509
NFKBIA	1.87E-07	-0.36876	0.865	0.916	0.006828
COX6C	1.97E-07	0.315265	0.968	0.945	0.0072
NDUFA1	2.66E-07	0.253355	0.932	0.861	0.009739
RPS15A	3.93E-07	0.262184	0.968	0.958	0.014376
RPS10	8.88E-07	-0.25933	0.955	0.937	0.032501

Female Ctrl vs Untr Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
RPS28	3.76E-12	0.492423	0.964	0.966	1.38E-07
ATP5ME	1.83E-11	0.520281	0.938	0.903	6.7E-07
RPS27	1.91E-10	0.505271	0.946	0.962	6.98E-06
COX7C	8.60E-10	0.434656	0.964	0.954	3.15E-05

Female Ctrl vs Untr Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
RPL37	1.30E-09	0.427818	0.964	0.97	4.76E-05
MT-ND3	1.30E-09	0.524482	0.964	0.979	4.76E-05
ATP5F1E	4.93E-09	0.378813	0.964	0.937	0.00018
RPL37A	5.78E-09	0.31636	0.964	0.975	0.000212
RPL39	1.06E-08	0.421201	0.982	0.958	0.000388
NDUFB1	1.28E-08	0.449788	0.938	0.907	0.00047
RPL41	9.01E-08	0.327672	0.982	0.954	0.003296
RPS29	1.66E-07	0.376677	0.955	0.937	0.006067
RPL38	2.03E-07	0.288763	0.955	0.945	0.007434
ATP5MD	2.95E-07	0.372947	0.946	0.932	0.010783
ATP5MPL	1.19E-06	0.281638	0.955	0.945	0.043449

Male BAP vs Untr BAP	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
RPS29	4.11E-22	-0.533634952	0.961	0.969	1.50474E-17
RPL38	1.05E-20	-0.461118654	0.937	0.969	3.83445E-16
S100A9	4.73E-20	0.989374933	0.615	0.274	1.73089E-15
RPS27	1.87E-19	-0.487194309	0.964	0.982	6.83532E-15
APOA4	3.71E-18	0.263664201	0.874	0.744	1.35849E-13
ATP5ME	6.84E-14	-0.382603689	0.91	0.937	2.5018E-09
RPL37A	5.27E-13	-0.336657499	0.978	0.973	1.92809E-08
SAA1	1.13E-11	1.123619873	0.472	0.233	4.13796E-07
S100A8	7.13E-11	0.429091718	0.305	0.081	2.60977E-06
RPL37	1.49E-10	-0.324024683	0.985	0.973	5.45474E-06
ROMO1	2.45E-10	-0.293954387	0.855	0.919	8.98081E-06
NDUFA3	2.56E-09	-0.279210849	0.86	0.933	9.36437E-05
SLPI	2.71E-09	0.965617271	0.436	0.224	9.91403E-05
RPS28	2.82E-09	-0.321522501	0.983	0.964	0.000103259
NDUFB1	3.40E-09	-0.327853581	0.898	0.928	0.000124389
MT-ND3	4.76E-08	-0.389043663	0.99	0.982	0.001740673
RPL41	1.30E-07	-0.25140408	0.993	0.991	0.00475818
RPS21	2.17E-07	-0.311376416	0.959	0.964	0.00794786
POLR2L	1.09E-06	-0.260455596	0.937	0.955	0.039742053

Female BAP vs Untr BAP	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
MT-ND3	4.52E-20	-0.98511908	0.953	0.982	1.65527E-15
ATP5ME	4.23E-18	-0.566374469	0.859	0.937	1.54643E-13
RPS29	4.79E-17	-0.604602422	0.946	0.969	1.75319E-12
RPS27	3.16E-16	-0.568783678	0.98	0.982	1.15578E-11
RPL38	5.73E-15	-0.505110836	0.906	0.969	2.09904E-10
RPL37A	7.11E-15	-0.505543259	0.953	0.973	2.60166E-10
PET100	1.26E-12	-0.439338894	0.658	0.892	4.59567E-08

Female BAP vs Untr BAP	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
NDUFB1	5.21E-12	-0.474410919	0.826	0.928	1.90627E-07
ROMO1	3.03E-11	-0.388175752	0.832	0.919	1.10909E-06
TOMM7	4.10E-10	-0.314280548	0.819	0.942	1.49987E-05
ATP5F1E	1.71E-09	-0.315202966	0.94	0.973	6.25888E-05
RPL39	9.11E-09	-0.324872663	0.966	0.973	0.000333323
NDUFA3	1.68E-08	-0.318087868	0.792	0.933	0.000614781
RPL7A	2.17E-08	0.336163737	0.973	0.973	0.000796032
POLR2L	2.81E-08	-0.345787469	0.913	0.955	0.001028855
MT-ATP8	4.66E-08	-0.294362415	0.705	0.906	0.001704569
RPL37	5.90E-08	-0.371169194	0.98	0.973	0.002159258
RPL29	1.00E-07	0.338091103	0.98	0.973	0.003667142
SLIRP	1.19E-07	-0.290121324	0.765	0.928	0.004361239
RPL18	1.49E-07	0.350778768	0.966	0.969	0.005444352
RPS4X	3.03E-07	0.338933922	0.987	0.987	0.011084103
COX7C	3.23E-07	-0.306671787	0.94	0.96	0.011812816
SEC62	3.68E-07	-0.328378238	0.893	0.964	0.0134585
NACA	1.12E-06	0.309650489	0.98	0.973	0.041091516
RPL10	1.20E-06	0.379854442	0.987	0.987	0.043841858

Untr BAP vs Untr Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
RPS27	8.06E-38	0.860539	0.982	0.962	2.95E-33
ATP5ME	4.35E-35	0.772068	0.937	0.903	1.59E-30
RPS29	1.21E-34	0.810621	0.969	0.937	4.41E-30
RPL37A	2.98E-32	0.66002	0.973	0.975	1.09E-27
PHGR1	3.82E-32	0.589605	0.852	0.409	1.4E-27
RPL38	3.77E-31	0.657633	0.969	0.945	1.38E-26
RPL39	1.49E-29	0.689405	0.973	0.958	5.46E-25
RPS28	3.34E-28	0.710028	0.964	0.966	1.22E-23
WWOX	2.24E-27	-1.0512	0.767	0.949	8.22E-23
RPL37	3.33E-26	0.656644	0.973	0.97	1.22E-21
MT-ND3	8.93E-26	0.962429	0.982	0.979	3.27E-21
NDUFB1	3.53E-25	0.684443	0.928	0.907	1.29E-20
ATP5F1E	7.29E-22	0.521584	0.973	0.937	2.67E-17
CYP1B1	2.12E-21	0.76213	0.439	0.068	7.77E-17
FHIT	3.58E-20	-0.89553	0.812	0.966	1.31E-15
RPL34	5.24E-19	0.54049	0.978	0.954	1.92E-14
RPS21	5.42E-19	0.630514	0.964	0.958	1.98E-14
TTR	5.53E-19	0.53872	0.825	0.506	2.02E-14
FABP1	5.85E-19	0.4831	0.924	0.726	2.14E-14
NDUFA3	2.31E-17	0.457424	0.933	0.848	8.44E-13
POLR2L	4.15E-17	0.50971	0.955	0.924	1.52E-12
PRKN	5.97E-17	-0.71219	0.543	0.827	2.19E-12
ROMO1	6.62E-17	0.445084	0.919	0.903	2.42E-12

Untr BAP vs Untr Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
ATP5MD	7.89E-17	0.483169	0.942	0.932	2.89E-12
MACROD2	1.18E-16	-0.74213	0.587	0.84	4.32E-12
APOC3	5.75E-16	0.445396	0.803	0.468	2.11E-11
RPL41	1.53E-15	0.43384	0.991	0.954	5.59E-11
PET100	2.42E-15	0.418014	0.892	0.764	8.87E-11
MT-ND4L	1.73E-14	0.574989	0.933	0.84	6.32E-10
RPL36	2.37E-14	0.564855	0.987	0.962	8.66E-10
COX7C	2.25E-13	0.403785	0.96	0.954	8.23E-09
AF305872.2	5.54E-13	-0.73971	0.717	0.848	2.03E-08
C16orf74	1.86E-12	0.548763	0.717	0.494	6.8E-08
RPL36A	2.22E-12	0.478881	0.96	0.941	8.13E-08
LRMDA	3.58E-12	-0.49144	0.646	0.844	1.31E-07
SEC62	5.93E-12	0.391158	0.964	0.958	2.17E-07
FABP2	6.33E-12	0.264666	0.525	0.232	2.32E-07
UQCR11	1.09E-11	0.376804	0.937	0.958	3.98E-07
MGLL	2.17E-11	0.593615	0.753	0.608	7.93E-07
BBS9	2.68E-11	-0.51223	0.726	0.89	9.82E-07
FABP6	4.30E-11	0.28578	0.628	0.346	1.57E-06
SLIRP	6.34E-11	0.311663	0.928	0.861	2.32E-06
RPS26	7.51E-11	0.386076	0.955	0.966	2.75E-06
NDUFA1	9.49E-11	0.312359	0.906	0.861	3.48E-06
AC034195.1	2.03E-10	-0.47549	0.332	0.612	7.44E-06
C4orf48	2.18E-10	0.4061	0.933	0.886	8E-06
DAB1	2.53E-10	-0.35484	0.184	0.426	9.27E-06
TOMM7	2.75E-10	0.300894	0.942	0.911	1.01E-05
EXOC4	3.35E-10	-0.35506	0.91	0.992	1.23E-05
ATP5MPL	3.54E-10	0.299053	0.924	0.945	1.29E-05
UBL5	6.12E-10	0.300187	0.937	0.932	2.24E-05
CCSER1	8.11E-10	-0.66454	0.713	0.865	2.97E-05
AL049828.1	8.34E-10	-0.44814	0.453	0.684	3.05E-05
CDKN1A	2.62E-09	0.642884	0.691	0.485	9.6E-05
MSRA	3.44E-09	-0.2898	0.413	0.633	0.000126
LGALS3	3.62E-09	0.43309	0.749	0.549	0.000132
GMDS	4.74E-09	-0.48894	0.7	0.878	0.000173
TMEM258	5.81E-09	0.289449	0.933	0.928	0.000213
TG	6.29E-09	-0.6372	0.973	1	0.00023
SLA	7.68E-09	-0.66324	0.843	0.949	0.000281
NFE2L2	1.08E-08	0.314589	0.973	0.954	0.000397
MRPL33	1.32E-08	0.284951	0.879	0.814	0.000482
ESRRG	1.38E-08	-0.41729	0.628	0.84	0.000506
SND1	1.56E-08	-0.2846	0.83	0.92	0.000571
MDM2	1.78E-08	0.676426	0.776	0.629	0.000652
UQCRQ	1.80E-08	0.306243	0.919	0.924	0.000658
RPS27L	2.01E-08	0.526241	0.924	0.92	0.000735
SOX4	2.10E-08	0.566261	0.906	0.865	0.00077

Untr BAP vs Untr Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
COPS9	2.40E-08	0.282449	0.919	0.899	0.00088
CNTN5	3.20E-08	-0.6618	0.583	0.785	0.001172
WDPCP	3.45E-08	-0.41921	0.623	0.789	0.001264
AC021231.1	4.40E-08	-0.26769	0.265	0.481	0.00161
SON	5.83E-08	0.25125	0.973	0.949	0.002133
BCAS3	6.23E-08	-0.39561	0.888	0.945	0.00228
ITGB1	6.80E-08	0.270239	0.964	0.949	0.002489
FARS2	7.08E-08	-0.27958	0.789	0.878	0.002593
ULK4	8.77E-08	-0.27686	0.601	0.806	0.00321
CALR	1.21E-07	-0.41595	0.964	0.954	0.004417
HSPA5	1.25E-07	-0.48833	0.973	0.958	0.004574
MT-CYB	1.32E-07	-0.34779	0.996	0.983	0.004829
SMYD3	1.52E-07	-0.42123	0.906	0.962	0.005551
TRAPPC9	1.54E-07	-0.38546	0.883	0.966	0.005636
GDF15	2.01E-07	0.865122	0.287	0.105	0.007362
LINC01473	2.02E-07	-0.29958	0.735	0.895	0.007389
TCERG1L	2.28E-07	-0.46082	0.489	0.692	0.008338
ZRANB2-AS2	2.48E-07	-0.27229	0.359	0.586	0.00907
SNHG6	3.07E-07	0.256556	0.87	0.831	0.011233
SCFD2	3.12E-07	-0.3218	0.493	0.688	0.011409
RBFOX1	3.46E-07	-0.53115	0.184	0.384	0.012671
SORCS1	3.69E-07	-0.45736	0.57	0.759	0.013511
DIAPH2	3.72E-07	-0.32183	0.785	0.903	0.013622
NPAS3	3.72E-07	-0.51003	0.408	0.595	0.013626
ERC2	3.76E-07	-0.33945	0.444	0.65	0.01378
UQCRB	3.92E-07	0.25208	0.964	0.932	0.014335
SNX29	3.99E-07	-0.26351	0.717	0.844	0.014592
NQO1	4.17E-07	0.556589	0.735	0.633	0.01525
SYNE2	5.51E-07	-0.3418	0.874	0.924	0.020157
VTI1A	8.15E-07	-0.29586	0.883	0.92	0.029828
RAD51B	9.19E-07	-0.47729	0.516	0.705	0.033654
COX17	1.07E-06	0.298368	0.87	0.806	0.039171
CDKAL1	1.11E-06	-0.39453	0.892	0.945	0.040628

Male BAP vs Male Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
S100A9	8.25E-31	0.580759728	0.615	0.131	3.01931E-26
WWOX	5.22E-25	-0.848719074	0.758	0.955	1.91097E-20
FHIT	2.73E-21	-0.71769047	0.789	0.946	1.00018E-16
PRKN	1.91E-19	-0.576626491	0.506	0.806	6.97813E-15
CYP1B1	8.55E-19	0.808189661	0.458	0.126	3.12768E-14
MACROD2	3.39E-17	-0.702933638	0.562	0.797	1.24202E-12
MGLL	2.52E-16	0.651592378	0.755	0.581	9.21041E-12
NFKBIA	1.47E-15	0.577828427	0.927	0.865	5.39852E-11
CXCL2	7.08E-14	0.792495865	0.804	0.599	2.58997E-09

Male BAP vs Male Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
CYP1A1	2.91E-13	0.314740935	0.228	0.009	1.06346E-08
BAX	1.29E-12	0.48600958	0.709	0.572	4.72098E-08
GDF15	3.28E-12	0.499099292	0.308	0.068	1.1992E-07
IER3	4.68E-12	0.611690387	0.77	0.608	1.71175E-07
PHGR1	5.33E-12	0.593511456	0.806	0.635	1.94902E-07
SQSTM1	6.81E-12	0.58846163	0.971	0.968	2.4921E-07
SHANK2	2.69E-11	-0.489789082	0.729	0.883	9.84472E-07
NQO1	5.81E-10	0.546425526	0.707	0.604	2.12747E-05
SLA	8.92E-10	-0.599413502	0.852	0.928	3.26633E-05
IER2	9.68E-10	0.501475551	0.908	0.86	3.54141E-05
CCSER1	1.15E-09	-0.428586741	0.695	0.878	4.19259E-05
AC034195.1	2.03E-09	-0.412125524	0.375	0.604	7.42016E-05
APOC3	2.27E-09	0.447937868	0.816	0.739	8.29839E-05
CALR	3.58E-09	-0.397343386	0.964	0.973	0.000130939
HSP90B1	3.64E-09	-0.419953552	0.964	0.982	0.00013314
NFE2L2	6.10E-09	0.317580786	0.947	0.964	0.000223365
IMMP2L	6.32E-09	-0.308402872	0.797	0.946	0.000231398
CNTN5	7.92E-09	-0.730292276	0.6	0.784	0.00028977
BCAS3	8.77E-09	-0.35841189	0.843	0.946	0.000320959
LINGO2	9.62E-09	-0.422242343	0.591	0.802	0.000351923
HSPA5	1.09E-08	-0.43482556	0.949	0.982	0.000400709
CDKN1A	1.12E-08	0.590528341	0.622	0.491	0.000409739
NFKBIZ	1.68E-08	0.402145281	0.717	0.554	0.000615195
WDPCP	1.92E-08	-0.309532904	0.622	0.806	0.000702786
DIAPH2	2.01E-08	-0.420446694	0.69	0.856	0.00073439
AL049828.1	2.10E-08	-0.300613438	0.484	0.694	0.000767826
SAA1	3.62E-08	1.192160864	0.472	0.266	0.001325755
ZNF804B	3.82E-08	-0.677195495	0.119	0.288	0.001399333
FABP2	4.15E-08	0.367108049	0.554	0.369	0.001517295
IL32	5.60E-08	0.305460937	0.504	0.311	0.002048163
SDF2L1	5.67E-08	-0.36194245	0.78	0.896	0.002074151
CXCL3	6.17E-08	0.497367889	0.441	0.243	0.002258226
IL1R1	6.24E-08	0.400730335	0.77	0.649	0.002283775
FARS2	7.90E-08	-0.287827891	0.729	0.887	0.002892902
NFKB1	9.70E-08	0.526290713	0.85	0.829	0.003549638
RBFOX1	1.13E-07	-0.360923274	0.182	0.356	0.004132601
AF305872.2	1.39E-07	-0.381069536	0.705	0.842	0.005075227
MANF	1.47E-07	-0.33637377	0.821	0.905	0.005367078
RYR2	1.50E-07	-0.425333317	0.763	0.883	0.005476985
APOA4	1.81E-07	0.289780459	0.874	0.793	0.006634095
GMDS	1.83E-07	-0.326702404	0.63	0.811	0.006690353
SMYD3	1.84E-07	-0.385987999	0.942	0.964	0.006730953
LITAF	1.89E-07	0.325593248	0.92	0.937	0.006912953
DLG2	1.91E-07	-0.33804886	0.409	0.626	0.006993066
EXOC4	2.72E-07	-0.283643063	0.898	0.982	0.009957542

Male BAP vs Male Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
C16orf74	3.17E-07	0.430367101	0.579	0.459	0.011609963
KRT17	3.40E-07	0.603162324	0.441	0.257	0.012451991
LGALS3	4.89E-07	0.36089463	0.772	0.667	0.017907663
CPQ	6.26E-07	-0.296720304	0.797	0.896	0.02292487
RORA	6.67E-07	-0.402460861	0.787	0.887	0.024415502
CNTN4	7.14E-07	-0.433073838	0.574	0.775	0.026142651
MAPK6	7.72E-07	0.268828604	0.852	0.802	0.028266712
BBS9	9.28E-07	-0.293319472	0.714	0.86	0.033956608
PPIB	1.03E-06	-0.283851416	0.954	0.968	0.037690825
ADAMTSL4-AS1	1.05E-06	0.340127158	0.659	0.527	0.038557401
TG	1.32E-06	-0.450244752	0.973	0.991	0.04825757
SH3RF3	1.34E-06	-0.2992562	0.329	0.514	0.049081114
S100A9	8.25E-31	0.580759728	0.615	0.131	3.01931E-26
WWOX	5.22E-25	-0.848719074	0.758	0.955	1.91097E-20
FHIT	2.73E-21	-0.71769047	0.789	0.946	1.00018E-16
PRKN	1.91E-19	-0.576626491	0.506	0.806	6.97813E-15
CYP1B1	8.55E-19	0.808189661	0.458	0.126	3.12768E-14

Female BAP vs Female Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
CYP1B1	3.90E-16	1.589700879	0.544	0.08	1.42691E-11
ZC3H12B	3.81E-09	-0.318346454	0.168	0.5	0.000139568
FHIT	4.34E-09	-0.821589212	0.779	0.964	0.000159012
WWOX	4.56E-09	-0.803560906	0.752	0.92	0.000166907
CYP1A1	5.45E-09	1.177659708	0.262	0	0.000199633
CCSER1	5.86E-09	-0.749249518	0.664	0.875	0.000214461
NEB	1.87E-08	-0.671504384	0.322	0.679	0.000684304
MT-ND3	5.38E-08	-0.547172551	0.953	0.964	0.001969436
ERC2	8.71E-08	-0.35671122	0.295	0.607	0.003186801
PRKN	9.57E-08	-0.517140712	0.51	0.812	0.003503268
AREG	1.76E-07	0.838230537	0.43	0.17	0.006452598
BCAS3	2.59E-07	-0.451536419	0.772	0.946	0.009494997
NQO1	3.05E-07	0.800427509	0.725	0.554	0.011165614
SUMF1	3.77E-07	-0.302302993	0.671	0.929	0.013799057
LINC01572	4.12E-07	-0.437137033	0.51	0.812	0.015090413
RORA	4.19E-07	-0.589703167	0.711	0.911	0.01534479
BBS9	7.69E-07	-0.395655954	0.617	0.875	0.028158199
SLA	1.02E-06	-0.755535479	0.852	0.893	0.037197373
MACROD2	1.10E-06	-0.521016949	0.564	0.804	0.04021396
HSPA5	1.34E-06	-0.604656518	0.94	0.982	0.04909294
FARS2	1.36E-06	-0.27725397	0.644	0.92	0.049919485

Male PCB153 vs Male Ctrl	p-value	Average log2(Fold Change)	pct.1	pct.2	Adjusted p-value
AFP	2.74E-12	-0.3088321	0.431	0.734	1.00434E-07
APOA1	7.00E-11	-0.471154513	0.85	0.959	2.56359E-06

BAX	7.58E-09	0.304446201	0.8	0.572	0.000277476
TTR	1.03E-07	-0.362161171	0.656	0.815	0.003784397
S100A9	1.14E-07	-0.650652469	0.375	0.131	0.004175226
APOA4	2.55E-07	-0.42016351	0.662	0.793	0.009321013

6. Data Availability

The sequencing data used in this manuscript has been deposited in BioStudies under the accession E-MTAB-13502.

7. Funding

The SCREENED project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825745.

8. References

1. Moleti M, Sturniolo G, Di Mauro M, Russo M, Vermiglio F. Female Reproductive Factors and Differentiated Thyroid Cancer. Front Endocrinol (Lausanne). 2017;8:111, doi:https://doi.org/10.3389/fendo.2017.00111.
2. dos Santos Silva I, Swerdlow AJ. Sex differences in the risks of hormone-dependent cancers. Am J Epidemiol. 1993;138(1):10-28, doi:https://doi.org/10.1093/oxfordjournals.aje.a116773.
3. Suteau V, Munier M, Briet C, Rodien P. Sex Bias in Differentiated Thyroid Cancer. Int J Mol Sci. 2021;22(23), doi:https://doi.org/10.3390/ijms222312992.
4. Mori M, Naito M, Watanabe H, Takeichi N, Dohi K, Ito A. Effects of sex difference, gonadectomy, and estrogen on N-methyl-N-nitrosourea induced rat thyroid tumors. Cancer research. 1990;50(23):7662-7.
5. Merrill SJ, Mu Y. Thyroid autoimmunity as a window to autoimmunity: An explanation for sex differences in the prevalence of thyroid autoimmunity. J Theor Biol. 2015;375:95-100, doi:https://doi.org/10.1016/j.jtbi.2014.12.015.
6. COMMISSION REGULATION (EU) 2018/605. Amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties. 19 April 2018, doi:https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2018:101:TOC.
7. Hyunok Choi, Roy Harrison, Hannu Komulainen, Saborit JMD. Polycyclic aromatic hydrocarbons. 2010. Geneva: World Health Organization. WHO Guidelines for Indoor Air Quality: Selected Pollutants. Available from: https://www.ncbi.nlm.nih.gov/books/NBK138709/.
8. Agency for Toxic Substances and Disease Registry (ATSDR) (US). 6 POTENTIAL FOR HUMAN EXPOSURE. 2000. In: Toxicological Profile for Polychlorinated Biphenyls (PCBs) [Internet]. Atlanta (GA).
9. Agency for Toxic Substances and Disease Registry (ATSDR) (US). 3.6 ENDOCRINE DISRUPTION. 2000. In: Toxicological Profile for Polychlorinated Biphenyls (PCBs) [Internet]. Atlanta (GA). Available from: https://www.ncbi.nlm.nih.gov/books/NBK587429/.
10. Palanza P, Nagel SC, Parmigiani S, Vom Saal FS. Perinatal exposure to endocrine disruptors: sex, timing and behavioral endpoints. Curr Opin Behav Sci. 2016;7:69-75, doi:https://doi.org/10.1016/j.cobeha.2015.11.017.
11. Bowman R, Frankfurt M, Luine V. Sex differences in cognition following variations in endocrine status. Learn Mem. 2022;29(9):234-45, doi:https://doi.org/10.1101/lm.053509.121.
12. Rebuli ME, Patisaul HB. Assessment of sex specific endocrine disrupting effects in the prenatal and pre-pubertal rodent brain. J Steroid Biochem Mol Biol. 2016;160:148-59, doi:https://doi.org/10.1016/j.jsbmb.2015.08.021.
13. Braniste V, Jouault A, Gaultier E, Polizzi A, Buisson-Brenac C, Leveque M, et al. Impact of oral bisphenol A at reference doses on intestinal barrier function and sex differences after

perinatal exposure in rats. Proc Natl Acad Sci U S A. 2010;107(1):448-53, doi:https://doi.org/10.1073/pnas.0907697107.
14. Liu Y, Chen L, Yu J, Ye L, Hu H, Wang J, Wu B. Advances in Single-Cell Toxicogenomics in Environmental Toxicology. Environ Sci Technol. 2022;56(16):11132-45, doi:https://doi.org/10.1021/acs.est.2c01098.
15. Romitti M, Eski SE, Fonseca BF, Gillotay P, Singh SP, Costagliola S. Single-Cell Trajectory Inference Guided Enhancement of Thyroid Maturation In Vitro Using TGF-Beta Inhibition. Frontiers in Endocrinology. 2021;12, doi:https://doi.org/10.3389/fendo.2021.657195.
16. Romitti M, Tourneur A, de Faria da Fonseca B, Doumont G, Gillotay P, Liao X-H, et al. Transplantable human thyroid organoids generated from embryonic stem cells to rescue hypothyroidism. Nature Communications. 2022;13(1):7057, doi:https://doi.org/10.1038/s41467-022-34776-7.
17. Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao XH, Dumitrescu AM, et al. Generation of functional thyroid from embryonic stem cells. Nature. 2012;491(7422):66-71, doi:https://doi.org/10.1038/nature11525.
18. 10X Genomics. Chromium Single Cell 3' Reagent Kits (v3.1 Chemistry Dual Index) User Guide Rev E 2022 [Available from: https://cdn.10xgenomics.com/image/upload/v1668017706/support-documents/CG000315_ChromiumNextGEMSingleCell3-_GeneExpression_v3.1_DualIndex__RevE.pdf.
19. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87 e29, doi:https://doi.org/10.1016/j.cell.2021.04.048.
20. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat Commun. 2022;13(1):1246, doi:https://doi.org/10.1038/s41467-022-28803-w.
21. Marsh SE. scCustomize: Custom Visualizations & Functions for Streamlined Analyses of Single Cell Sequencing 2021 [Available from: https://doi.org/10.5281/zenodo.5706430.
22. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141, doi:https://doi.org/10.1016/j.xinn.2021.100141.
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9, doi:https://doi.org/10.1038/75556.
24. Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1), doi:https://doi.org/10.1093/genetics/iyad031.
25. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25(2):288-9, doi:https://doi.org/10.1093/bioinformatics/btn615.
26. Ferri FF. Ferri's Clinical Advisor 2019. Philadelphia, PA: Elsevier; 2019. p. 1872.

27. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165(3):535-50, doi:10.1016/j.cell.2016.03.014.
28. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419, doi:http://dx.doi.org/10.1126/science.1260419.
29. Chen GG, Vlantis AC, Zeng Q, van Hasselt CA. Regulation of cell growth by estrogen signaling and potential targets in thyroid cancer. Curr Cancer Drug Targets. 2008;8(5):367-77, doi:https://doi.org/10.2174/156800908785133150.
30. Kansakar E, Chang Y-J, Mehrabi M, Mittal V. Expression of Estrogen Receptor, Progesterone Receptor, and Vascular Endothelial Growth Factor-A in Thyroid Cancer. The American Surgeon. 2009;75(9):785-9, doi:https://doi.org/10.1177/000313480907500908.
31. Egawa C, Miyoshi Y, Iwao K, Shiba E, Noguchi S. Quantitative analysis of estrogen receptor-alpha and -beta messenger RNA expression in normal and malignant thyroid tissues by real-time polymerase chain reaction. Oncology. 2001;61(4):293-8, doi:https://doi.org/10.1159/000055336.
32. Taylor AH, Al-Azzawi F. Immunolocalisation of oestrogen receptor beta in human tissues. J Mol Endocrinol. 2000;24(1):145-55, doi:https://doi.org/10.1677/jme.0.0240145.
33. Kawabata W, Suzuki T, Moriya T, Fujimori K, Naganuma H, Inoue S, et al. Estrogen receptors (alpha and beta) and 17beta-hydroxysteroid dehydrogenase type 1 and 2 in thyroid disorders: possible in situ estrogen synthesis and actions. Mod Pathol. 2003;16(5):437-44, doi:https://doi.org/10.1097/01.MP.0000066800.44492.1B.
34. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. Mod Pathol. 2008;21 Suppl 2:S8-S15, doi:https://doi.org/10.1038/modpathol.2008.34.
35. Rossi R, Franceschetti P, Maestri I, Magri E, Cavazzini L, degli Uberti EC, del Senno L. Evidence for androgen receptor gene expression in human thyroid cells and tumours. J Endocrinol. 1996;148(1):77-85, doi:https://doi.org/10.1677/joe.0.1480077.
36. Stanley JA, Aruldhas MM, Chandrasekaran M, Neelamohan R, Suthagar E, Annapoorna K, et al. Androgen receptor expression in human thyroid cancer tissues: a potential mechanism underlying the gender bias in the incidence of thyroid cancers. J Steroid Biochem Mol Biol. 2012;130(1-2):105-24, doi:https://doi.org/10.1016/j.jsbmb.2012.02.004.
37. Bertoni AP, Brum IS, Hillebrand AC, Furlanetto TW. Progesterone Upregulates Gene Expression in Normal Human Thyroid Follicular Cells. Int J Endocrinol. 2015;2015:864852, doi:https://doi.org/10.1155/2015/864852.
38. Chen D, Qi W, Zhang P, Guan H, Wang L. Expression of the estrogen receptor alpha, progesterone receptor and epidermal growth factor receptor in papillary thyroid carcinoma tissues. Oncol Lett. 2015;10(1):317-20, doi:https://doi.org/10.3892/ol.2015.3223.
39. Yavropoulou MP, Panagiotou G, Topouridou K, Karayannopoulou G, Koletsa T, Zarampoukas T, et al. Vitamin D receptor and progesterone receptor protein and gene expression in papillary thyroid carcinomas: associations with histological features. J Endocrinol Invest. 2017;40(12):1327-35, doi:https://doi.org/10.1007/s40618-017-0700-4.

40. Reinecke F, Smeitink JA, van der Westhuizen FH. OXPHOS gene expression and control in mitochondrial disorders. Biochim Biophys Acta. 2009;1792(12):1113-21, doi:https://doi.org/10.1016/j.bbadis.2009.04.003.
41. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004;101(16):6062-7, doi:https://doi.org/10.1073/pnas.0400782101.
42. Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI. BioGPS: building your own mash-up of gene annotations and expression profiles. Nucleic Acids Res. 2016;44(D1):D313-6, doi:https://doi.org/10.1093/nar/gkv1104.
43. Storch J, Thumser AE. Tissue-specific functions in the fatty acid-binding protein family. J Biol Chem. 2010;285(43):32679-83, doi:https://doi.org/10.1074/jbc.R110.135210.
44. Vieira M, Saraiva MJ. Transthyretin: a multifaceted protein. Biomol Concepts. 2014;5(1):45-54, doi:https://doi.org/10.1515/bmc-2013-0038.
45. Wieczorek E, Ozyhar A. Transthyretin: From Structural Stability to Osteoarticular and Cardiovascular Diseases. Cells. 2021;10(7), doi:https://doi.org/10.3390/cells10071768.
46. White JT, Kelly JW. Support for the multigenic hypothesis of amyloidosis: the binding stoichiometry of retinol-binding protein, vitamin A, and thyroid hormone influences transthyretin amyloidogenicity in vitro. Proceedings of the National Academy of Sciences. 2001;98(23):13019-24, doi:https://doi.org/10.1073/pnas.241406698.
47. Giusti L, Iacconi P, Ciregia F, Giannaccini G, Donatini GL, Basolo F, et al. Fine-needle aspiration of thyroid nodules: proteomic analysis to identify cancer biomarkers. J Proteome Res. 2008;7(9):4079-88, doi:https://doi.org/10.1021/pr8000404.
48. Vogel CFA, Haarmann-Stemmann T. The aryl hydrocarbon receptor repressor - More than a simple feedback inhibitor of AhR signaling: Clues for its role in inflammation and cancer. Curr Opin Toxicol. 2017;2:109-19, doi:https://doi.org/10.1016/j.cotox.2017.02.004.
49. Dietrich C. Antioxidant Functions of the Aryl Hydrocarbon Receptor. Stem Cells Int. 2016;2016:7943495, doi:https://doi.org/10.1155/2016/7943495.
50. Diani-Moore S, Ram P, Li X, Mondal P, Youn DY, Sauve AA, Rifkind AB. Identification of the aryl hydrocarbon receptor target gene TiPARP as a mediator of suppression of hepatic gluconeogenesis by 2,3,7,8-tetrachlorodibenzo-p-dioxin and of nicotinamide as a corrective agent for this effect. J Biol Chem. 2010;285(50):38801-10, doi:https://doi.org/10.1074/jbc.M110.131573.
51. Hwang DY, Cho JS, Chae KR, Kang TS, Hwang JH, Lim CH, et al. Differential expression of the tetracycline-controlled transactivator-driven human CYP1B1 gene in double-transgenic mice is due to androgens: application for detecting androgens and antiandrogens. Arch Biochem Biophys. 2003;415(2):137-45, doi:https://doi.org/10.1016/S0003-9861(03)00218-2.
52. Sanada N, Gotoh Y, Shimazawa R, Klinge CM, Kizu R. Repression of activated aryl hydrocarbon receptor-induced transcriptional activation by 5alpha-dihydrotestosterone in human prostate cancer LNCaP and human breast cancer T47D cells. J Pharmacol Sci. 2009;109(3):380-7, doi:https://doi.org/10.1254/jphs.08328FP.

53. Tsuchiya Y, Nakajima M, Kyo S, Kanaya T, Inoue M, Yokoi T. Human CYP1B1 is regulated by estradiol via estrogen receptor. Cancer Res. 2004;64(9):3119-25, doi:https://doi.org/10.1158/0008-5472.CAN-04-0166.
54. Angus WG, Larsen MC, Jefcoate CR. Expression of CYP1A1 and CYP1B1 depends on cell-specific factors in human breast cancer cell lines: role of estrogen receptor status. Carcinogenesis. 1999;20(6):947-55, doi:https://doi.org/10.1093/carcin/20.6.947.
55. Spink DC, Katz BH, Hussain MM, Pentecost BT, Cao Z, Spink BC. Estrogen regulates Ah responsiveness in MCF-7 breast cancer cells. Carcinogenesis. 2003;24(12):1941-50, doi:https://doi.org/10.1093/carcin/bgg162.
56. Gelboin HV. Benzo[alpha]pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes. Physiol Rev. 1980;60(4):1107-66, doi:https://doi.org/10.1152/physrev.1980.60.4.1107.
57. Engeland K. Cell cycle regulation: p53-p21-RB signaling. Cell Death Differ. 2022;29(5):946-60, doi:https://doi.org/10.1038/s41418-022-00988-z.
58. Vousden KH. p53: death star. Cell. 2000;103(5):691-4, doi:https://doi.org/10.1016/S0092-8674(00)00171-9.
59. Wang A, Gu J, Judson-Kremer K, Powell KL, Mistry H, Simhambhatla P, et al. Response of human mammary epithelial cells to DNA damage induced by BPDE: involvement of novel regulatory pathways. Carcinogenesis. 2003;24(2):225-34, doi:https://doi.org/10.1093/carcin/24.2.225.
60. Xing Y, Zhou F, Wang J. Subset of genes targeted by transcription factor NF-kappaB in TNFalpha-stimulated human HeLa cells. Funct Integr Genomics. 2013;13(1):143-54, doi:https://doi.org/10.1007/s10142-012-0305-0.
61. Decombis S, Papin A, Bellanger C, Sortais C, Dousset C, Le Bris Y, et al. The IL32/BAFF axis supports prosurvival dialogs in the lymphoma ecosystem and is disrupted by NIK inhibition. Haematologica. 2022;107(12):2905-17, doi:https://doi.org/10.3324/haematol.2021.279800.
62. DeCastro BR, Korrick SA, Spengler JD, Soto AM. Estrogenic activity of polychlorinated biphenyls present in human tissue and the environment. Environ Sci Technol. 2006;40(8):2819-25, doi:https://doi.org/10.1021/es051667u.
63. Agency for Toxic Substances and Disease Registry (ATSDR) (US). Toxicological Profile for Polychlorinated Biphenyls (PCBs). Department of Health and Human Services, Public Health Service; 2000. Report No.: NTIS/02928747_a.
64. Li MH, Zhao YD, Hansen LG. Multiple dose toxicokinetic influence on the estrogenicity of 2,2',4,4',5,5'-hexachlorobiphenyl. Bull Environ Contam Toxicol. 1994;53(4):583-90, doi:https://doi.org/10.1007/BF00199030.
65. Desaulniers D, Leingartner K, Wade M, Fintelman E, Yagminas A, Foster WG. Effects of acute exposure to PCBs 126 and 153 on anterior pituitary and thyroid hormones and FSH isoforms in adult Sprague Dawley male rats. Toxicol Sci. 1999;47(2):158-69, doi:https://doi.org/10.1093/toxsci/47.2.158.
66. Santin AP, Furlanetto TW. Role of estrogen in thyroid function and growth regulation. J Thyroid Res. 2011;2011:875125, doi:https://doi.org/10.4061/2011/875125.

Chapter 6

Discussion

One of the main aims of this thesis was investigating the direct effects of the four studied endocrine disrupting chemical (EDC) classes on *in vitro* models of thyroid using omics approaches, with a particular focus on transcriptomics. For this discussion, we could play the role of devil's advocate and put our approach under the spotlight, to investigate the relevance of some of its aspects.

Exploring EDC Effects on Thyrocytes: Insights from Four Chemical Classes

As explained in the general introduction to this thesis, there are hundreds of EDCs, and for this project, sixteen were selected; one could ask themselves what the criteria for choosing these specific classes were. They all contaminate the environment, and the general population is exposed to them daily mostly via inhalation and ingestion (1, 2). For example, phthalates are ingested via contaminated food or water, as they are released from the matrix they are contained in (1). Following this route, they are metabolized in the intestine by pancreatic enzymes and by liver first (3), and their first and second metabolites are the ones ending up in the systemic circulation and henceforth in contact with the thyroid, and these are the compounds responsible for the toxic cellular effects, rather than the unmodified chemical (3-7). A direct exposure to unmetabolized phthalates could instead occur via intravenous medical devices with plastic tubing, or with the use of syringes, to provide some examples. The way our experiments were performed, treating the models with primary phthalates, would be representative of this second case, showing relevance for the exposure of a smaller, but no less important, population.

Production of polychlorinated biphenyls (PCBs) was phased out in the 1970s, because of recognized toxicity (8). Nonetheless, they are studied to this day. What would the purpose of adding them to our panel be if they are already categorized as toxic substances? Our current knowledge is that PCBs' interference with the thyroid system occurs at the level of thyroid hormone (TH) transport, signaling or metabolism, while very few reports exist of direct effects on the thyroid (9). Unfortunately, the thyroid is an organ often understudied in toxicology. In October 2019, we performed an exhaustive review on public data repositories of transcriptomics data obtained from thyroid tissue or thyroid models exposed to the EDC panel we used in SCREENED. We only found 1 hit, a microarray dataset deriving from rats exposed to perchlorate (Figure 1). For this reason, mechanistic insight into the effects of PCBs on the thyrocytes is lacking. Notably, together with the exposures to polycyclic aromatic hydrocarbons (PAHs), we observed that thyrocytes can respond to aryl hydrocarbon receptor (AHR) agonists by upregulating the genes of the CYP450 members 1B1 and 1A2, with the induced levels of the former being higher than the latter. To our knowledge, this has been described rarely in the

thyroid, and mostly in cancer-derived cell lines. This is relevant because PCBs tend to accumulate in the liver and lipid-rich body compartments, and, for example, neonatal exposure via breastfeeding can have deleterious developmental effects. Also, absorption via inhalation or skin contact would transfer both PCBs and PAHs directly to the systemic circulation (10).

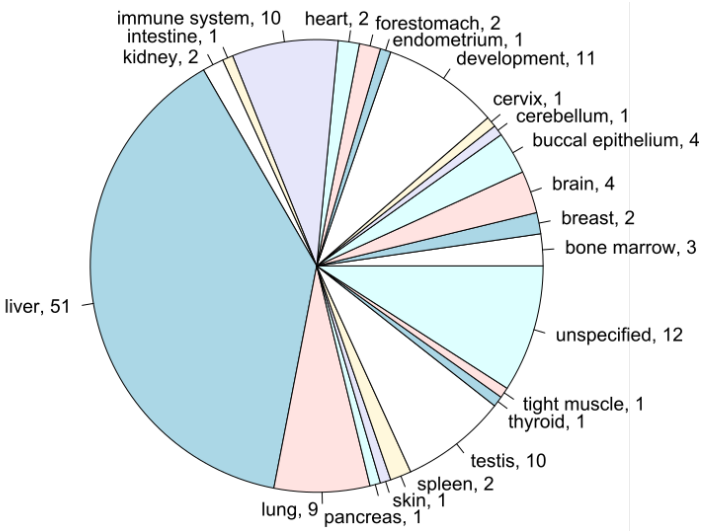


Figure 1. Number of toxicological transcriptomics datasets derived from exposure to the SCREENED endocrine disrupting chemicals panel. The number of datasets per tissue/system is reported (142 datasets in total as of October 2019). The search has been conducted on the following public repositories: Gene Expression Omnibus (GEO), ArrayExpress and the European Nucleotide Archive (ENA). Next Generation Sequencing (NGS) methods, Affymetrix or Agilent arrays were the selected platforms.

Organophosphate flame retardants (OPFRs) have been the most elusive class we tested. Unlike PAHs or PCBs, they do not have known targets (like the AHR) that they can bind or no specific biomarker for exposure (11). The full OPFRs panel was used for two screenings in the project: the one in **Chapter 4** and a similar one, not presented in this thesis, where we used the microbioreactor developed by our collaborators and only two doses (1 nM and 10 uM). Transcriptomics analyses of these datasets have given inconsistent results in terms of gene dysregulation, meaning that we could not identify compounds that are more "reactive" or toxic in terms of number of DEGs (like for PAHs or PCBs, whose reported potency in the literature was reflected on the number of dysregulated genes and the level of AHR target genes induction). Despite the different chemical structures within the four OPFRs we tested, in **Chapter 4** we identified one miRNA (mmu-miR-182-5p) upregulated at the class level. The other two we identified (mmu-miR-3076-3p and mmu-miR-6939-3p) were detected in all replicates of TPP- and TDCPP-treated samples, but in 0, 1 or 2 for DMMP and BADP: this could be both a biological effect (given, as just stated, the difference in chemical structure of the four

compounds), or a sampling bias, since the normalized read counts were less than 100 (a possible shortcoming of Combo-Seq libraries discussed in **Chapter 2**). Among the four classes, it is the one for which we could also find the least peer-review literature, especially for BADP and DDMP, making it difficult to place our results in a broader context: indeed, according to a report by the American Agency for Toxic Substances and Disease Registry (ATSDR), many health or toxicity studies on these compounds performed or sponsored by the industry remain unpublished (11).

In this context, the European project SCREENED, that financed the work carried out in this thesis and other related experiments not presented here, generated transcriptomics, proteomics and genomics datasets that we made (or will make) publicly available, filling a data gap currently present in the field of thyroid toxicology, understudied compared to other organs like the liver and intestine. To our knowledge, no extensive transcriptomics and proteomics EDC screening has been performed to date on thyroid *in vitro* models: they are instead mostly performed *in vivo* and in a smaller scale (12, 13). Larger-scale screening often use Zebrafish embryos but analyze the expression of thyroid markers (like thyroid peroxidase *TPO*, thyroid stimulating hormone receptor *TSHR*, thyroid receptor *TR*, thyroglobulin *TG*, diodinases *DIO*) or parameters like TPO activity, TR agonism/antagonist or transthyretin (TTR) binding potency (14-17).

Integrating *In Vitro* Models and Omics Approaches for Thyroid Toxicology

This thesis places itself at the crossroads of endocrine disruption testing, *in vitro* methodologies and omics approaches. It was written with the overarching thought of "can we use *in vitro* models to study the effects of endocrine disrupting chemicals using omics"? As all complex questions, it requires a complex answer. In the chapters, we observed a limited effect directly on thyroid markers or genes/proteins important for the synthesis of the TH: Tpo was upregulated at the protein level in response to DIDP treatment and *Tshr* displayed a non-monotonic dysregulation induced by sodium perchlorate in **Chapter 3**. In **Chapter 4** we observed *TG* upregulation in the presence of benzo[a]pyrene (BAP), either alone or in conjunction with "male" sex hormones, compared to the respective controls. At a first impression, if we did not have any previous knowledge of these compounds, we would probably not be able to conclude that they are disrupting chemicals with a direct effect on the TH synthesis within the thyroid. In toxicology, the duration of exposure is critical. In *in vivo* experiments, short-term testing parallels acute human or mammalian exposure, whereas long-term treatments are performed to observe the effects of chronic exposure. *In vitro* testing, on the other hand,

allows to dissect the molecular and cellular responses to chemicals (18). The duration of the culture can be limited by several factors, like the onset of cell senescence and death when using primary cells, or the limited growth space in the case of cell lines. Nonetheless, while it cannot last for months or even years like in *in vivo* systems, it can be prolonged for up to two to four weeks, depending on the cell type. Additionally, innovative approaches like 3D culturing and organ-on-a-chip methodologies offer more physiologically representative growth conditions, encompassing structure, cell-cell contact, specific microenvironments, and a continuous flow of media (19, 20). In the experiments we performed, the exposure to EDCs lasted 24 hours, a choice aimed at capturing the early effects on the transcriptome. However, to observe an impairment of the TH synthesis machinery, we would have likely needed to prolong the incubation time, considering that the half-life of TH is 5-8 days (21): in other experiments performed for the consortium (not presented in this thesis), we treated mouse embryonic stem cell-derived follicles to benzo[a]pyrene (BAP) for 10 days in the abovementioned microbioreactor and observed a dysregulation of several thyroid genes.

The Organization for Economic Co-operation and Development (OECD) tests mentioned in the general introduction measure five endpoints as indicators of thyroid disrupting activity: histopathologic changes in the thyroid (increased follicular cell height and decrease of colloid area), increased thyroid weight, decrease of serum T3 and T4 levels, increase in serum thyroid stimulating hormone (TSH) levels, interference with (embryonic) development or metamorphosis. However, in the OECD test guidelines themselves, it is stated that the relevance of some endpoints is not discriminative of a compound's thyroid activity. As can be noted, these endpoints do not provide any mechanistic insight into thyroid disruption, as this does not concern regulatory risk assessment: what matters is a visible (deleterious) effect on the hormone levels or on the organ physiology. If we consider the cited retrospective evaluation of 128 studies performed according to these test guidelines, it becomes apparent how the sole observation of altered TH serum levels is not an indicative endpoint of thyroid disruption, due to their fluctuating levels and high level of standard deviation (22) and indeed, several other thyroid endpoints are evaluated. In the case of *in vitro* assays, on the other hand, the absence of a system (i.e. the body) that affects hormone levels removes this element of complexity, making evaluation of changes in TH production as an effect of chemical treatment more straightforward.

Our data alone does not provide conclusive evidence of endocrine disruption, underscoring the complexity of elucidating such effect. If we observe a cellular alteration in thyroid cells, does it mean that this will reflect in an alteration of the TH synthesis mechanism? Is it worth spending time trying to investigate the mechanistic effect of toxic compounds on the thyroid? In the case of application of transcriptomics to *in vitro* testing, comprehensive understanding of gene expression changes necessitates an evaluation of their impact on thyroid functionality, and incorporating assays, such as measurements of TH, is necessary to determine

how transcriptomic alterations reflect into alterations of the TH synthesis machinery. An example of such approach is the work by Song et al. (2011) (23), who used microarray data of the human follicular thyroid carcinoma FTC-238 line transformed with human recombinant TPO exposed to TPO-interfering compounds to identify genes that would allow to determine TPO activity. Combining one-way analysis of variance (ANOVA) and class prediction via k-nearest neighbors, they selected 362 genes that could classify a compound as able to increase or decrease TPO activity. A validation with the guaiacol assay, which measures TPO activity, showed 66.7% accuracy.

An additional, different, approach is the use of omics in the context of Adverse Outcome Pathways (AOPs): they can help connect changes in gene or protein expression to an adverse effect irrespective of the compound that is being studied, being AOPs cell- and compound-agnostic. This could be applied, for example, to the categorization of compounds: if their action can be traced back to a known Molecular Initiating Event (MIE) and eventual Adverse Outcome (AO), they could be classified as toxic or not, and could be an alternative to read-across methods. The strong point of transcriptomics, which has been the main focus of this thesis, is its ability to provide a snapshot of the expression status of all genes in a cell population at a given time, allowing to generate hypotheses that can be further investigated, since it is the combined evidence provided by several approaches, the reproducibility of results and the continuous accumulation of evidence that contributes to elucidating the mechanisms of toxicity.

In Vitro Thyroid Models

To be employed in toxicological risk assessment, *in vitro* systems need to retain the functional characteristics of the organ(s) they are aiming at substituting (for example, in the case of the thyroid, production of the TH at a measurable level). Additionally, there needs to be little batch to batch variability, which ensures reproducibility of the results. The nature of the TH synthesis mechanism requires very polarized cells and a lumen where to store the colloid, and for this reason the 3D structure is essential for proper thyroid modelling, which explains the great effort to develop thyroid organoids as an *in vitro* system. They can be derived from two cellular sources: thyroid tissue from biopsies (24, 25), and embryonic (26-28) or (induced) pluripotent (29, 30) stem cells. Both models present advantages and limitations: the cellular population derived from tissues is highly enriched in thyrocytes (considering that more than 99% of the thyroid cell population is constituted by thyrocytes (31)). However, the starting material is limited, especially when coming from patients' biopsies, which also requires the proximity to hospitals and the collaboration with medical personnel. On the other hand, stem cell-derived organoids can be differentiated from cell lines (although not in every country due to ethical regulations). Unfortunately, such models can be constituted by more than the desired cell type

due to the complexity of tightly controlling the differentiation process. This can vary from batch to batch, and in case of induced pluripotent stem cells (iPSC), the variability in differentiation efficiency due to the donor's genetic background needs to be added. This is the challenge we faced in the work of this thesis. In the models we used for the experiments in the previous chapters, the percentage of thyrocytes is between 50% and 10% and varies from batch to batch. Additionally, we observed the expression of some genes (like *TTR*) not native to the thyroid. The other cell types do not express thyroid biomarkers (Tshr, Nis, Tpo, Tg), so they are valid models for studying the disruption of the TH synthesis pathway if it directly affects those genes. However, if the disruption is a consequence of interference with a more "common" cellular pathway, which involves genes expressed by other cell types as well, but for which the thyrocytes have a particular sensitivity, we will not be able to observe that. This is not an implausible scenario: for example, Porreca *et al.* showed how treatment of the immortalized thyrocyte cell line PCCl3 with two known thyroid disrupting chemicals ethylenethiourea and chlorpyrifos affected pathways and genes not thyroid-specific, while observing reduction of serum free T4 and decrease of *Tg* expression (13). Only techniques that allow to isolate the thyrocyte population, such as fluorescence-activated cellular sorting (FACS) or single cell RNA-Seq would be able to answer the question, which is the reason why we adopted the latter approach in **Chapter 5**. Unfortunately, these techniques are not yet scalable at the level required by high throughput toxicological screenings. Nonetheless, the number of available thyroid *in vitro* models is limited, and what we used are among the best to date.

A Role for *In Vitro* Omics in Regulatory Testing

Returning to the initial question "Can we predict thyroid disruption using *in vitro* models using omics approaches?" we could answer "Not at present, not with the existing models, but potentially within the next five to ten years". The paradigm shift in cosmetic testing, for which the OECD, in 2009, banned *in vivo* testing and introduced *in vitro* methods for the assessment of skin and eye irritation and skin sensitization unless technically infeasible serves as a noteworthy example (32, 33). The OECD-approved test using a transcriptomic panel is a success story and shows the way forward. A possible example of how such a panel for thyroid disruption testing could be built is provided by the work of Haggard *et al.* (2018), who performed a screening using TR agonists on Zebrafish embryos and identified 23 transcripts whose expression increased after all treatments (34).

The discussion of animal testing will always be a divisive one and, particularly in some circumstances, I do not envision an equivalent *in vitro* testing able to substitute it, like the case of reproductive effects (although even that has some limitations if we consider the infamous example of thalidomide). Even in five or ten years, I would not expect *in vivo* thyroid testing to

be completely superseded by *in vitro* approaches, however I believe plausible the introduction of some tests for thyroid endocrine disruption testing using cell models, considering that some are in the evaluation phase by the OECD. Even though OECD's Testing Guidelines represent suggested, and not compulsory, guidelines, the fact that some *in vitro* tests are in the development phase shows a certain degree of confidence towards them: completely forsaking animal testing is unlikely to happen (soon or ever I cannot say), but it is a change that happens in small steps, bit by bit, test after test.

References

1. Wittassek M, Koch HM, Angerer J, Bruning T. Assessing exposure to phthalates - the human biomonitoring approach. Mol Nutr Food Res. 2011;55(1):7-31, doi:https://doi.org/10.1002/mnfr.201000121.
2. Versar Inc. Review of Exposure Data and Assessments for Select Dialkyl Ortho-PhthalatesFebruary 2010. Available from: https://www.cpsc.gov/s3fs-public/pthalexp.pdf.
3. Zhang YJ, Guo JL, Xue JC, Bai CL, Guo Y. Phthalate metabolites: Characterization, toxicities, global distribution, and exposure assessment. Environ Pollut. 2021;291:118106, doi:https://doi.org/10.1016/j.envpol.2021.118106.
4. Frederiksen H, Skakkebaek NE, Andersson AM. Metabolism of phthalates in humans. Mol Nutr Food Res. 2007;51(7):899-911, doi:https://doi.org/10.1002/mnfr.200600243.
5. Peck CC, Albro PW. Toxic potential of the plasticizer Di(2-ethylhexyl) phthalate in the context of its disposition and metabolism in primates and man. Environ Health Perspect. 1982;45:11-7, doi:https://doi.org/10.1289/ehp.824511.
6. Sjoberg P, Lindqvist NG, Ploen L. Age-dependent response of the rat testes to di(2-ethylhexyl) phthalate. Environ Health Perspect. 1986;65:237-42, doi:https://doi.org/10.1289/ehp.65-1474698.
7. Babich MA. Overview of phthalates toxicity. US Consumer Product Safety Commission, Bethesda, MD 20814 [Internet]. April 2010. Available from: https://www.cpsc.gov/s3fs-public/phthalover.pdf.
8. Agency for Toxic Substances and Disease Registry (ATSDR) (US). Toxicological Profile for Polychlorinated Biphenyls (PCBs). Department of Health and Human Services, Public Health Service; 2000. Report No.: NTIS/02928747_a.
9. Liu J, Liu Y, Barter RA, Klaassen CD. Alteration of thyroid homeostasis by UDP-glucuronosyltransferase inducers in rats: a dose-response study. J Pharmacol Exp Ther. 1995;273(2):977-85.
10. Agency for Toxic Substances and Disease Registry (ATSDR) (US). 3 HEALTH EFFECTS. 2000. In: Toxicological Profile for Polychlorinated Biphenyls (PCBs) [Internet]. Atlanta (GA). Available from: https://www.ncbi.nlm.nih.gov/books/NBK587429/.
11. Agency for Toxic Substances and Disease Registry (ATSDR) (US). 3 HEALTH EFFECTS. 2012. In: Toxicological Profile for Phosphate Ester Flame Retardants [Internet]. Atlanta (GA). Available from: https://www.ncbi.nlm.nih.gov/books/NBK592277/.
12. Lu L, Wu H, Cui S, Zhan T, Zhang C, Lu S, et al. Pentabromoethylbenzene Exposure Induces Transcriptome Aberration and Thyroid Dysfunction: In Vitro, in Silico, and in Vivo Investigations. Environ Sci Technol. 2020;54(19):12335-44, doi:https://doi.org/10.1021/acs.est.0c03308.
13. Porreca I, D'Angelo F, De Franceschi L, Matte A, Ceccarelli M, Iolascon A, et al. Pesticide toxicogenomics across scales: in vitro transcriptome predicts mechanisms and outcomes of exposure in vivo. Sci Rep. 2016;6:38131, doi:https://doi.org/10.1038/srep38131

14. Collet B, Simon E, van der Linden S, El Abdellaoui N, Naderman M, Man HY, et al. Evaluation of a panel of in vitro methods for assessing thyroid receptor beta and transthyretin transporter disrupting activities. Reprod Toxicol. 2020;96:432-44, doi:https://doi.org/10.1016/j.reprotox.2019.05.011.
15. Leusch FDL, Aneck-Hahn NH, Cavanagh JE, Du Pasquier D, Hamers T, Hebert A, et al. Comparison of in vitro and in vivo bioassays to measure thyroid hormone disrupting activity in water extracts. Chemosphere. 2018;191:868-75, doi:https://doi.org/10.1016/j.chemosphere.2017.10.109.
16. Guo Y, Zhou B. Thyroid endocrine system disruption by pentachlorophenol: an in vitro and in vivo assay. Aquat Toxicol. 2013;142-143:138-45, doi:https://doi.org/10.1016/j.aquatox.2013.08.005.
17. Jarque S, Ibarra J, Rubio-Brotons M, Garcia-Fernandez J, Terriente J. Multiplex Analysis Platform for Endocrine Disruption Prediction Using Zebrafish. Int J Mol Sci. 2019;20(7), doi:https://doi.org/10.3390/ijms20071739.
18. John Timbrell, Barile FA. Introduction to Toxicology. 4th Edition ed Boca Raton: CRC Press; 2023 27 February 2023. doi:https://doi.org/10.1201/9781003188575.
19. Ma C, Peng Y, Li H, Chen W. Organ-on-a-Chip: A New Paradigm for Drug Development. Trends Pharmacol Sci. 2021;42(2):119-33, doi:https://doi.org/10.1016/j.tips.2020.11.009.
20. Antoni D, Burckel H, Josset E, Noel G. Three-dimensional cell culture: a breakthrough in vivo. Int J Mol Sci. 2015;16(3):5517-27, doi:https://doi.org/10.3390/ijms16035517.
21. Luster M, Duntas LH, Wartofsky L. The Thyroid and Its Diseases: Springer; 2019.
22. Beekhuijzen M, Rijk JCW, Meijer M, de Raaf MA, Pelgrom S. A critical evaluation of thyroid hormone measurements in OECD test guideline studies: Is there any added value? Reprod Toxicol. 2019;88:56-66, doi:https://doi.org/10.1016/j.reprotox.2019.07.014.
23. Song M, Kim YJ, Song MK, Choi HS, Park YK, Ryu JC. Identification of classifiers for increase or decrease of thyroid peroxidase activity in the FTC-238/hTPO recombinant cell line. Environ Sci Technol. 2011;45(18):7906-14, doi:https://doi.org/10.1021/es200475k.
24. van der Vaart J, Bosmans L, Sijbesma SF, Knoops K, van de Wetering WJ, Otten HG, et al. Adult mouse and human organoids derived from thyroid follicular cells and modeling of Graves' hyperthyroidism. Proc Natl Acad Sci U S A. 2021;118(51), doi:https://doi.org/10.1073/pnas.2117017118.
25. Ogundipe VML, Groen AH, Hosper N, Nagle PWK, Hess J, Faber H, et al. Generation and Differentiation of Adult Tissue-Derived Human Thyroid Organoids. Stem Cell Reports. 2021;16(4):913-25, doi:https://doi.org/10.1016/j.stemcr.2021.02.011.
26. Ma R, Latif R, Davies TF. Human embryonic stem cells form functional thyroid follicles. Thyroid. 2015;25(4):455-61, doi:https://doi.org/10.1089/thy.2014.0537.
27. Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao XH, Dumitrescu AM, et al. Generation of functional thyroid from embryonic stem cells. Nature. 2012;491(7422):66-71, doi:https://doi.org/10.1038/nature11525.

28. Romitti M, Tourneur A, de Faria da Fonseca B, Doumont G, Gillotay P, Liao X-H, et al. Transplantable human thyroid organoids generated from embryonic stem cells to rescue hypothyroidism. Nature Communications. 2022;13(1):7057, doi:https://doi.org/10.1038/s41467-022-34776-7.
29. Ma R, Shi R, Morshed SA, Latif R, Davies TF. Derivation and 97% Purification of Human Thyroid Cells From Dermal Fibroblasts. Front Endocrinol (Lausanne). 2020;11:446, doi:https://doi.org/10.3389/fendo.2020.00446.
30. Kurmann AA, Serra M, Hawkins F, Rankin SA, Mori M, Astapova I, et al. Regeneration of Thyroid Function by Transplantation of Differentiated Pluripotent Stem Cells. Cell Stem Cell. 2015;17(5):527-42, doi:https://doi.org/10.1016/j.stem.2015.09.004.
31. Schmid KW. Histopathology of C Cells and Medullary Thyroid Carcinoma. Recent Results Cancer Res. 2015;204:41-60, doi:https://doi.org/10.1007/978-3-319-22542-5_2.
32. EC. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on Cosmetic Products: OJ L342; 2009. Volume 64. Document 32009R1223. <http://data.europa.eu/eli/reg/2009/1223/oj>.
33. EC. Consolidated text: Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products (recast): OJ L342; 2023. Document 02009R1223-20230816. <http://data.europa.eu/eli/reg/2009/1223/2023-08-16>.
34. Haggard DE, Noyes PD, Waters KM, Tanguay RL. Transcriptomic and phenotypic profiling in developing zebrafish exposed to thyroid hormone receptor agonists. Reprod Toxicol. 2018;77:80-93, doi:https://doi.org/10.1016/j.reprotox.2018.02.006.

Addendum

Curriculum Vitae

Marta Nazzari (1994) was born in Marcon, a town in the province of Venice, Italy. At the University of Padua (Italy) she obtained her bachelor's degree in Molecular Biology and her master's degree in Medical Biotechnologies. For her master's thesis, she worked at the Veneto Institute of Oncology (IOV) in Padua on the project "Targeted Drug-Polymer Conjugates for Combined Dual Therapy of Ovarian Cancer" for which she tested and characterized the cytotoxicity of two bioconjugate drugs on in vitro models of ovarian cancer with cytotoxicity assays and flow cytometry.

After graduating, she spent one year and a half at the Center for Molecular Medicine at the Karolinska Institute in Stockholm, Sweden, first as an intern with the European Erasmus Project for six months, and later as a researcher. There, she worked on testing the specific T-cell reactivity in different cohorts of Multiple Sclerosis patients for the detection of novel autoantigens, and on the development of a personalized adoptive cell therapy strategy for the treatment of bladder cancer.

In 2019, she started her PhD at the University of Maastricht, the Netherlands, under the supervision of Associate Professor Florian Caiment on the project "Leveraging Multi-Omics Technologies for Studying the Effects of Endocrine Disrupting Chemicals on Thyroid *in Vitro* Models".

Scientific output

Publications

Marta Nazzari, Mírian Romitti, Anna M Kip, Rick Kamps, Sabine Costagliola, Twan van den Beucken, Lorenzo Moroni, Florian Caiment. **Impact of Endocrine Disrupting Chemicals and Sex Hormones on Human ESC-Derived Thyroid Follicles Using Single Cell Transcriptomics**. 2023. https://dx.doi.org/10.2139/ssrn.4623146 (preprint in *Environment International*, under review)

Marta Nazzari, Mírian Romitti, Duncan Hauser, Daniel J Carvalho, Stefan Giselbrecht, Lorenzo Moroni, Sabine Costagliola, Florian Caiment. **Investigation of the effects of phthalates on in vitro thyroid models with RNA-Seq and ATAC-Seq.** *Frontiers in Endocrinology.* 2023; 14, 1200211. https://doi.org/10.3389/fendo.2023.1200211

Daniel J Carvalho, Anna M Kip, Mírian Romitti, **Marta Nazzari,** Andreas Tegel, Matthias Stich, Christian Krause, Florian Caiment, Sabine Costagliola, Lorenzo Moroni, Stefan Giselbrecht. **Thyroid-on-a-Chip: An Organoid Platform for In Vitro Assessment of Endocrine Disruption.** *Advanced Healthcare Materials.* 2023; 12, 2201555. https://doi.org/10.1002/adhm.202201555

Marta Nazzari, Duncan Hauser, Marcel van Herwijnen, Mírian Romitti, Daniel J Carvalho, Anna M Kip, Florian Caiment. **CODA: a combo-Seq data analysis workflow.** *Briefings in Bioinformatics.* 2023;24(1), bbac582. https://doi.org/10.1093/bib/bbac582

Lorenzo Moroni, Barbaro Fulvio, Caiment Florian, Coleman Orla, Costagliola Sabine, Di Conza Giusy, Elviri Lisa, Giselbrecht Stefan, Krause Christian, Mota Carlos, **Marta Nazzari**, Stephen R Pennington, Annette Ringwald, Monica Sandri, Simon Thomas, James Waddington, Roberto Toni. **SCREENED: A Multistage Model of Thyroid Gland Function for Screening Endocrine- Disrupting Chemicals in a Biologically Sex-Specific Manner.** *International Journal of Molecular Sciences.* 2020, 21(10):3648. https://doi.org/10.3390/ijms21103648

Abstracts/posters

Marta Nazzari*, James Waddington*, Orla Coleman, Duncan Hauser, Stephen R. Pennington & Florian Caiment. **Proteogenomics Analysis of Endocrine Disruptors on an *In Vitro* Human Thyroid Model**. (*contributed equally) (Tissue Engineering and Regenerative Medicine International Society [TERMIS] 2021)

JC. Waddington*, **M. Nazzari***, O. Coleman, R. Zhou, L. Moroni, S. Pennington, F. Caiment. **Transcriptomic and Proteomic Analysis of Endocrine Disrupting Chemicals on an *In Vitro* Human Thyroid Model**. (*contributed equally) (EURION Annual Meeting 2022)

Marta Nazzari, Mírian Romitti, Daniel J. Carvalho, Duncan Hauser, Florian Caiment. **Investigation of the epigenetic effects of phthalates on in vitro thyroid models with RNA-Seq and ATAC-Seq**. (EURION Annual Meeting 2023, Amsterdam)

Acknowledgements

This thesis may have only my name on it, but it has not been a solo journey. Along the way, I have met many awesome and talented people, and I would like to give back just a fraction of what I have received.

First and foremost, I must thank my supervisor and promoter **Florian**, who has been the best mentor I could ever ask for. You have always been supportive but have also given me the independence of doing things my way. I have learnt some much from you: bioinformatics, how to stand up for myself and my ideas, and that it is okay to make mistakes, but the key is owning them and fixing them without worry. I admire your optimism; you say that you are a pessimist by nature, but your PhD in Silver Lining proves the opposite. You also taught me that gambling can be a good thing after all, but I am not sure this was among your objectives.

Second, I want to thank my co-promoter **Twan**, whose precious technical knowledge and advice has been key to the success of my few but important! experiments and contributed to making me prepare the best single-cell RNA-Seq libraries I have ever made.

A special thanks to the members of my Assessment Committee for their time spent reading and evaluating my thesis: Prof. **Karine Audouze**, Prof. Dr. **Maarteen Honing**, Prof. Dr. Ir. **Juliette Legler** and Dr. **Susan L. M. Steinbusch-Coort**.

Thank you **Christa**, **Marcel**, and **Duncan**, you have always helped me with kindness and have been a great support for me and my sometimes silly questions.

To **Marcha**, who has been one of my role models at TGX, I admire you a lot; and **Rossella**, who is now maintaining the contractual quota of Italians required in every workplace. Your gluten-free lasagna made me tear up of joy a bit. I want to thank also the other TGX staff members, **Bert**, **Danyel**, **Florence**, **Jacco**, **Julian**, **Simone** and **Theo**, your expertise and knowledge keeps the scientific level of our department always high. **Erika**, **Iris**, **Margot**, **Patrick**, **Rick** and **Sean**, always providing expert support and help, you make TGX work as a perfectly oiled machine.

I have shared my office with many people, and everyone has been lovely. Thanks to my former office mates **Tim**, who understood my passion for great coffee, **Rajinder**, that nothing seemed to upset, and **Juan**, my Mediterranean brother in arms, full of joy and words to say. Thank you **Somaieh**, who is in a better place above us (she is doing her postdoc on the 5th floor, what were you thinking?), you are so sweet and kind; **Toni**, TGX authority in the field of manga's, **Nicolaj**, whose style I love; **Zhiling** and **Zheping**, you all made my days at the office always a delight.

To my next-door neighbors **Nikki** and **Jelmer**, who are among the sweetest people I know and made me discover that at UM there is coffee made by real people. **Nikki** I will never

forget when you rescued me when I was stranded in Roermond. **Kaidi**, **Yueqin** and **Wen**, whose Chinese food specialties I have loved. **Na**, you have always been very kind and interested in my weird Italian lunches. **Saad**, who has eaten mangoes we people wouldn't believe, and now I cannot look at our own mangoes without a bit of sadness, and **Brian**, thank you for all your fascinating stories from a faraway place, who thought Meerssen, Limburg could be so interesting.

Thanks to my friends old and new, who have made me laugh and feel like I belong. **Sara**, **Erica**, **Elena**, **Chiara** and **Roberta**, who are proof friendships can live across borders, you are all so special. **Alessandro**, **Lorenzo** and **Andrea**, we all moved from Padova to the Netherlands, and we will continue scheming within the Shadow Government, with **Matteo** as field agent in Italy. **Julia** and **Frank**, who have become precious friends.

To my mum **Rossella**, my dad **Davide** and my big brother **Andrea**, who still do not understand what it is I do but love me, nonetheless. I love you too.

And last, but just because it is the hardest to put into words, to **Giulio**, my partner who I love more than anything I have ever loved, you are the reason I am the person I am now. You are my reference point, and the reason I strive to always be the best version of myself I can be. This PhD would not have been possible without you, but after all nothing would.