

## **Blood transcriptional and microRNA responses to short-term exposure to disinfection by-products in a swimming pool**

Almudena Espín-Pérez<sup>1</sup>, Laia Font-Ribera<sup>2</sup>, Karin van Veldhoven<sup>3</sup>, Julian Krauskopf<sup>1</sup>, Lutzen Portengen<sup>4</sup>, Marc Chadeau-Hyam<sup>3</sup>, Roel Vermeulen<sup>4</sup>, Joan O. Grimalt<sup>5</sup>, Cristina M. Villanueva<sup>2</sup>, Paolo Vineis<sup>3</sup>, Manolis Kogevinas<sup>2</sup>, Jos C Kleinjans<sup>1</sup>, Theo M. de Kok<sup>1</sup>

1 Department of Toxicogenomics, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, the Netherlands

2 Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain

3 MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK.

4 Institute for Risk Assessment Sciences, Utrecht, the Netherlands

5 Institute of Environmental Assessment and Water Research (IDAEA-CSIC). Barcelona. Catalonia. Spain.

**Corresponding author:** Almudena Espín-Pérez, Department of Toxicogenomics, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, the Netherlands . Tel +31-43-3883986. E-mail: a.espin@maastrichtuniversity.nl

**Running title:** Transcriptional and microRNA responses to DBPs

**Acknowledgements:** This work was supported by the grant FP7 of the European Commission ‘Enhanced exposure assessment and omic profiling for high priority environmental exposures in Europe’ (no. 308610).

**Conflict of interest:** None declared

## Abstract

### Background

Swimming in a chlorinated pool results in high exposure levels to disinfection by-products (DBPs), which have been associated with an increased risk of bladder cancer.

### Objectives

By studying molecular responses at the blood transcriptome level we examined the biological processes associated with exposure to these compounds.

### Methods

Whole-genome gene expression and microRNA analysis was performed on blood samples collected from 43 volunteers before and 2 hours after 40 minutes swimming in an indoor chlorinated pool (PISCINAII study). Exposure to THMs was measured in exhaled breath. Heart rate and kcal expenditure were measured as proxies for physical activity. Associations between exposure levels and gene expression were assessed using multivariate normal models (MVN), correcting for age, body mass index and sex. A Bonferroni threshold at 5% was applied.

### Results

MVN-models for the individual exposures identified 1,778 genes and 23 microRNAs that were significantly associated with exposure to at least one DBP. Due to co-linearity it was not possible to statistically disentangle responses to DBP exposure from those related to physical activity. However, after eliminating previously reported transcripts associated with physical activity a large number of hits remained associated with DBP exposure. Among those, 9 were linked with bladder and 31 with colon cancer. Concordant microRNA/mRNA expressions were identified in association with DBP exposure for hsa-mir-22-3p and hsa-miR-146a-5p and their targets RCOR1 and TLR4, both related to colon cancer in association with DBP exposure.

### Conclusions

Short-term exposure to low levels of DBPs shows genomics responses that may be indicative of increased cancer risk.

**Keywords:** disinfection by-products, transcriptomics, microRNA, multivariate normal models

## Introduction

Chlorination is a widely used method to disinfect drinking water as well as swimming pool water. Apart from its important role in preventing diseases due to microbial infections, particularly drinking water chlorination has also been associated with serious adverse health effects, such as bladder cancer and fetal growth retardation [1-4]. Chlorine in combination with organic matter yields so called disinfection by-products (DBPs), compounds that can be ingested, absorbed by the skin or inhaled [5]. Chloroform ( $\text{CHCl}_3$ ), dibromochloromethane (DBCM), bromodichloromethane (BDCM) and bromoform ( $\text{CHBr}_3$ ) belong to the most common group of DBPs, defined as Trihalomethanes (THMs). THMs are volatile and skin permeable compounds and can be quantified with validated analytical methods in biological samples, including exhaled air. More than 600 DBPs have been identified, including THMs, haloacetic acids (HAAs), trihaloacetic acids (THAAs), dichloroacetonitrile ( $\text{CHCl}_2\text{CN}$ ), chloramines, mutagen X (MX), N-nitrosodimethylamine (NDMA), haloacetamides (HAcAms), haloacetonitriles (HANs), chloral hydrate, etc [6]. The International Agency for Research on Cancer (IARC) has classified several of them as possibly carcinogenic to humans [7], which led to the implementation of guidelines for DBPs in some countries. The maximum allowed level of total trihalomethanes in the U.S. is 80  $\mu\text{g/L}$  and in the EU 100  $\mu\text{g/L}$ , although the cutoff differs between EU countries.

Evidence of DBP toxicity is based on experiments using transgenic bacteria, mammal, rodents and *in vitro* human cells [6]. THMs are formed at the highest concentrations after chlorination [2] and long-term exposure has been associated with increased risk of bladder cancer in human epidemiological studies [7, 8]. As bladder cancer is the 11nd most common cancer worldwide [9], investigating the impact of DBP exposure on bladder cancer risk is highly relevant. Notably also an association between DBP exposure and risk of colon cancer has been suggested [2, 10-12].

In view of the potential carcinogenic effects of DBPs further investigation of the biological responses to human exposure is essential for risk assessment and management. In order to support evidence-based policy making, a better understanding of the potential molecular mechanisms that underpin cancer risks associated with DBP exposure is needed [13]. Studying molecular responses on the blood transcriptome level might elucidate in more detail the biological processes associated with exposure to these compounds. Additionally, measurements of microRNAs, small non-protein-coding RNA molecules that can regulate the expression of particular gene products by for instance transcript binding [14], may provide deeper insight into underlying gene expression regulatory mechanisms.

Although THMs represent only a subgroup of DBP, levels of brominated compounds determined in exhaled breath after swimming are known to correlate well with concentrations of some DBPs in the pool water and also to trichloramine in air. For this reason THM levels have been suggested as a non-invasive biomarker of DBP exposure after swimming in chlorinated pools [15]. These

findings on exhaled bromoform and not exhaled chloroform are consistent with other studies [16, 17].

In the present study, we aim to study the molecular responses at the blood transcriptome level by investigating the biological processes associated with exposure to THMs. In view of the existing literature supporting increased risk of bladder cancer associated with THMs, we hypothesize that short-term exposure to THMs can already trigger biological responses that may contribute to an increased risk of bladder and potentially also colon cancer.

## **Methods**

### ***Study design***

In total, 116 healthy and non-smoking subjects participated in the experiment, 43 of which were included in the transcriptomics analysis. The sample size is sufficient to identify strong effects in a regression analysis according to a power calculation test (“pwr” R package). This subpopulation was selected following the criteria of not having missing values in the exposure and physical activity measurements and having a balanced numbers of females and males (22 females and 21 males) in order to reduce potential for confounding in the association between sex-dependent responses and exposure. These 43 non-smoking volunteers in the age between 18 and 40 years swam for 40 minutes in the same indoor chlorinated pool in Barcelona (Spain). The exposure time of 40 minutes is consistent with a previous study on exposure of swimmers to disinfection by products [17]. The subjects, non-professional swimmers, were allowed to rest from swimming but only while staying in the pool ( $6.6 \pm 6.7$  minutes of resting). Blood was collected from the subjects around 30 minutes before and 2 hours after swimming. Blood samples were mixed with RNAlater after collection and stored at  $-80^{\circ}\text{C}$  within two hours. To control the intake of THMs during the experiment, volunteers were asked not to swim for one week before the study and avoid showering on the day of the experiment (as it is an important source of THM exposure). During the experiment, participants were not allowed to drink tap water. Age ( $24.0 \pm 5.1$ ), gender (51% females 49% males) and body mass index (BMI) ( $24.1 \pm 2.8$ ) measurements of the subjects as well as temperature ( $27.9 \pm 0.3$ ), and pH ( $7.5 \pm 0.2$ ) of the swimming water were registered.

Swimming distance and swimming time were registered and volunteers provided a subjective score which was intended to reflect their level of physical fatigue and breathlessness after swimming. Intensity of physical activity was recorded using a heart rate (HR) monitor (Polar RCX5) during the swimming time. The percentage of HR values above 69 percent of the “HRmax”, defined as the theoretical maximum HR according to sex and age [18, 19], indicates high intensity

physical activity (60.5±30.3). In addition, the number of kilocalories (kcal) spent during the exercise (210.6±75.2) was estimated with the following formula (Ainsworth et al. 2000):

$$Kcal = weight (kg) \times \frac{distance\ swam (m)}{min} \times time\ swimming (hr) \times 8.3 \left( \frac{kcal/kg}{hr} \right) / 46 (m/min)$$

The experiment was approved by the ethics committee (Ethics Committee of Clinical Research Parc de Salut Mar, 2013/5034/I) and written informed consent forms from the volunteers were obtained. The participants were not informed about their own results.

### **DBP Exposure**

Chloroform (CHCl<sub>3</sub>), bromodichloromethane (BDCM), dibromochloromethane (DBCM) and bromoform (CHBr<sub>3</sub>) were measured in exhaled breath, before and right after swimming. The volunteers breathed deeply three times, retained the air for 10 seconds and breathed into the disposable piece of a Bio-VOC™ sampler to obtain 150 ml of alveolar air. Once this alveolar air was collected, a screw-in plunger was used to steadily discharge the sample into a sorbent tube trap. These stainless steel tubes were 3.5" (89 mm) long and ¼" o.d. They were packed with 200 mg sorbent Tenax TA 35/60 mesh (Markes International Ltd, UK) which was preconditioned using helium of 5N grade at 100 ml/min at 320°C for 2 hours and then at 335°C for 30 minutes. In posterior conditioning cycles they were reconditioned at 335°C for 20 minutes and the same flow carrier gas. One cleaned, the tubes were capped with brass storage caps fitted with PTFE ferrules and were stored at 4°C in a clean environmental free-solvent atmosphere.

The Bio-VOC sampler was also used to take the indoor air samples (room and swimming pool). THMs from indoor air were collected by pulling air through the Bio-VOC sampler using the plunger for sample to sorbent tube transfer. This process was repeated 4 times to obtain a total volume of 600 ml. After collection, the sorbent tubes were capped with brass storage caps fitted with PTFE ferrules and were stored at 4°C in a free-solvent environment air until its analysis.

THMs were then desorbed and concentrated in a thermal desorption unit equipped with a Unity Series 2 Thermal Desorber and an Ultra 50:50 Multi-tube Auto-sampler (Markes International Ltd.). They were desorbed at 300 °C for 5 min and the vapors were concentrated into a cold trap at -20 °C at 40 ml/min carrier flow (Helium 5N grade) without split. The cold trap, packed with a graphitized carbon sorbent for General Purpose by Markes International Ltd., was rapidly heated to 300 °C for 5 min with a 7.5 ml/min flow of helium and the THMs were transferred to the gas chromatograph (GC) with a split flow of 6 ml/min. Only 20% of the samples in the sorbent tube was introduced into the GC column by heating at 200°C.

The THMs were finally measured in a Gas Chromatograph 7890 (Agilent Technologies) equipped with a DB-5MS UI capillary column (60 m x 0.32 mm, 1  $\mu$ m; Agilent J&W GC Columns) at 1.5 ml/ml of carrier gas flow (Helium, 5N grade). The GC oven temperature program was hold for 10 min at 40 °C; 5 °C/min to 150 °C; hold for 10 min; 15 °C/min to 210 °C; and finally hold for 10 min. This instrument was coupled to a Mass Spectrometer 5975C Inert XL MSD with a Source in Electronic Impact Mode (Agilent Technologies). The GC-MS heat transfer line was heated at 280 °C. The MS Source and MS Quadrupol temperatures were 230° C and 150 °C, respectively. The quantitative determination was carried out in Selected Ion Monitoring Mode. Dwell time was set at 50 ms for each ion scanned. THMs concentrations were expressed as micrograms per cubic meter [15].

Correlation between the exposures and between exposure and physical activity measurements was calculated using the Spearman method (“cor( )” function from the R package “stats”).

### ***RNA isolation and microarray hybridization***

Total RNA was isolated from samples (400  $\mu$ l of whole blood and 1600  $\mu$ l of RNA later) by using RiboPure<sup>TM</sup>-Blood (Ambion), according to the manufacturer’s instructions. RNA isolated from a total of 86 samples was hybridized on Agilent 8x60K Whole Human Genome microarrays for mRNA and Sureprint G3 Human V19 miRNA 8 x 60K microarrays for microRNA. Raw data on pixel intensities were extracted using Agilent Feature Extraction Software. Four samples failed in the mRNA hybridization process.

### ***Pre-processing of mRNA and microRNA***

The normalization of the gene expression was performed using the Bioconductor ([github.com/BiGCAT-UM/arrayQC\\_Module](https://github.com/BiGCAT-UM/arrayQC_Module)), followed by selection of genes with less than 30% flagged bad spots and imputation by k-nearest neighbors (k-NN; k-value 15). The final number of probes was 24,893 transcripts.

For microRNA pre-processing, AgiMicroRna was applied [20]. Control features and microRNAs with more than 34% low intensity spots were filtered out, resulting in a total of 278 microRNAs.

### ***Quality control factor correction***

The study design includes two quality control samples per batch of microarray hybridization for both transcriptomics and microRNA analysis. These quality control samples were identical biological samples that were distributed across batches, and enable correction of batch-related influences on gene expression levels. The pre-processing of quality control samples was performed together with the study samples except for the imputation step which was omitted in the case of quality control samples. A ratio was calculated by subtracting (since expression values were  $\log_2$  transformed) the median expression of the quality control samples of one day from

the median expression of all quality control samples of the study. The study samples were corrected by addition of these ratios to the samples belonging to the same batch.

### ***Linear mixed models (LMM) and Multivariate normal model (MVN)***

Before applying the MVN approach, we removed the residual technical noise induced by microarray technology from the gene/microRNA expression data: for this, we performed a linear mixed model analysis using the “lmer( )” function from the “lme4” R package, including DBP exposures as variables of interest together with the random variables (e.g. the date of RNA isolation, date of dye labelling and date of microarray hybridization). The residuals, intercept and estimated betas of this model were estimated gene-wise using linear mixed models. Then, once the parameters from the right-hand side of the linear regression equation are known for each gene, their values were used in a second equation in order to calculate the de-noised gene/microRNA expression values. A Principal Component Analysis (PCA) of the transcriptomics data and the microRNA data was performed in order to visualize the pre- and post-swimming condition.

We ran a MVN using the “glsl( )” function from the R package “nlme” to find associations between gene/ microRNAs expression and THMs exposure. The multivariate normal distribution was parameterized using a mean vector ( $\mu$ ) and a covariance matrix ( $\Sigma$ ).  $\Sigma$  represents the variance-covariance across the two observations per subject (before and after swimming).

In the model  $Y \sim MVN(\mu, \Sigma)$ ,  $Y$  represents the gene/ microRNAs expression, where  $Y_{pre}$  refers to the expression of the samples collected before swimming and  $Y_{post}$  to the samples after swimming. The THMs were modelled individually, so  $CHCl_3$ , BDCM, DBCM,  $CHBr_3$ , and as a sum of all brominated compounds, BrTHMs ( $BrTHMs = BDCM + DBCM + CHBr_3$ ) as well as the grand total of all THMs measured, tTHM ( $tTHM = CHCl_3 + BDCM + DBCM + CHBr_3$ ). The measured exposure levels were centered around the mean exposure level before swimming. The model formulation includes the individual mean of each exposure and the increase of the exposure between the individual mean and the centered mean. The last one is the variable of interest that has been used for further analysis. In order to implement the MVN, the IDs from the different subjects were used as the grouping factor in the general correlation structure (“corSymm”), and the pre-/post-swimming condition was implemented by the constant variance function structure (“varIdent”).

In addition, we ran a model with experiment (pre- vs post-swimming) as variable of interest instead of the individual THMs in order to assess the changes in gene expression due to 40 minutes of swimming in the chlorinated swimming pool. The analysis gave the almost the same number of hits associated with pre-/post-swimming with individual exposures analysis and therefore these genes were not further analyzed.

Sensitivity analysis assessing potential confounders such as temperature and pH of water and temperature of air was performed by running models with and without such variables.

Regression coefficients and p-values were estimated from each model (individual exposures and sum of them). P-values were adjusted using Bonferroni threshold at 5%.

Due to the relatively small number of subjects in the experiment, there is a risk that some outliers drive the results. In order to avoid this potential risk, the highest and lowest expression levels were truncated 5%, 10% and 20% (capping). This sensitivity analysis was performed for both mRNA and microRNA datasets.

### ***Integration of genome-wide gene and microRNA expression data***

We generated lists of significant hits in association with DBP exposures identified in the MVN model, for transcriptomics and microRNA analysis independently. miRTarBase database was used to identify the relationships between significantly modified microRNAs and gene expressions. The database only includes microRNA-target gene interactions that are collected by manually surveying relevant literature and have been validated experimentally [21]. The significantly expressed microRNAs from the MVN model were used to identify target mRNA candidates. These mRNA candidates were exported and compared with the list of significant transcripts generated by the MVN model. These genes were displayed in Cytoscape [22] in order to visualize the interactions between up- or down-regulated mRNAs and microRNAs in association with DBP exposure. Significantly expressed genes with negative correlations between the mRNA and microRNA expressions were selected for further evaluation.

### ***Genes associated with bladder and colon cancer***

The genes that were identified to be significantly expressed in association with THM exposure were used as input for the Comparative Toxicogenomics Database in order to assess their potential role in the development of bladder and colon cancer. The search was focused on *Homo sapiens*. References with “direct evidence”, defined as “a gene that may be a biomarker of a disease or play a role in the etiology of a disease”, were selected [23].

### ***Genes associated with physical activity and THM exposure***

In order to enable correction for genes that are potentially modulated by physical activity, the biomedical mindmap “Coremine” (<https://www.coremine.com/>) was used to find connections between gene expression and physical activity using the term “motor activity (alias Physical

Activity)''' . The same search was performed for each individual DBP exposure, using the terms 'Chloroform (chemical)', 'bromodichloromethane (chemical)', 'chlorodibromomethane (alias dibromochloromethane)' and 'bromoform (chemical)'. Next, we assessed the overlap with the significant hits from the MVN by comparing both lists in a venn diagram.

### ***Pathway analysis of gene expression***

For the biological interpretation of genes deregulated in association with THM exposure, we used the lists of significant transcripts for each model after Bonferroni correction as input for pathway analysis using ConsensusPathDB [24]. A background list comprising the total list of genes from the dataset after pre-processing was used.

## **Results**

### ***DBP exposure***

Exposures to brominated compounds showed considerable inter-correlation (ranging from 0.64-0.91). Also, the correlation between exposures to chloroform and to the sum of THM was very high (0.96 and 0.99 before and after swimming respectively) (Supplementary material Table S1). All exhaled air THM levels were significantly increased after swimming (paired T-test) (Table 1). The correlation between THM exposure levels and the various parameters of physical activity was also high (0.73-0.86 for Kcal and heart rates) (Supplementary material Table S2). Since it appeared impossible to disentangle physical activity from DBP exposure due to the high correlation between both variables, the final MVN model was adjusted only for the confounders' sex, age and BMI. The water quality parameters of the swimming pool are described in Supplementary Material Table S3.

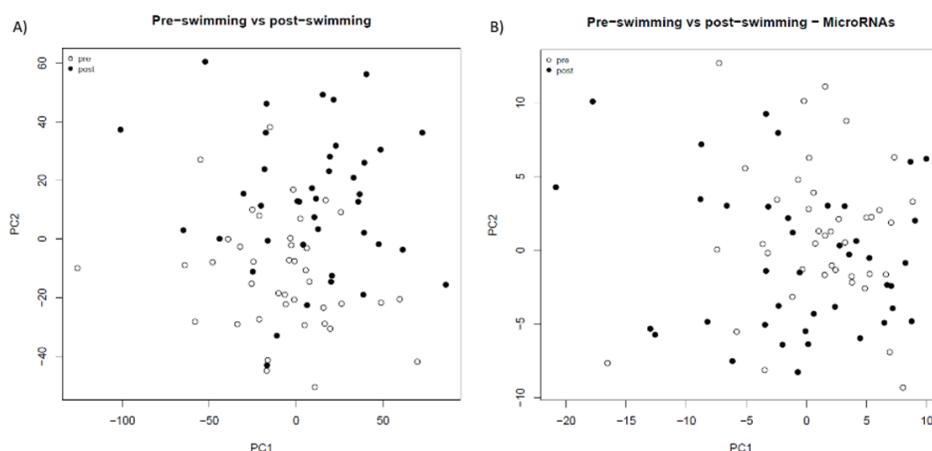
**Table 1.** *Exposure levels measured in exhaled breath before and after swimming (N=43).*

Compounds	Pre-swimming	Post-swimming	Paired t-test
	Mean $\pm$ SD ( $\mu\text{g}/\text{m}^3$ )	Mean $\pm$ SD ( $\mu\text{g}/\text{m}^3$ )	P-value

CHCl <sub>3</sub>	0.41±0.26	11.58±4.78	< 2.2e-16
CHBr <sub>3</sub>	0.03±0.02	0.11±0.07	< 2.2e-09
BDCM	0.06±0.06	2.49±1.08	< 2.2e-16
DBCM	0.02±0.03	0.53±0.28	< 2.2e-14
BrTHMs	0.12±0.10	3.13±1.41	< 2.2e-16
tTHMs	0.52±0.34	14.72±5.90	< 2.2e-16

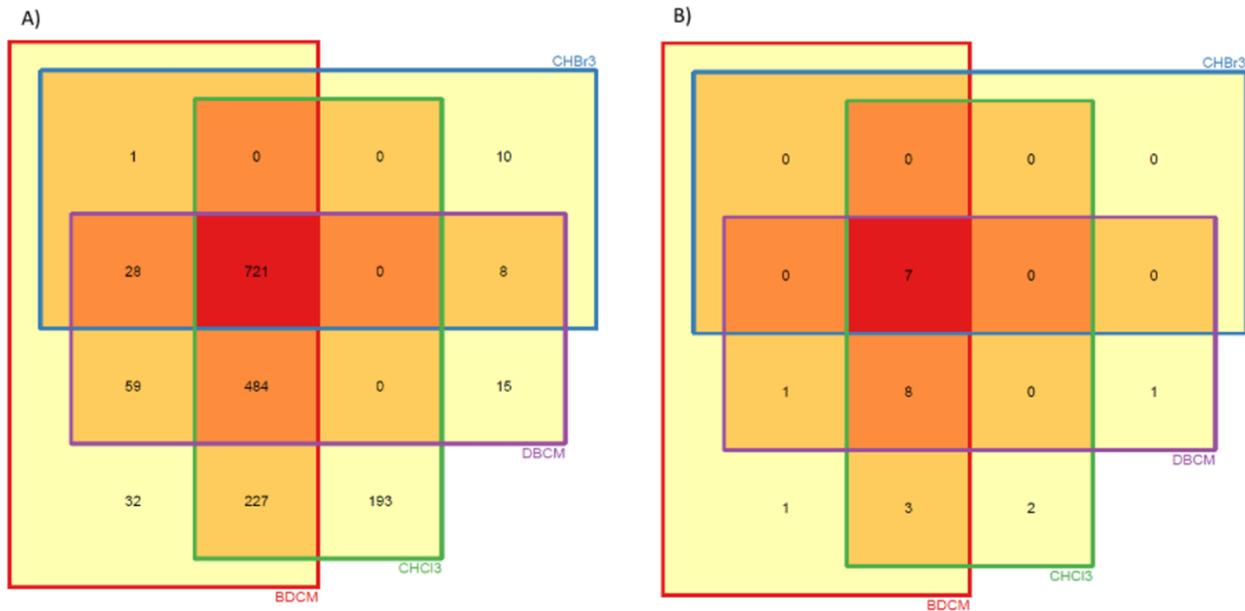
### **Changes in transcriptome and microRNA levels as a consequence of swimming**

Figure 1 A) shows a Principal Component Analysis (PCA) of the transcriptomics data (24 893 transcripts) and Figure 1 B) of the microRNA data (278 microRNAs). The expression values were corrected for batch effect using linear mixed model analysis as described in the methods section. To some degree, with respect to the transcriptome data the condition ‘pre-swimming’ (white) can be discriminated from the condition ‘post-swimming’ (black) while this is relatively less clear for the microRNA dataset.



**Figure 1.** PCA plots of the gene expression before swimming (white) and after swimming (black). A) for mRNA and B) for microRNA. N=41 in A) and N=43 in B).

The number of Bonferroni significant hits associated with each individual THM exposure using the confounder adjusted MVN models are displayed in Figure 2 (A) for transcriptomics and B) for microRNAs). A considerable number of similar hits are shared among the associations with all THM exposures (721 genes and 7 microRNAs). A relatively large number of hits are in common for brominated exposures (153 genes and 8 microRNAs). To a lesser extent, hits uniquely associated with each individual exposure were found. BrTHMs show only a few hits not common with the individual brominated compounds so it has been excluded from Figure 2 in order to make it more readable.



**Figure 2.** Venn diagram of Bonferroni significant hits, adjusted for sex, age and BMI.  $\text{CHCl}_3$  = chloroform; BDCM = bromodichloromethane; DBCM = dibromochloromethane;  $\text{Br}_3\text{CH}$  = bromoform (N=41).

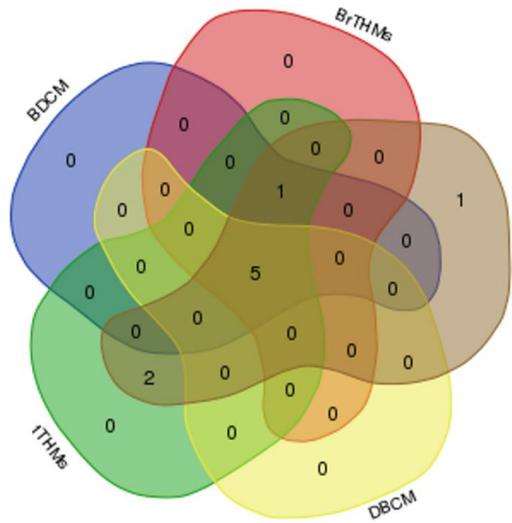
### Capping

Truncation of the omic distribution demonstrates robustness of the MVN hits for both transcriptomics and microRNA analysis. The total number of hits per capping parameter (5, 10, 20%) is described in Supplementary Material Table S4.

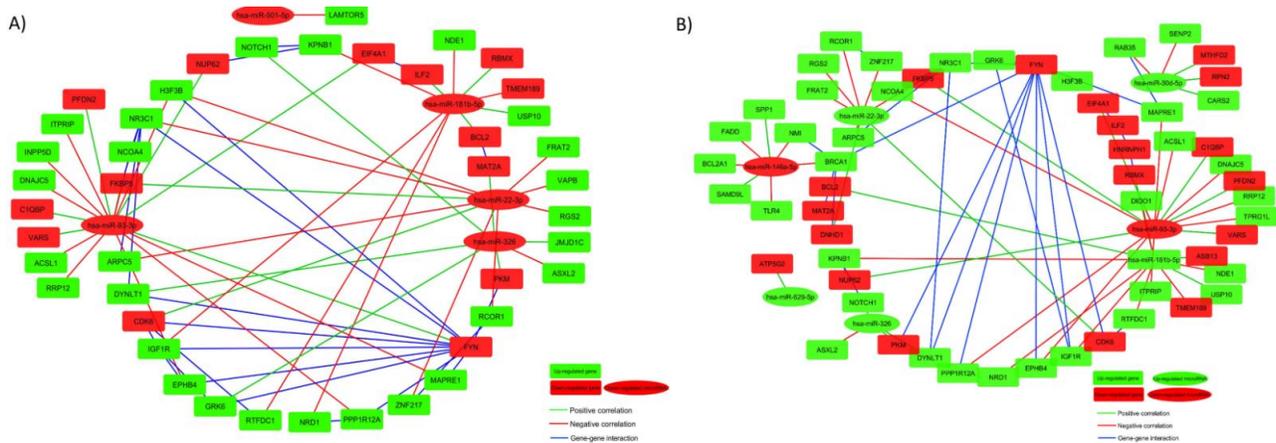
### MicroRNA/transcript integration

Figure 3 shows the numbers of Bonferroni-significant expressions associated with each individual exposure for microRNAs that target genes significantly expressed in association with THM exposures. In total 9 microRNAs out of a total of 23 microRNAs were found that were significantly associated with at least one exposure: 63 for  $\text{CHCl}_3$ , 43 for BDCM, 29 for DBCM, 43 for BrTHMs and 57 for tTHMs. Heatmaps of the correlations between Bonferroni significant transcripts and microRNAs are displayed in Supplementary Material Figure S1 to Figure S6. Around 60% of the expressed genes significantly associated with BrTHMs that are also targeted by significantly expressed microRNAs show an inverse mRNA-microRNA relationship. In the case of  $\text{CHCl}_3$  the proportion of such negative correlations is around 50%. Some of these significant genes (targeted by significant microRNAs) are involved in pathways related to cancer processes (Supplementary Material Table S5), such as apoptosis modulation and signaling, transcriptional misregulation in cancer, and  $\text{TNF-}\alpha$  pathways. Figure 4 presents the Cytoscape networks for the microRNAs and

the mRNA targets that were significantly associated with exposures to BrTHMs and to CHCl<sub>3</sub>. These networks present IGF1R, NCOA4, MAPRE1, ITPRIP, EPHB4, ACSL1, ZNF217, FRAT2, ARPC5, KPNB1 and NRD1, all up-regulated in association with DBP exposure and involved in cancer-related processes, and showing a negative correlation with the expression of the microRNAs that target them.



**Figure 3.** Number of Bonferroni-significant gene expressions associated with individual exposures (N=41).



**Figure 4.** Cytoscape network showing interactions of microRNAs and genes. Up-/down-regulation in association with exposure for BrTHMs (A) and CHCl<sub>3</sub> (B) and positive/negative correlation between mRNA-microRNA. Can be found at:

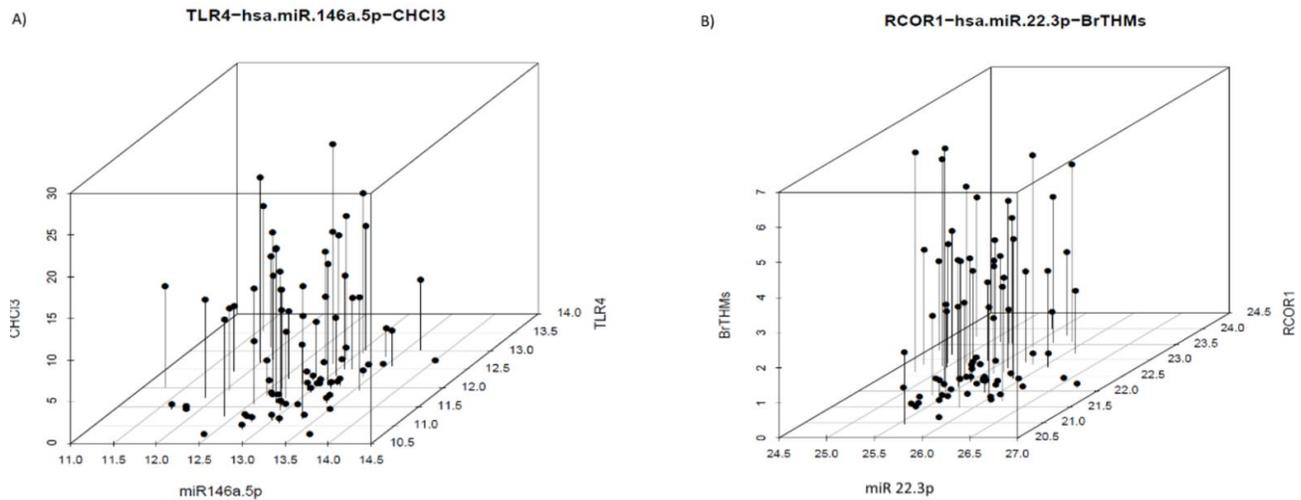
[http://web.tqx.unimaas.nl/shared\\_files/aespin/Piscina/Figure\\_4](http://web.tqx.unimaas.nl/shared_files/aespin/Piscina/Figure_4)

**Genes associated with bladder cancer and colon cancer**

Some of the significantly expressed genes in association with DBP exposures have previously been linked to bladder cancer (9 genes) and colon cancer (31 genes). A hypergeometric test shows a significant overlap for colon cancer (p-value 0.004) but not for bladder cancer (p-value 0.227). Supplementary Material Table S6 displays the regression coefficients for each exposure and presents correlation levels with their corresponding p-values for bladder and colon cancer.

The 9 significant genes also associated with bladder cancer are SLC19A1 [25], CREBBP [26], GSTP1 [27-29], STAG2 [30-32], KMT2C [26], ASXL2 [30], SOD2 [33, 34] and RASSF1 [35].

In case of colon cancer, the 31 genes which are found to be involved in cancer are involved in the RAGE-, FoxO-, HIF-1-, DNA damage- and interleukin-related pathways. Two genes deregulated in association with THM exposure, were negatively correlated with two microRNAs and at the same time regulated into the direction of colorectal cancer risk. Firstly, RCOR1 is a colon cancer-related gene that is significantly associated with BrTHMs exposure and shows a negative correlation with the microRNA hsa-miR-22-3p, also significantly associated with BrTHMs exposure (correlation gene-microRNA= -0.33, p-value 2.33E-03). Secondly, TLR4 is also related to colon cancer and is significantly associated with CHCl<sub>3</sub> exposure. It shows a negative correlation with microRNA 146a-5p, which is significantly associated with CHCl<sub>3</sub> exposure (correlation gene-microRNA= -0.08, p-value 0.50). Figure 5 shows a 3D plot with the TLR4 and RCOR1 gene expression and hsa-miR-146a-5p and hsa-miR-22-3p expression against exposure levels of CHCl<sub>3</sub> (Figure 5 A) and BrTHMs (Figure5B).



**Figure 5.** 3D plot for TLR4 gene and hsa-miR-146a-5p, both significantly associated with CHCl<sub>3</sub> (A) and RCOR1 and hsa-miR-22-3p, both significantly associated with BrTHMs exposure (B) (N=41).

**Genes associated with physical activity and THM exposure according to literature**

We found some overlap between the hits reported in literature for physical activity and the significant genes from MVN (between 17% and 19% depending on the exposure). Only significant genes associated with  $\text{CHCl}_3$  exposure showed an overlap: 32 significant genes associated with  $\text{CHCl}_3$  were found also in the literature and do not overlap with physical activity.

### ***Pathway analysis***

In view of the difficulty to separate the exposure effects from effects induced by physical activity, we deleted the genes that have been associated with physical activity in the literature (Coremine mindmap) from the input list for pathway analysis. Table 2 shows the resulting list of pathways found in association with DBP exposure, grouped according to common pathways among exposures. ConsensusPathDB analysis provided a large number of deregulated cancer-related pathways such as miR-targeted genes in squamous cell carcinoma, and various cancer-related pathways (SUMO, RHO GTPases, WNT, etc). Table 3 shows the pathways associated with chloroform exposure that were common with those from literature and did not overlap with physical activity. We found relevant pathways related to nucleotide excision repair and p73 (protein related to the p53 tumor protein). Supplementary Material Table S7 displays the full list (before excluding genes related to physical activity).

## **Discussion**

Exposure to disinfection by-products present in swimming water has been associated with adverse health outcomes such as increased risk of bladder cancer, and in order to gain improved understanding of the potential mechanisms involved in this association we evaluated the transcriptome and microRNA responses before and after swimming in a chlorinated indoor pool. After denoising the data from technical variables using linear mixed models, associations between gene/ microRNA expression levels and exposures to DBPs were established using multivariate normal models. The stringent multiple hypothesis testing method and cutoff chosen for the analysis (Bonferroni  $<0.05$ ) reduces the chance of identifying false positives. In view of the consistently reported link between bladder cancer risk and DBPs, we evaluated the established molecular responses in the context of this association. Previous publications reported the impact of THMs long-term exposure [7, 8]. In this study, we aim to assess the molecular changes occurring as a direct consequence of short-term exposure. Besides the relatively small population, significant genomics responses that induce up or down regulation in cancer-related genes have been identified.

From the total of list of genes that were significantly associated with DBPs, we identified nine genes significantly associated with DBP exposure that have previously been linked to bladder

cancer (Table S5). The protein encoded by SLC19A1 is a transporter of folate, and the expression of this gene was approximately 9-fold higher expressed in bladder tumor relative to normal bladder tissue [25]. In our study, the SLC19A1 expression was highly correlated with exposure to tTHMs (0.52, p-value 5.293e-07). CREBBP, a gene involved in the transcriptional co-activation of a large variety of transcription factors, was found to have genetic aberrations in individuals with bladder cancer [26]. According to the literature, CREBBP up-regulates hTERT activity, promoting lung cancer growth [36]. In our study we found a positive association between CREBBP and all DBP exposures (range of regression coefficients 0.05-6.37). Although this gene is linked to bladder cancer, having no information on the directionality of the gene activity makes it difficult to interpret the consequence of a positive correlation with exposure with regard to the impact on cancer risk. GSTP1 is a phase II enzyme which is involved in the detoxification of metabolites with carcinogenic properties. Down-regulation of GSTP1 has been suggested as a possible early event in prostate cancer development [37]. On the other hand, *in vitro* studies have shown an over-expression of GST in resistant human bladder cancer cells [29]. Its expression has also been found to be low in brain cancer patients [38]. GSTP1 has negative regression coefficients in our findings (ranging from -0.03 to -3.54). A down regulation of GSTP1 detoxification is generally interpreted as a carcinogenic risk in view of the reduced capacity to eliminate reactive and potentially carcinogenic metabolites. SOD2 is a member of the superoxide dismutase family. The stage and grade of bladder cancer was found to correlate positively with SOD2 gene expression levels [33]. We found a positive association between SOD2 and exposure levels (regression coefficients ranging from 0.03 to 4.36), which suggests that exposure to DBP drives SOD2 activity in the same direction as observed in bladder cancer cells. Finally, RASSF1 is a tumor suppressor gene and epigenetic silencing of this gene has been associated with advanced tumor stages in bladder cancer patients [35]. The gene shows a negative association with DBP exposure levels (regression coefficients from -0.02 to -1.88), indicating a down-regulation of the gene expression when subjects are exposed to high levels of DBPs. It should be noted that in our study we are measuring gene expression changes after short-term exposure to DBP, whereas the gene expression data reported in the literature in most cases report on basal levels in cancerous tissues and cells. Even when the observed transcriptome responses seem to go in the same direction as found in the literature, interpretation of gene expression changes in terms of increased cancer risk should be done with some reservation.

The study of microRNAs as gene regulators in cancer development has tremendously improved the diagnosis and personalized care of patients. As microRNAs are very stable and much more resistant to RNAase degradation as compared to mRNAs, these small molecules have a great potential as biomarkers of early disease [39]. In view of the important function of microRNAs in the regulation of oncogenes and tumor suppressor genes, studying co-expressed and targeted genes may further enhance the understanding of molecular mechanisms that could explain the observed link between DBP exposure and cancer risk.

In our analyses, we found that the bladder cancer associated gene ASXL2, was positively associated with DBP-exposure, whereas the microRNA targeting this gene, hsa-miR-326, was negatively associated with exposure. Hsa-miR-326 is known to exert an anticancer effect [40-45] and is found to be downregulated in primary human lung cancer and non-small lung cancer cell lines [46]. ASXL2 encodes a chromatin protein involved in transcriptional activation and silencing. Although it may promote breast cancer [47] and cervical cancer [48], the impact of increased ASXL2 expression or its mutational inactivation on bladder cancer risk is inconclusive [30, 49].

Even though our study was primarily focused on the identification of molecular changes linking exposure to DBP to bladder cancer, we also observed expression changes of genes that have previously been associated with colon cancer risk. Given the fact that the association between exposure and colon cancer risk is less consistent in literature, it is remarkable that we found even more genes linked to colon cancer as compared to bladder cancer (p-value 0.004 from a hypergeometric test). The 31 genes we identified are involved in RAGE-, FoxO-, HIF-1-, DNA damage- and interleukin-related pathways. In view of the large number of genes it is difficult to interpret the overall outcome in terms of directionality towards cancer risk. In addition to this transcriptome response, we found 4 microRNAs that were significantly associated with DPB exposure. Two of these microRNAs may target genes from the list of 31 genes associated with colon cancer. This applies for the microRNAs hsa-miR-22-3p and hsa-miR-146a-5p and their target genes RCOR1 and TLR4, respectively.

The microRNA hsa-miR-22-3p has a tumor suppressor role [50] and it was found to be down-regulated in colon cancer [51, 52], hepatocellular carcinoma [53] and osteosarcoma [54]. As shown in Table S5 B, the expression of this microRNA in association with THM exposure goes into the opposite direction, suggesting that cancer risk would be increasing with increasing exposure. Furthermore, the microRNA hsa-miR-22-3p is expressed in the opposite direction to its target gene RCOR1, an oncogene that has also been linked with colon cancer [55]. Thus, increased exposure is found to result in higher expression of RCOR1, which is again supporting a potentially increased cancer risk as a consequence of exposure to DBPs.

Hsa-miR-146a-5p acts as a tumor-suppressor gene by inhibiting proliferation of colon cancer cells [56]. The oncogene TLR4 is targeted by hsa-miR-146a-5p and both TLR4 and hsa-miR-146a-5p show a negative correlation. Increased TLR4 expression has been associated with colon cancer [57-59] and in our study, TLR4 expression increased with increasing exposure levels to all DBPs. This implies that high exposure to DBPs changes the expression of hsa-miR-146a-5p and TLR4 the same direction as observed in colon cancer.

Exposures to brominated compounds show the highest regression coefficients for genes associated with bladder and colon cancer (Supplementary Material Table S6). In addition, also the most significant pathways found in ConsensusPathDB are associated with exposure to these

compounds. These findings suggest a stronger toxic effect of brominated compounds as compared to chloroform. Previous studies have also reported an increased frequency of micronucleated lymphocytes after swimming linked to brominated compounds and not chloroform [17].

THMs is one of the most prevalent classes of DBPs, reason why it has been widely used in epidemiological studies as surrogates of total DBP content, together with validated methods of analysis from biological samples including measurements in exhaled air [2, 60]. In view of the large number of DBP species, THMs represent however only a small but relevant subgroup. The genomics responses observed in this study are identified in correlation analysis with THMs exposure and therefore, these findings cannot be translated directly to other DBPs with important toxicological effects [4]. Nevertheless, increased levels of BrTHMs in exhaled breath were significantly correlated to total organic carbon (Spearman correlation=-0.22) and some DBPs from the pool water such as BrTHMs (0.48), THAAs (-0.34),  $\text{CHCl}_2\text{CN}$  (-0.40),  $\text{C}_3\text{H}_3\text{Cl}_3\text{O}$  (-0.42), chloramines, total organic chlorine and total organic bromine [15]. This demonstrates the potential of BrTHMs in exhaled breath as a valid biomarker of DBPs exposure measurements in swimming pools containing bromide source water. Regarding the contribution of DBPs to cancer risk, further analysis including a wider range of DPBs species should be carried out in order to provide a more specified assessment.

Separating DBP exposure effects from those induced by physical activity represents a challenge in this experimental design. As 59% of the genes that we identified have also been associated with physical activity, we removed all physical activity-related genes reported in literature. In view of the large number of hits remaining after this removal and the relatively large number of cancer-related pathways generated from this corrected gene list, we conclude that these findings reflect the response to DPB exposure rather than to physical activity.

## Conclusions

In conclusion, the results of this study show that a short period of swimming in a chlorinated pool can already produce changes in gene and microRNA expression. A total of 1,778 unique transcripts were significantly associated with exposures to  $\text{CHCl}_3$ , BDCM, DBCM,  $\text{CHBr}_3$  and the sum of BrTHMs and THMs. The concomitant analysis of microRNAs resulted in the identification of 23 unique microRNAs associated with the same exposures. Nine significantly expressed genes could be linked to bladder cancer risk and 31 to colon cancer risk. Both microRNA hsa-miR-22-3p and its target gene RCOR1 were significantly related to DBP exposure, and regulated into the same direction as in colon cancer. The same applies to the significantly regulated hsa-miR-146a-5p and its target gene TLR4.

This study thus shows that gene expression and miRNA changes induced by short-term exposure to low levels of DBPs in a chlorinated pool may be indicative for increased risk of colon and bladder cancer.

## **Acknowledgements**

This work was supported by the EXPOsOMICS European FP7 research projects (Grants number 308610).

## **Supplementary material**

**Table S1.** *Correlation between the different THMs in exhaled breath, before and after swimming (Spearman method), N=43.*

Pre-swimming							Post-swimming						
Compound	CHCl <sub>3</sub>	BDCM	DBCM	CHBr <sub>3</sub>	BrTHMs	tTHMs	Compound	CHCl <sub>3</sub>	BDCM	DBCM	CHBr <sub>3</sub>	BrTHMs	tTHMs
CHCl <sub>3</sub>	1.00						CHCl <sub>3</sub>	1.00					
BDCM	0.57	1.00					BDCM	0.79	1.00				
DBCM	0.22	0.67	1.00				DBCM	0.53	0.91	1.00			
CHBr <sub>3</sub>	0.27	0.70	0.64	1.00			CHBr <sub>3</sub>	0.37	0.75	0.91	1.00		
BrTHMs	0.46	0.95	0.81	0.83	1.00		BrTHM	0.71	0.99	0.96	0.81	1.00	
tTHMs	0.96	0.73	0.40	0.46	0.66	1.00	tTHM	0.99	0.86	0.62	0.45	0.79	1.00

**Table S2.** Correlation between the different DBP compounds and the different physical activity measurements (Spearman method), N=43. A) Before and after swimming. B) After swimming.

A)

Variables	CHCl <sub>3</sub>	BDCM	DBCM	CHBr <sub>3</sub>	BrTHMs	tTHMs	Kcal	HR max 69	Breathless	Fatigue	Swim time	Distance
CHCl <sub>3</sub>	1.00											
BDCM	0.92	1.00										
DBCM	0.84	0.95	1.00									
CHBr <sub>3</sub>	0.78	0.90	0.92	1.00								
BrTHMs	0.90	0.99	0.97	0.93	1.00							
tTHMs	0.99	0.95	0.88	0.82	0.93	1.00						
Kcal	0.86	0.86	0.84	0.76	0.85	0.86	1.00					
HR max 69	0.83	0.84	0.83	0.73	0.84	0.84	0.87	1.00				
Breathless	0.62	0.66	0.64	0.59	0.65	0.64	0.54	0.70	1.00			
Fatigue	0.71	0.74	0.74	0.68	0.74	0.72	0.74	0.77	0.68	1.00		
Swim time	0.81	0.83	0.83	0.75	0.83	0.81	0.93	0.85	0.52	0.68	1.00	
Distance	0.86	0.86	0.85	0.76	0.85	0.86	0.98	0.87	0.56	0.74	0.94	1.00

B)

Variables	CHCl <sub>3</sub>	BDCM	DBCM	CHBr <sub>3</sub>	BrTHMs	tTHMs	Kcal	HR max 69	Breathless	Fatigue	Swim time	Distance
CHCl <sub>3</sub>	1.00											
BDCM	0.79	1.00										
DBCM	0.53	0.91	1.00									
CHBr <sub>3</sub>	0.37	0.75	0.91	1.00								
BrTHMs	0.71	0.99	0.96	0.81	1.00							
tTHMs	0.99	0.86	0.62	0.45	0.79	1.00						
Kcal	0.47	0.40	0.32	0.19	0.36	0.47	1.00					
HR max 69	0.38	0.41	0.34	0.16	0.39	0.41	0.21	1.00				
Breathless	0.33	0.46	0.41	0.42	0.44	0.38	-0.01	0.48	1.00			
Fatigue	0.13	0.22	0.25	0.23	0.22	0.17	0.07	0.29	0.42	1.00		
Swim time	0.05	0.17	0.23	0.16	0.18	0.06	0.48	0.08	-0.11	-0.20	1.00	
Distance	0.46	0.44	0.35	0.21	0.40	0.47	0.88	0.23	0.01	0.04	0.61	1.00

**Table S3.** Characteristics and exposure levels measured in swimming pool water

Swimming pool water		
Compound family	Compound	Mean±SD
Trihalomethanes (µg/L)	Chloroform (CHCl <sub>3</sub> )	37.76±7.53

	Bromodichloromethane (BDCM)	8.17±2.34
	Dibromochloromethane (DBCM)	2.81±1.06
	Bromoform (CHBr <sub>3</sub> )	1.07±0.54
	Brominated THMs (BrTHM)	10.88±5.15
	Total THMs (TTHM)	49.82±9.14
Haloacetic acids (µg/L)	Dichloroacetic acid (DCIAA)	29.67±10.25
	Trichloroacetic acid (TCIAA)	59.95±11.11
	Bromochloroacetic acid (BrCIAA)	5.1±1.68
	Dibromoacetic acid (DBrAA)	1.58±0.71
	Dibromochloroacetic acid (DCIBrAA)	12.22±5.3
Haloacetonitriles (µg/L)	Dichloroacetonitrile (C <sub>2</sub> HCl <sub>2</sub> N)	7.03±1.95
	Bromochloroacetonitrile (CHBrClCN)	3.73±0.76
Chloramines (mg/L)	Monochloramine (NH <sub>2</sub> Cl)	0.31±0.24
	Dichloramine (NHCl <sub>2</sub> )	0.25±0.2
	Trichloramine (NCl <sub>3</sub> )	0.11±0.3
Others	Haloketone (µg/L) (C <sub>3</sub> H <sub>3</sub> Cl <sub>3</sub> O)	2.31±0.85
	Total organic carbon (mg/L) (NPOC)	2.76±1.35
	Free chlorine (mg/L) (FCI)	1.33±0.38
	Total Organic Chlorine (TOCl)	0.07±0.01
	Total Organic Halides (TOX)	0.47±0.03
	N-nitrosodimethylamine (NDMA)	10.84±1.6
	pH	7.54±0.24
	Temperature	27.89±0.31

**Table S4.** Sensitivity analysis for capping. Bonf=Bonferroni correction, N=41 in A) and N=43 in B).

A) Transcriptomics

Exposure	All transcripts	Capping 95%-5%	Capping 90%-10%	Capping 80%-20%	Full vs 95%-5%	Full vs 90%-10%	Full vs 80%-20%
	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05
CHCl <sub>3</sub>	1625	1471	1185	606	1349	1069	537
BDCM	1552	1387	1065	495	1272	976	444
DBCM	1315	1138	865	403	1037	773	336
CHBr <sub>3</sub>	768	595	430	237	510	316	136
BrTHMs	1517	1334	1034	474	1207	948	421
tTHMs	1657	1492	1191	608	1350	1081	538

## B) MicroRNA

Exposure	All transcripts	Capping 95%-5%	Capping 90%-10%	Capping 80%-20%	Full vs 95%-5%	Full vs 90%-10%	Full vs 80%-20%
	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05	Bonf<0.05
CHCl <sub>3</sub>	20	17	12	7	15	11	6
BDCM	20	18	11	4	15	8	2
DBCM	17	12	8	3	10	6	2
CHBr <sub>3</sub>	7	6	2	2	5	2	1
BrTHMs	19	17	9	4	14	6	2
tTHMs	19	18	13	7	14	11	5

**Figure S1.** Figure S1. Heatmap showing correlations between Bonferroni significant transcripts and microRNAs for CHCl<sub>3</sub>. Can be found at:

[http://web.tqx.unimaas.nl/shared\\_files/aespin/Piscina/Figure\\_S1.docx](http://web.tqx.unimaas.nl/shared_files/aespin/Piscina/Figure_S1.docx)

**Figure S2.** Figure S2. Heatmap showing correlations between Bonferroni significant transcripts and microRNAs for BDCM. Can be found at:

[http://web.tqx.unimaas.nl/shared\\_files/aespin/Piscina/Figure\\_S2.docx](http://web.tqx.unimaas.nl/shared_files/aespin/Piscina/Figure_S2.docx)

**Figure S3.** Figure S3. Heatmap showing correlations between Bonferroni significant transcripts and microRNAs for DBCM. Can be found at:

[http://web.tqx.unimaas.nl/shared\\_files/aespin/Piscina/Figure\\_S3.docx](http://web.tqx.unimaas.nl/shared_files/aespin/Piscina/Figure_S3.docx)

**Figure S4.** Figure S4. Heatmap showing correlations between Bonferroni significant transcripts and microRNAs for CHBr<sub>3</sub>. Can be found at:

[http://web.tqx.unimaas.nl/shared\\_files/aespin/Piscina/Figure\\_S4.docx](http://web.tqx.unimaas.nl/shared_files/aespin/Piscina/Figure_S4.docx)

**Figure S5.** Figure S5. Heatmap showing correlations between Bonferroni significant transcripts and microRNAs for tTHMs. Can be found at:

[http://web.tqx.unimaas.nl/shared\\_files/aespin/Piscina/Figure\\_S5.docx](http://web.tqx.unimaas.nl/shared_files/aespin/Piscina/Figure_S5.docx)

**Figure S6.** Figure S6. Heatmap showing correlations between Bonferroni significant transcripts and microRNAs for BrTHMs. Can be found at:

[http://web.tqx.unimaas.nl/shared\\_files/aespin/Piscina/Figure\\_S6.docx](http://web.tqx.unimaas.nl/shared_files/aespin/Piscina/Figure_S6.docx)

**Table S5.** Full list of pathway from significant genes targeted by significant microRNAs using ConsensusPathDB.

q-value	pathway
0.045747	il-7 signal transduction
0.045747	nf-kb signaling pathway
0.045747	Apoptosis Modulation and Signaling
0.045747	Transcriptional misregulation in cancer - Homo sapiens
0.045747	IKK complex recruitment mediated by RIP1
0.045747	FSH
0.045747	Prion diseases - Homo sapiens
0.045747	miR-targeted genes in lymphocytes - TarBase
0.045747	TNFalpha
0.045747	Measles - Homo sapiens
0.045747	TRIF-mediated programmed cell death
0.045747	TLR NFkB
0.045747	Hepatitis B - Homo sapiens
0.045747	Signaling Pathways in Glioblastoma
0.045747	miR-targeted genes in epithelium - TarBase
0.047017	IL2
0.047811	EPHB-mediated forward signaling
0.047811	TRAF6 mediated induction of TAK1 complex
0.047811	Toll-like receptor signaling pathway - Homo sapiens
0.047811	CXCR4-mediated signaling events
0.047811	NF-kappa B signaling pathway - Homo sapiens
0.047811	PECAM1 interactions
0.047811	IRAK1 recruits IKK complex upon TLR7/8 or 9 stimulation
0.047811	mechanism of protein import into the nucleus
0.047811	Downstream TCR signaling
0.047811	TRAF6 mediated IRF7 activation in TLR7/8 or 9 signaling
0.047811	TRIF-mediated TLR3/TLR4 signaling
0.047811	MyD88-independent TLR3/TLR4 cascade
0.047811	Toll Like Receptor 3 (TLR3) Cascade
0.047811	B Cell Receptor Signaling Pathway
0.047811	Pathways in cancer - Homo sapiens
0.047811	Focal adhesion - Homo sapiens
0.047811	IL3
0.047811	Nucleotide-binding domain
0.047811	miR-targeted genes in muscle cell - TarBase
0.047811	Pathogenic Escherichia coli infection - Homo sapiens
0.047811	Ligand-dependent caspase activation

0.048355	Apoptosis
0.051554	Ephrin signaling
0.051554	Programmed Cell Death
0.051665	Influenza Life Cycle
0.051665	Endocytosis - Homo sapiens
0.051926	Toxoplasmosis - Homo sapiens
0.051926	Kit receptor signaling pathway
0.051926	Leptin
0.051926	Activated TLR4 signalling
0.051926	PI3K-Akt signaling pathway - Homo sapiens
0.054684	Apoptosis Modulation by HSP70
0.055185	Influenza Infection
0.055185	regulation of bad phosphorylation
0.055185	EPH-ephrin mediated repulsion of cells
0.055185	Primary Focal Segmental Glomerulosclerosis FSGS
0.055185	EPH-Ephrin signaling
0.055185	Coregulation of Androgen receptor activity
0.055185	TCR signaling
0.055185	Methionine and cysteine metabolism
0.055185	Regulation of toll-like receptor signaling pathway
0.055185	Toll Like Receptor 4 (TLR4) Cascade
0.055879	IGF-Core

**Table S6. Regression coefficients and correlation levels for genes associated with bladder cancer (A) and colon cancer (B) from literature and the targeted microRNA.**

A)

Features	CHCl <sub>3</sub>	BDCM	DBCM	CHBr <sub>3</sub>	BrTHMs	tTHMs	Correlation tTHMs	p-value	Correlation BrTHMs	p-value
SLC19A1	0.051	0.228	1.002	4.858	0.180	0.040	0.521	5.29E-07	0.521	5.29E-07
CREBBP	0.062	0.285	1.293	6.367	0.227	0.049	0.509	1.06E-06	0.509	1.06E-06
GSTP1	-0.033	-0.146	-0.665	-3.530	-0.117	-0.026	-0.352	1.20E-03	-0.352	1.20E-03
STAG2	0.053	0.254	1.170	5.858	0.203	0.043	0.441	3.43E-05	0.441	3.43E-05
KMT2C	0.037	0.166	0.751	-	0.132	0.029	0.426	6.62E-05	0.426	6.62E-05
ASXL2	0.035	0.162	0.748	-	0.130	0.028	0.472	7.41E-06	0.472	7.41E-06
SOD2	0.042	0.196	0.898	-	0.157	0.034	0.377	4.90E-04	0.377	4.90E-04
RASSF1	-0.019	-0.087	-	-	-0.069	-0.015	-0.364	7.65E-04	-0.364	7.65E-04
POR	-	0.152	0.723	-	0.123	-	-	-	0.397	2.22E-04
hsa-miR-326	-0.034	-0.139	-0.558	-2.598	-0.108	-0.026	-0.251	2.26E-02	0.251	2.30E-02

B)

Features	CHCl <sub>3</sub>	BDCM	DBCM	CHBr <sub>3</sub>	BrTHMs	tTHMs	Correlation tTHMs	p-value	Correlation BrTHMs	p-value
ABCA1	0.027	0.119	0.526	2.587	0.094	0.021	0.500	1.78E-06	0.500	1.78E-06
ABCA1	0.053	0.246	1.112	5.599	0.196	0.042	0.516	6.85E-07	0.516	6.85E-07
ABCC5	0.040	0.184	0.845	7.235	0.147	0.032	0.431	5.28E-05	0.431	5.28E-05
ABCC5	0.067	0.310	1.415	6.401	0.247	0.054	0.492	2.61E-06	0.492	2.61E-06
ABCG1	0.066	0.293	1.299	6.108	0.232	0.052	0.571	2.06E-08	0.571	2.06E-08
ALOX5	0.059	0.269	1.216	-	0.214	0.047	0.481	4.88E-06	0.481	4.88E-06
AMACR	0.054	0.248	1.116	-	0.197	0.043	0.380	4.30E-04	0.380	4.30E-04
ATP7A	0.042	0.195	0.875	4.450	0.155	0.034	0.474	6.75E-06	0.474	6.75E-06
BCL2	-0.036	-0.157	-0.672	-3.397	-0.123	-0.028	-0.348	1.34E-03	-0.348	1.34E-03
CALR	0.034	-0.122	-0.558	-2.777	-0.097	-0.020	-0.413	1.13E-04	-0.413	1.13E-04
CDKN1B	-0.043	0.160	0.725	-	0.128	0.028	0.399	2.05E-04	0.399	2.05E-04
FADS1	-0.029	-0.187	-	-	-0.148	-0.034	-0.401	1.92E-04	-0.401	1.92E-04
FZD2	0.065	-0.129	-	-	-0.102	-0.023	-0.323	3.08E-03	-0.323	3.08E-03
IL1RN	0.064	0.296	1.316	6.420	0.235	0.052	0.547	1.08E-07	0.547	1.08E-07
JUN	0.041	0.292	1.300	6.620	0.231	0.051	0.491	2.81E-06	0.491	2.81E-06
KDM5B	0.064	0.180	0.788	3.782	0.142	0.032	0.476	6.05E-06	0.476	6.05E-06
LRG1	0.082	0.293	1.327	6.650	0.233	0.051	0.530	3.04E-07	0.530	3.04E-07
MMP9	0.071	0.368	1.627	8.033	0.291	0.065	0.474	6.85E-06	0.474	6.85E-06
NAMPT	0.053	0.336	1.541	7.684	0.268	0.057	0.495	2.24E-06	0.495	2.24E-06
NOTCH1	0.053	0.234	1.036	-	0.185	0.042	0.525	4.11E-07	0.525	4.11E-07
RCOR1	0.019	0.252	1.143	5.601	0.201	0.043	0.424	7.19E-05	0.424	7.19E-05
RNF34	-0.081	0.085	-	-	0.067	0.015	0.383	3.80E-04	0.383	3.80E-04
SH3TC1	0.051	-0.365	-1.623	-8.284	-0.290	-0.064	-0.513	8.52E-07	-0.513	8.52E-07
SLC19A1	0.033	0.228	1.002	4.858	0.180	0.040	0.521	5.29E-07	0.521	5.29E-07
SLC22A1	0.052	0.150	0.650	-	0.118	0.026	0.445	2.82E-05	0.445	2.82E-05
SLC22A15	0.042	0.242	1.089	5.435	0.192	0.042	0.394	2.51E-04	0.394	2.51E-04
SOD2	0.052	0.196	0.898	-	0.157	0.034	0.377	4.90E-04	0.377	4.90E-04
STAT3	0.030	0.228	-	-	0.180	0.041	0.506	1.24E-06	0.506	1.24E-06
TGFBR2	0.046	0.138	0.616	-	0.109	0.024	0.365	7.56E-04	0.365	7.56E-04
TLR4	0.052	0.242	1.107	-	0.193	0.042	0.409	1.35E-04	0.409	1.35E-04
TNFSF10	0.040	0.192	0.877	-	0.153	0.033	0.395	2.40E-04	0.395	2.40E-04
TP53TG1	-0.029	-0.133	-0.590	-2.993	-0.106	-0.023	-0.404	1.67E-04	-0.404	1.67E-04
ZMIZ1	0.033	0.152	0.681	-	0.121	0.027	0.405	1.59E-04	0.405	1.59E-04
hsa-miR-326	-0.034	-0.139	-	-	-0.108	-0.026	-0.271	1.38E-02	-0.259	1.87E-02
hsa-miR-181b-5p	-0.029	-0.125	-0.525	-	-0.098	-0.023	-0.273	1.32E-02	-0.288	8.80E-03
hsa-miR-146a-5p	-0.030	-0.127	-	-	-	-	-0.310	4.63E-03	-0.336	2.01E-03
hsa-miR-22-3p	-0.020	-0.083	-0.342	-	-0.065	-0.015	-0.317	3.69E-03	-0.322	3.20E-03

**Table S7.** Full list of pathway from Bonferroni significant hits using ConsensusPathDB.

Names	Pathways	
BDCM BrTHMs CHBr <sub>3</sub> CHCl <sub>3</sub> DBCM tTHMs	Validated targets of C-MYC transcriptional repression	
	FAS pathway and Stress induction of HSP regulation	
	mapkinase signaling pathway	
	Apoptosis Modulation and Signaling	
	Osteoclast differentiation - Homo sapiens	
	Direct p53 effectors	
	Influenza A - Homo sapiens	
	Insulin Signaling	
BDCM BrTHMs CHCl <sub>3</sub> DBCM tTHMs	JAK STAT pathway and regulation	
	TNF alpha Signaling Pathway	
	CXCR4-mediated signaling events	
	Transcriptional misregulation in cancer - Homo sapiens	
	transcription regulation by methyltransferase of carm1	
	Integrated Cancer pathway	
	Apoptosis Modulation by HSP70	
	Fas	
	Apoptosis	
	Natural killer cell mediated cytotoxicity - Homo sapiens	
	TRAIL signaling pathway	
	miR-targeted genes in epithelium - TarBase	
	Caspase activation via extrinsic apoptotic signalig pathway	
	IL6	
	Interferon type I signaling pathways	
Fc-epsilon receptor I signaling in mast cells		
Hepatitis B - Homo sapiens		
IL-3 Signaling Pathway		
ceramide signaling pathway		
IL2		
BDCM BrTHMs CHBr <sub>3</sub> CHCl <sub>3</sub> tTHMs	Epithelial cell signaling in Helicobacter pylori infection - Homo sapiens	
	BDCM BrTHMs CHBr <sub>3</sub> DBCM tTHMs	Meiosis
		Disease
	BDCM BrTHMs	HIF-1 signaling pathway - Homo sapiens

CHCl <sub>3</sub> DBCM	
BDCM BrTHMs CHCl <sub>3</sub> tTHMs	miR-targeted genes in squamous cell - TarBase
	RHO GTPases Activate NADPH Oxidases
	role of mitochondria in apoptotic signaling
	EGFR1
	Processing and activation of SUMO
	RAC1 signaling pathway
BDCM BrTHMs CHBr <sub>3</sub> DBCM	keratinocyte differentiation
	Regulation of toll-like receptor signaling pathway
	NOTCH1 Intracellular Domain Regulates Transcription
	Fc epsilon RI signaling pathway - Homo sapiens
BDCM BrTHMs DBCM tTHMs	Oxidative Stress Induced Senescence
	Tuberculosis - Homo sapiens
	Toxoplasmosis - Homo sapiens
	miR-targeted genes in lymphocytes - TarBase
BDCM BrTHMs CHCl <sub>3</sub>	HIV-1 Nef: Negative effector of Fas and TNF-alpha
	Signaling by Rho GTPases
	hiv-1 nef: negative effector of fas and tnf
	Vitamin E metabolism
BDCM CHCl <sub>3</sub> tTHMs	IL4
	Coregulation of Androgen receptor activity
	JAK STAT MolecularVariation 2
BDCM BrTHMs DBCM	Cellular Senescence
	Hemostasis
	Toll-like receptor signaling pathway
	Toll-like receptor signaling pathway - Homo sapiens
	p38 MAPK signaling pathway
	Cellular responses to stress
	MAPK Cascade
	links between pyk2 and map kinases
	Senescence-Associated Secretory Phenotype (SASP)
	Chagas disease (American trypanosomiasis) - Homo sapiens
	Amyotrophic lateral sclerosis (ALS) - Homo sapiens
	Factors involved in megakaryocyte development and platelet production
	Angiogenesis overview
EGF-Core	
BCR	
BDCM BrTHMs tTHMs	Endocytosis - Homo sapiens
	IL5

BrTHMs CHBr <sub>3</sub> DBCM	ATF6-alpha activates chaperone genes
BrTHMs DBCM tTHMs	B Cell Receptor Signaling Pathway
BDCM CHCl <sub>3</sub>	Nanomaterial induced apoptosis
CHCl <sub>3</sub> DBCM	Programmed Cell Death
CHCl <sub>3</sub> tTHMs	Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha)
	mechanism of gene regulation by peroxisome proliferators via ppara
	Steroid biosynthesis - Homo sapiens
	miR-targeted genes in adipocytes - TarBase
BDCM BrTHMs	Innate Immune System
	Regulation of p38-alpha and p38-beta
	CDC42 signaling events
	Nucleotide-binding domain
	Kit receptor signaling pathway
	tnf/stress related signaling
BDCM tTHMs	Regulation of Androgen receptor activity
CHBr <sub>3</sub> DBCM	Glucocorticoid Receptor Pathway
BrTHMs DBCM	Measles - Homo sapiens
	TSH signaling pathway
	Mitochondrial Beta-Oxidation of Long Chain Saturated Fatty Acids
	Class IB PI3K non-lipid kinase events
	Pathways in cancer - Homo sapiens
	EGF
	repression of WNT target genes
	Non-small cell lung cancer - Homo sapiens
	Nicotinate and Nicotinamide Metabolism
	Leishmaniasis - Homo sapiens
BDCM tTHMs	TCR
CHCl <sub>3</sub>	Regulation of actin dynamics for phagocytic cup formation
BDCM	formation of the beta-catenin:TCF transactivating complex
	RHO GTPases activate PKNs
	Retinoic acid receptors-mediated signaling
	p38 mapk signaling pathway
	TNFalpha

	EGF-EGFR Signaling Pathway
DBCM	Vitamin B3 (nicotinate and nicotinamide) metabolism
	IL8- and CXCR2-mediated signaling events
	IL3
	rac1 cell motility signaling pathway
	degradation of the rar and rxr by the proteasome
	Regulation of KIT signaling
	fc epsilon receptor i signaling in mast cells
	miR-targeted genes in muscle cell - TarBase
	nf-kb signaling pathway
	Ligand-dependent caspase activation
	Chronic myeloid leukemia - Homo sapiens
CHBr <sub>3</sub>	Aurora C signaling
	Signaling events mediated by focal adhesion kinase
	HDMs demethylate histones
	Synthesis of PG
	Sialic acid metabolism
	Nuclear Receptors Meta-Pathway
	Degradation of beta-catenin by the destruction complex
BrTHMs	miR-targeted genes in leukocytes - TarBase
	S1P2 pathway
	Regulation of RAC1 activity
	Leukocyte transendothelial migration - Homo sapiens
tTHMs	Alzheimer
	JAK STAT MolecularVariation 1
	Oxidative phosphorylation
	Jak-STAT signaling pathway - Homo sapiens
	EPO Receptor Signaling

## References

1. Nieuwenhuijsen, M.J., et al., *Chlorination disinfection by-products in drinking water and congenital anomalies: review and meta-analyses*. Environ Health Perspect, 2009. **117**(10): p. 1486-93.
2. Villanueva, C.M., et al., *Overview of Disinfection By-products and Associated Health Effects*. Curr Environ Health Rep, 2015. **2**(1): p. 107-15.
3. Villanueva, C.M. and L. Font-Ribera, *Health impact of disinfection by-products in swimming pools*. Annali dell'Istituto superiore di sanita, 2012. **48**(4): p. 387-96.
4. Zwiener, C., et al., *Drowning in disinfection byproducts? Assessing swimming pool water*. Environ Sci Technol, 2007. **41**(2): p. 363-72.
5. Xu, X. and C.P. Weisel, *Dermal uptake of chloroform and halo ketones during bathing*. J Expo Anal Environ Epidemiol, 2005. **15**(4): p. 289-96.
6. Nieuwenhuijsen, M.J., et al., *The epidemiology and possible mechanisms of disinfection by-products in drinking water*. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 2009. **367**(1904): p. 4043-76.
7. Villanueva, C.M., et al., *Disinfection byproducts and bladder cancer: a pooled analysis*. Epidemiology, 2004. **15**(3): p. 357-67.
8. Villanueva, C.M., et al., *Bladder cancer and exposure to water disinfection by-products through ingestion, bathing, showering, and swimming in pools*. Am J Epidemiol, 2007. **165**(2): p. 148-56.
9. Ferlay J, S.I., Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F., *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. Lyon, France: International Agency for Research on Cancer. Available from: <http://globocan.iarc.fr>, accessed on 14/09/2016.
- , 2013.
10. Rahman, M.B., et al., *Disinfection by-products in drinking water and colorectal cancer: a meta-analysis*. International journal of epidemiology, 2010. **39**(3): p. 733-45.
11. Bove, G.E., Jr., P.A. Rogerson, and J.E. Vena, *Case control study of the geographic variability of exposure to disinfectant byproducts and risk for rectal cancer*. International journal of health geographics, 2007. **6**: p. 18.
12. Villanueva, C.M., et al., *Colorectal Cancer and Long-Term Exposure to Trihalomethanes in Drinking Water: A Multicenter Case-Control Study in Spain and Italy*. Environ Health Perspect, 2017. **125**(1): p. 56-65.
13. Grellier, J., et al., *Assessing the human health impacts of exposure to disinfection by-products--a critical review of concepts and methods*. Environ Int, 2015. **78**: p. 61-81.
14. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-97.
15. Font-Ribera, L., et al., *Environmental and personal determinants of the uptake of disinfection by-products during swimming*. Environ Res, 2016. **149**: p. 206-15.
16. Font-Ribera, L., et al., *Short-term changes in respiratory biomarkers after swimming in a chlorinated pool*. Environ Health Perspect, 2010. **118**(11): p. 1538-44.
17. Kogevinas, M., et al., *Genotoxic effects in swimmers exposed to disinfection by-products in indoor swimming pools*. Environ Health Perspect, 2010. **118**(11): p. 1531-7.
18. Kubesch, N.J., et al., *Respiratory and inflammatory responses to short-term exposure to traffic-related air pollution with and without moderate physical activity*. Occup Environ Med, 2015. **72**(4): p. 284-93.

19. Gulati, M., et al., *Heart rate response to exercise stress testing in asymptomatic women: the st. James women take heart project*. *Circulation*, 2010. **122**(2): p. 130-7.
20. Lopez-Romero, P., *Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library*. *BMC Genomics*, 2011. **12**: p. 64.
21. Chou, C.H., et al., *miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database*. *Nucleic Acids Res*, 2016. **44**(D1): p. D239-47.
22. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape*. *Nature Protocols*, 2007. **2**(10): p. 2366-2382.
23. Davis, A.P., et al., *The Comparative Toxicogenomics Database's 10th year anniversary: update 2015*. *Nucleic Acids Res*, 2015. **43**(Database issue): p. D914-20.
24. Kamburov, A., et al., *The ConsensusPathDB interaction database: 2013 update*. *Nucleic Acids Res*, 2013. **41**(Database issue): p. D793-800.
25. Abdel-Haleem, A.M., et al., *Expression of RFC/SLC19A1 is associated with tumor type in bladder cancer patients*. *PloS one*, 2011. **6**(7): p. e21820.
26. Gui, Y., et al., *Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder*. *Nature genetics*, 2011. **43**(9): p. 875-8.
27. Lesseur, C., et al., *A case-control study of polymorphisms in xenobiotic and arsenic metabolism genes and arsenic-related bladder cancer in New Hampshire*. *Toxicology letters*, 2012. **210**(1): p. 100-6.
28. Ma, Q.W., et al., *Polymorphism of glutathione S-transferase T1, M1 and P1 genes in a Shanghai population: patients with occupational or non-occupational bladder cancer*. *Biomedical and environmental sciences : BES*, 2002. **15**(3): p. 253-60.
29. Singh, S.V., et al., *Mechanism of cross-resistance to cisplatin in a mitomycin C-resistant human bladder cancer cell line*. *International journal of cancer*, 1995. **61**(3): p. 431-6.
30. Balbas-Martinez, C., et al., *Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy*. *Nature genetics*, 2013. **45**(12): p. 1464-9.
31. Guo, G., et al., *Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation*. *Nature genetics*, 2013. **45**(12): p. 1459-63.
32. Solomon, D.A., et al., *Frequent truncating mutations of STAG2 in bladder cancer*. *Nature genetics*, 2013. **45**(12): p. 1428-30.
33. Hempel, N., et al., *Altered redox status accompanies progression to metastatic human bladder cancer*. *Free radical biology & medicine*, 2009. **46**(1): p. 42-50.
34. Hung, R.J., et al., *Genetic polymorphisms of MPO, COMT, MnSOD, NQO1, interactions with environmental exposures and bladder cancer risk*. *Carcinogenesis*, 2004. **25**(6): p. 973-8.
35. Marsit, C.J., et al., *Carcinogen exposure and gene promoter hypermethylation in bladder cancer*. *Carcinogenesis*, 2006. **27**(1): p. 112-6.
36. Qin, Y., et al., *RFPL3 and CBP synergistically upregulate hTERT activity and promote lung cancer growth*. *Oncotarget*, 2015. **6**(29): p. 27130-45.
37. Beer, T.M., et al., *Polymorphisms of GSTP1 and related genes and prostate cancer risk*. *Prostate cancer and prostatic diseases*, 2002. **5**(1): p. 22-7.
38. Wahid, M., et al., *Expression of CYP1A1 and GSTP1 in human brain tumor tissues in Pakistan*. *Asian Pacific journal of cancer prevention : APJCP*, 2013. **14**(12): p. 7187-91.
39. Paranjape, T., F.J. Slack, and J.B. Weidhaas, *MicroRNAs: tools for cancer diagnostics*. *Gut*, 2009. **58**(11): p. 1546-54.
40. Du, W.Z., et al., *Targeting the SMO oncogene by miR-326 inhibits glioma biological behaviors and stemness*. *Neuro-Oncology*, 2015. **17**(2): p. 243-253.

41. Liang, Z.X., et al., *Involvement of miR-326 in chemotherapy resistance of breast cancer through modulating expression of multidrug resistance-associated protein 1*. *Biochemical Pharmacology*, 2010. **79**(6): p. 817-824.
42. Qiu, S.W., et al., *Interactions of miR-323/miR-326/miR-329 and miR-130a/miR-155/miR-210 as prognostic indicators for clinical outcome of glioblastoma patients*. *Journal of Translational Medicine*, 2013. **11**.
43. Wang, S., et al., *Expression and clinical significance of microRNA-326 in human glioma miR-326 expression in glioma*. *Medical Oncology*, 2013. **30**(1).
44. Wu, J., K. Du, and X. Lu, *Elevated expressions of serum miR-15a, miR-16, and miR-17-5p are associated with acute ischemic stroke*. *International journal of clinical and experimental medicine*, 2015. **8**(11): p. 21071-9.
45. Zhou, J.X., et al., *MicroRNA-326 Functions as a Tumor Suppressor in Glioma by Targeting the Nin One Binding Protein (NOB1)*. *PloS one*, 2013. **8**(7).
46. Sun, C., et al., *Hsa-miR-326 targets CCND1 and inhibits non-small cell lung cancer development*. *Oncotarget*, 2016. **7**(7): p. 8341-59.
47. Park, U.H., et al., *ASXL2 promotes proliferation of breast cancer cells by linking ERalpha to histone methylation*. *Oncogene*, 2016. **35**(28): p. 3742-52.
48. Katoh, M., *Functional and cancer genomics of ASXL family members*. *British journal of cancer*, 2013. **109**(2): p. 299-306.
49. Daou, S., et al., *The BAP1/ASXL2 Histone H2A Deubiquitinase Complex Regulates Cell Proliferation and Is Disrupted in Cancer*. *Journal of Biological Chemistry*, 2015. **290**(48): p. 28643-28663.
50. Xiong, J., Q. Du, and Z. Liang, *Tumor-suppressive microRNA-22 inhibits the transcription of E-box-containing c-Myc target genes by silencing c-Myc binding protein*. *Oncogene*, 2010. **29**(35): p. 4980-8.
51. Alvarez-Diaz, S., et al., *MicroRNA-22 is induced by vitamin D and contributes to its antiproliferative, antimigratory and gene regulatory effects in colon cancer cells*. *Human Molecular Genetics*, 2012. **21**(10): p. 2157-2165.
52. Yamakuchi, M., et al., *MicroRNA-22 Regulates Hypoxia Signaling in Colon Cancer Cells*. *PloS one*, 2011. **6**(5).
53. Zhang, J., et al., *microRNA-22, downregulated in hepatocellular carcinoma and correlated with prognosis, suppresses cell proliferation and tumourigenicity*. *British journal of cancer*, 2010. **103**(8): p. 1215-20.
54. Wang, G.J., et al., *Downregulation of miR-22 acts as an unfavorable prognostic biomarker in osteosarcoma*. *Tumor Biology*, 2015. **36**(10): p. 7891-7895.
55. Yuhong Lu, N.W., Mitchell S. Turker, Peter M. Glazer, *Epigenetic silencing of the DNA mismatch repair gene, MLH1, induced by hypoxic stress in a pathway dependent on the histone demethylase, LSD1*. *Cell reports*, 2014. **8**(2): p. 501-513.
56. Zeng, C., et al., *[Expression of miR-146a in colon cancer and its significance]*. *Nan fang yi ke da xue xue bao = Journal of Southern Medical University*, 2014. **34**(3): p. 396-400.
57. Yesudhas, D., et al., *Multiple roles of toll-like receptor 4 in colorectal cancer*. *Frontiers in immunology*, 2014. **5**: p. 334.
58. Li, T.T., S. Ogino, and Z.R. Qian, *Toll-like receptor signaling in colorectal cancer: carcinogenesis to cancer therapy*. *World journal of gastroenterology*, 2014. **20**(47): p. 17699-708.
59. Doan, H.Q., et al., *Toll-like receptor 4 activation increases Akt phosphorylation in colon cancer cells*. *Anticancer research*, 2009. **29**(7): p. 2473-8.
60. Salas, L.A., et al., *Biological and statistical approaches for modeling exposure to specific trihalomethanes and bladder cancer risk*. *Am J Epidemiol*, 2013. **178**(4): p. 652-60.

