

SCIENTIFIC REPORTS



OPEN

Identification of Sex-Specific Transcriptome Responses to Polychlorinated Biphenyls (PCBs)

Almudena Espín-Pérez¹, Dennie G. A. J. Hebels^{1,2}, Hannu Kiviranta³, Panu Rantakokko³, Panagiotis Georgiadis⁴, Maria Botsivali⁴, Ingvar A. Bergdahl⁵, Domenico Palli⁶, Florentin Späth⁵, Anders Johansson⁵, Marc Chadeau-Hyam⁷, Soterios A. Kyrtopoulos⁴, Jos C. S. Kleinjans¹ & Theo M. C. M. de Kok¹

PCBs are classified as xenoestrogens and carcinogens and their health risks may be sex-specific. To identify potential sex-specific responses to PCB-exposure we established gene expression profiles in a population study subdivided into females and males. Gene expression profiles were determined in a study population consisting of 512 subjects from the EnviroGenomarkers project, 217 subjects who developed lymphoma and 295 controls were selected in later life. We ran linear mixed models in order to find associations between gene expression and exposure to PCBs, while correcting for confounders, in particular distribution of white blood cells (WBC), as well as random effects. The analysis was subdivided according to sex and development of lymphoma in later life. The changes in gene expression as a result of exposure to the six studied PCB congeners were sex- and WBC type specific. The relatively large number of genes that are significantly associated with PCB-exposure in the female subpopulation already indicates different biological response mechanisms to PCBs between the two sexes. The interaction analysis between different PCBs and WBCs provides only a small overlap between sexes. In males, cancer-related pathways and in females immune system-related pathways are identified in association with PCBs and WBCs. Future lymphoma cases and controls for both sexes show different responses to the interaction of PCBs with WBCs, suggesting a role of the immune system in PCB-related cancer development.

Chemicals such as polychlorinated biphenyls (PCBs) are synthetic xenoestrogens with high solubility in fats and lipids and a long half-life, which causes an overall accumulation in the environment. PCBs were widely used in diverse industrial applications from the 1930s to the 1980s. Despite the prohibition of their production due to association with cancer risk and infertility¹, PCBs can still be found in the environment. They have been classified by the International Agency for Research on Cancer (IARC) as human carcinogens based on epidemiological evidence². They furthermore interfere in processes controlled by hormones and modify growth factors³.

It has also been suggested that exposure to xenoestrogens including PCBs might explain the sex-specific differences in cancer incidence that have been observed, where males showed to be more prone to cancer than females⁴⁻⁶.

Furthermore, PCB exposure has been associated with adverse effects on the immune system⁷ and immunosuppressive effects in unborn children^{8,9} and also in animal studies¹⁰⁻¹⁴. Exposure to environmental pollutants such as PCBs is known to induce changes in gene expression^{4,15,16}. To investigate the molecular mechanisms involved in lymphoma risk in association with PCB exposure, we performed an analysis of genome-wide RNA expression established in blood samples of 649 subjects from the EnviroGenomarkers Project. This EU seventh

¹Department of Toxicogenomics, Maastricht University, Maastricht, The Netherlands. ²MERLN Institute for Technology-inspired Regenerative Medicine, Maastricht University, Maastricht, The Netherlands. ³Department of Health Protection, Chemicals and Health Unit, National Institute for Health and Welfare, Kuopio, Finland. ⁴Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation, Athens, Greece. ⁵Department of Biobank Research, and Occupational and Environmental Medicine, Umeå University, Umeå, Sweden. ⁶Molecular and Nutritional Epidemiology Unit, ISPO Cancer Prevention and Research Institute, Florence, Italy. ⁷Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. Correspondence and requests for materials should be addressed to A.E.-P. (email: a.espin@maastrichtuniversity.nl)

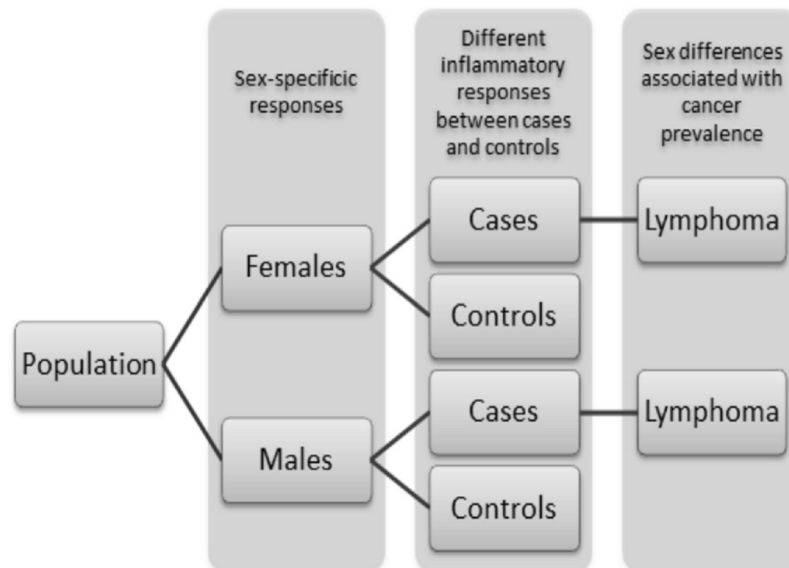


Figure 1. Stratification of the population for LMM analysis implementing PCB exposure and distribution of WBC, where each step address a different research question. *In addition to LMM implementing PCB exposure and distribution of WBC, LMM implementing PCB exposure alone has been performed.

framework project (<http://www.envirogenomarkers.net/>) aimed to develop ‘OMICS-based biomarkers’ for studying the role of environmental exposures in human health. The project applied different –omics technologies for the discovery and validation of novel biomarkers predictive of increased risks of chronic diseases in which the environment may play an important role, specifically breast cancer and Non Hodgkin’s lymphoma. Additionally, the EnviroGenomarkers project aimed to explore the association of such risk biomarkers with exposure to a number of high-priority or emerging environmental pollutants with carcinogenic, immunotoxic or hormone-like properties. The EnviroGenomarkers dataset presents some challenges in terms of the statistical analysis since it comprises two prospective cohorts from the Northern Sweden Health and Disease Study (NSHDS) and the European Prospective Investigation into Cancer and Nutrition (EPIC-Italy). In addition, significant batch effects need to be taken into account¹⁷. In order to address these challenges we analyzed the relationship between exposure to PCBs and gene expression using a linear mixed model (LMM), an increasingly prominent research tool in epidemiological data analysis because of its flexibility to correct for both fixed effects (such as demographic parameters) and technical variables (nuisance variables)¹⁸. We assessed the relationship between exposure to PCBs and gene expression using the LMM approach thereby adjusting for these potential confounders.

Since PCBs may affect the immune system, it is important to also have information on the distribution of white blood cells (WBC) in order to establish the effect of PCB exposure on human cancer risk. Information about the proportion of white blood cells (WBCs) is available from the EnviroGenomarkers dataset. The number of WBCs in the blood is adjusted in the LMM as a confounder but also as an interaction term together with PCBs due to its potential role in the causal path between the exposure and the gene expression.

The current study aims to establish the potential sex-specific gene expression changes induced by exposure to a mixture of PCBs under daily life circumstances by analyzing blood cell gene expression profiles in the population subdivided into females and males. Secondly, it sets out to assess the transcriptomic changes possibly mediated by immunotoxic effects associated with differences in the distribution of WBC. Finally, sex-specific differences in gene expression responses to PCB exposures in the subpopulation developing lymphoma in later life are identified (Fig. 1).

Results

Exposure level. Exposure levels of PCB118, PCB156, PCB170 and PCB180 were significantly different between females and males (Student’s t-test, p-value < 0.05). Figure 2 presents individual exposure levels. Robust differences in the LMM signals obtained in the two sexes were still present after deleting those high/low exposed females that caused these statistically significant differences, thus suggesting a sex-specific transcriptional response to PCBs. Based on this and in view of the importance to evaluate a larger range of exposure, the entire female population was used for further analyses. Exposure levels of PCBs are not different when other variables such BMI are assessed.

LMM for PCBs and gene expression: Sex-specific responses. Lists of transcripts associated with a p-value were generated from LMM analysis by comparing the levels of significance of the two models with and without a variable of interest. Afterwards an FDR correction (FDR cutoff of 0.05) was applied on the p-value following the Benjamini–Hochberg procedure¹⁹. The list of genes is available in the Excel document from Supplementary Material. As shown in the list representing the sum of transcript lists for individual PCBs from Table 1, the response to PCB exposure was only observed in the female population (mostly associated with the

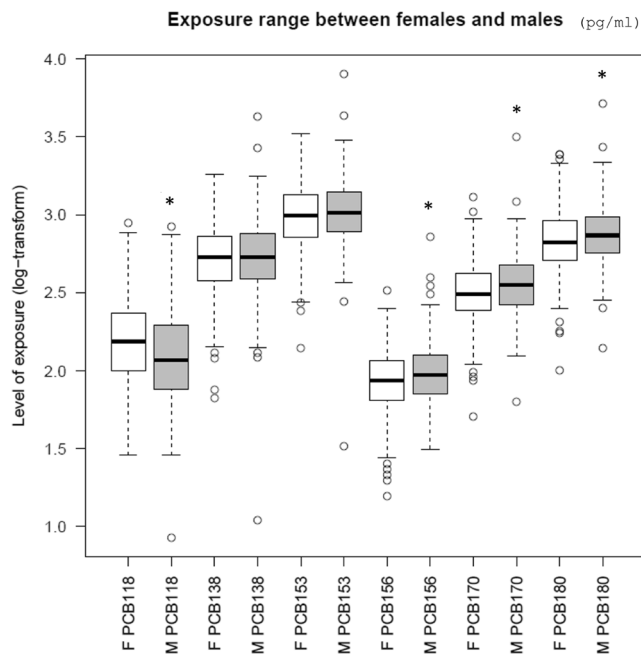


Figure 2. Range of PCB exposure for females (“F”, white boxplot) and males (“M”, grey boxplot). *p-value < 0.05 (Student’s t-test).

Population	Variable of interest	FDR < 0.05
Females	PCB118	0
Females	PCB138	0
Females	PCB153	504
Females	PCB156	4824
Females	PCB170	3939
Females	PCB180	4163
Males	PCB118	0
Males	PCB138	0
Males	PCB153	0
Males	PCB156	0
Males	PCB170	0
Males	PCB180	0

Table 1. Number of transcripts after linear mixed model analysis from the female and male population using only PCB as variable of interest.

NSHDS cohort, as the cohorts did not differ regarding BMI or other registered variables). When a less strict FDR cutoff (FDR < 0.2) was applied, some genes appeared induced by PCB170 also in males. The top 3 deregulated pathways in females in which these genes were involved are mainly generic and relate to gene expression, metabolism of proteins, translation and cell cycle (Supplementary Table S2).

The relatively large response to PCBs in the female population could potentially also be related to the average age of females 53.31 ± 7.81 years old, an age at which changes in the hormonal regulation occur involving for instance estrogen metabolism and menopausal processes. To test this hypothesis an analysis was conducted on 59 non-smoking females from the Italian cohort for whom age of menopause was registered. The LMM analysis specifically for PCB156, selected because it induces a large response (Table 1), controlling or not controlling for age of menopause as a confounder, gave different numbers of transcripts: with a FDR cutoff of 0.2, 53 hits for the not controlling approach and 134 hits for the controlling for age of menopause approach. The 53 genes were included within the 134 transcripts. The 81 additional hits that appeared after controlling for menopause age might be due to biological differences as a consequence of hormonal regulation. They were involved in pathways of DNA damage/telomere stress-induced senescence and in G1-to-S cell cycle control. Previous studies revealed associations between PCBs and breast cancer in postmenopausal women²⁰ and lower immune function²¹. However, the low number of samples with postmenopausal information available could also cause overfitting and therefore produce false positives.

LMM for analyzing the interaction between PCB exposure levels and distribution of WBC. We ran different LMM for the interaction between each individual PCB and WBC of the different individual WBC types in order to identify transcripts significantly associated with the combination of both PCBs and WBCs that are not found when looking at them individually. This means that the identified genes only respond to PCB exposure given a certain composition of the WBC population. The WBCs population itself might be also affected by the exposure, although this is not necessary required for the identification of these responding genes. The number of transcripts generated from each approach and appearing among men and women is presented in Supplementary Material Table S3. The list of genes is available in the Excel document from Supplementary Material. Interactions with B cell distribution appeared among both men and women, while interactions with natural killer cell and monocyte distribution appeared almost exclusively in women. The outcome of the pathway analysis is shown in Table 2, where cancer pathways specifically appear relatively stronger induced among males (8 out of 19 against 4 out of 18 among females).

The overlap of hits with the list of genes associated with PCB exposure as retrieved from CTD (Comparative Toxicogenomics Database) was around 10%, showing a high number of hits (531 for females, 191 for males) that were previously not known to be associated with PCB exposure but that were significant when we studied the interaction between PCBs and WBCs. This complex interaction may explain why these genes have not been identified before. For females, platelet-related pathways were identified from the 90% remaining genes, whereas cancer-related pathways were found for males. None of these new pathways were overlapping with the gene list found associated with PCBs only. Figure S1 from Supplementary Material shows the gene-gene interactions between the novel genes found by analyzing the interaction of WBCs and PCBs (90%) and the genes reported in literature to be associated with PCBs (10%). The genes from the cancer-related pathways identified in Table 2 are highlighted in the network.

LMM for evaluating the interaction between PCB exposure levels and distribution of WBC in control subjects and future lymphoma cases. The results from stratifying the population into healthy males and females that eventually developed lymphoma and healthy controls (males and females that remain healthy) showed very different numbers of transcripts (Table 3). The list of genes is available in the Excel document from Supplementary Material. There was no overlap between the lists of genes from the healthy and future disease population. Females showed highly significant pathways associated with the hits from the interaction PCBs and distribution of WBC only in the population of future cases and not in the controls (Supplementary Material Table S4). Males presented hits of cancer pathways associated with the interaction PCBs and B-cells in the future cases population and interaction PCBs and natural killers, CD4T and monocytes in the control population.

The column “LY Future cases” in Table 3 corresponds to the subpopulation of future lymphoma cases. Despite the reduction of the population size and therefore statistical power, the fact that most of the highly significant genes found within each separate analysis were involved in common pathways showed that it is rather unlikely that these genes represented only false positives.

Male cases showed only high response from the interaction model from B-cells. The interaction involving the rest of WBCs did not result in a large response. There was only one common pathway present in both sexes (females PCB118, females PCB138, females PCB153 and males PCB138) which was ubiquitin-mediated proteolysis in human, a pathway potentially involved in cancer prevention²². TGF-beta receptor-related pathways were most frequent among females while in males cancer-related and neuronal-related pathways were predominant (Supplementary Material Table S4).

Time of diagnosis. A highly significant but weak (low correlation coefficients) association between transcripts associated with PCB exposure identified in the PCBs and WBCs interaction LMM analysis and time of diagnosis was found. This implies that a higher or lower level of expression of transcripts is associated with a longer or shorter time to diagnosis. The correlation coefficients vary between -0.31 (p-value $5.05E-5$) and 0.19 (p-value 0.011) for females, -0.16 (p-value 0.082) and 0.25 (p-value 0.009) for males. The list of genes from the females population negatively correlated (p-value < 0.05) with time of diagnosis showed TGF-beta receptor and cell cycle-related pathways (Supplementary Material Table S5). No significant association was found between the distribution of WBC and time of diagnosis.

The survival analyses identified a few features relevant to future lymphoma status and time to diagnosis (Supplementary Material Table S6). The list of pathways associated with these features is shown in Supplementary Material Table S7. The ROC curves generated shows an important overlap with the expected line, suggesting a random separation between future lymphoma disease and future healthy, instead of a powerful predictive model (Supplementary Material Figure S2).

Lymphoma profile comparison. The genes corresponding to the top 5% over or under-expressed were compared to the list of genes identified by the LMM in association with PCB exposure. The significant hits from LMM did not show a substantial overlap with the top 5%, as shown in Supplementary Material Table S9, where it is displayed the number of genes and the overlap between the top 5% up/down-regulated genes from the profile and the significant genes from LMM (overlap expressed in percentages in the last four columns). This finding implies that the mechanisms found in association with PCBs might be different from the lymphoma disease specific signatures identified from the ArrayExpress database.

Names	Elements	Cancer related pathways
Females PCB153 Monocytes Females PCB156 Monocytes Females PCB170 Monocytes	Platelet degranulation	
	Response to elevated platelet cytosolic Ca ²⁺	
	Platelet activation, signaling and aggregation	
	Hemostasis	
Females PCB118 B-cell Females PCB153 B-cell	Regulation of cytoplasmic and nuclear SMAD2/3 signaling	X
Females PCB118 B-cell	miR-targeted genes in muscle cell - TarBase	
	Signaling by Hedgehog	X
	HH-Ncore	
Females PCB153 B-cell	Regulation of nuclear SMAD2/3 signaling	X
	Antiviral mechanism by IFN-stimulated genes	
	Inositol Metabolism	
	Unfolded Protein Response (UPR)	
	tgf beta signaling pathway	X
	ISG15 antiviral mechanism	
Females PCB156 Monocytes	mcalpain and friends in cell motility	
Females PCB170 Monocytes	Formation of Fibrin Clot (Clotting Cascade)	
	ADP signalling through P2Y purinoceptor 12	
	Common Pathway of Fibrin Clot Formation	
Males PCB156 B-cell	Retrograde endocannabinoid signaling - Homo sapiens (human)	
Males PCB170 B-cell	BDNF signaling pathway	
Males PCB180 B-cell	Pathways in cancer - Homo sapiens (human)	X
Males PCB153 B-cell	Angiogenesis	X
Males PCB156 B-cell		
Males PCB156 B-cell	Ca ²⁺ pathway	
	Proteoglycans in cancer - Homo sapiens (human)	X
	Circadian entrainment - Homo sapiens (human)	
	G-protein activation	X
	Opioid Signalling	
	G Protein Signaling Pathways	X
	GABAergic synapse - Homo sapiens (human)	
	Signaling by Wnt	X
	Morphine addiction - Homo sapiens (human)	
	Choline metabolism in cancer - Homo sapiens (human)	X
	Rap1 signaling pathway - Homo sapiens (human)	
beta-catenin independent WNT signaling	X	
Males PCB170 B-cell	Neural Crest Differentiation	
	Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds	
	Signaling pathways regulating pluripotency of stem cells - Homo sapiens (human)	

Table 2. Pathways (q-value < 0.05) for interaction between PCBs and WBCs in all females and all males (transcriptomics approach, FDR < 0.05).

Discussion

We investigated the impact of PCB exposure on gene expression profiles in a human population study by applying LMMs. In addition, we identified the effect of the interaction between individual PCBs and distribution of WBC. The small overlap observed between pathways affected by different individual PCBs suggests that also different biological mechanisms of action are activated by varying exposures (Supplementary Material Tables S3 and S4). In the same way, interaction between PCBs and different distribution of WBC resulted in different responses for each WBC, demonstrating the importance of considering individual congeners and distribution of WBC in data analyses. In addition, of the total number of 531 transcripts for females that were significantly associated with the interaction PCBs and WBCs, 90% of them have not previously been associated with PCB exposure. This implies that only by looking at PCBs induced-responses these novel targets cannot be identified. The interaction between PCB156 and B-cells resulted in a number of pathways directly related to cancer, especially for males. Deregulated cancer pathways among males were also observed in association with other PCBs (highlighted in the column “Cancer related pathways” from Table 2) and according to literature, lymphoma incidence among males is significantly higher than in females^{22,23}. Platelet-related pathways were found in the female population in the interaction with PCBs and monocytes, which is not surprising in view of the fact that it has been reported earlier in literature associations between PCBs and cardiovascular disease²⁴ and platelets and cardiovascular disease^{25,26}. Figure S1 from Supplementary Material shows the gene network of the novel genes and the genes reported in

literature, indicating the genes that are known to be associated with cancer disease from Table 2. There were two estrogen-related activity pathways for different PCBs in future lymphoma cases in females (Supplementary Material Table S4). In the male future cases subpopulation, the interaction between PCB138 and B-cells showed also a large effect on the pathway level while in females this occurred in the interaction between PCB118 and B-cells.

Individual PCB congeners have different physicochemical properties that lead to different mechanism of toxicity. Because of their structural differences, PCBs can be divided into dioxin-like and non dioxin-like compounds (co-planar and non-co-planar), although PCB118 can be considered as a mixed type inducer^{27,28}. PCBs are known to cause genotoxicity, sperm DNA damage, gene mutation, epigenetic effects and changes in gene expression². AhR, a ligand-activated transcription factor that regulates xenobiotic-metabolizing enzymes and suppression of apoptosis, is known to be activated by dioxin-like PCBs^{2,29}. The AhR is involved in most of the dioxin-induced immune suppression processes as it was previously demonstrated in a human population study⁸. Non dioxin like PCBs, do not activate the AhR and are known to be neurotoxic and immunotoxic. However, these effects are only observed at much higher concentrations. In our study, the AhR was only found to respond significantly in the interaction models for PCBs and monocytes in females but in contrast to what it was reported in literature, AhR was modulated by both dioxin-like and non-dioxin-like PCBs. However, in the interaction models the association between the AhR expression levels and interaction of exposure and WBCs was negative. This might be explained by the fact that PeCB (a PCB congener) can induce apoptosis of monocytes, and therefore AhR could be activated because the interaction term is low, due to low levels of monocytes³⁰. The AhR interacts with other transcription and signaling pathways such as MAPKs proteins and induces enzymes linked to bioactivation of xenobiotics such as CYPs, particularly the subfamily members of CYPs CYP1A and CYP2B². MAPKs and several CYP isoforms such as CYP11B1 (which are not induced by AhR agonists) were found to be significant in our PCBs models only for females. We have seen induction of CYP11B1, an effect that might be related to the steroid 11b-hydroxylase activity of aromatizing androstenedione to estrone. CYP11B1 has also previously been found to be associated with PCBs exposure³¹. NR3C and NR1D3 were also found to be associated with exposure to both dioxin-like and non-dioxin-like PCBs. Associations between exposure to dioxin-like PCBs and expression of C-Src, MAP Kinase, and CXCR4 have been described in literature³² but none of these genes is found in our study. In the same way, associations between NFKB1, STAT3, STAT5, GRIN2B, GRIA2, VEGF and FOXO3^{33–36} and non-dioxin-like PCBs are found in literature but not identified in our study. This lack of overlap might be a consequence of the species, tissue and dose differences between our study and those reported in the literature. However, the different pathways shown in the results in association with different PCBs suggest a variety of ways that each PCB interacts with genes and also independently of the classification (dioxin-like or non dioxin-like) relevant to the AhR-mediated activation^{37,38}. Other toxic effects associated with PCBs are endocrine disruption by blocking the thyroid system functioning and neurotoxicity. We found in our study TSHB (Thyroid Stimulating Hormone Beta) in females as well as several neural signaling pathways in both females and males.

PCB exposure has been associated with a variety of cancers³⁹. We found a high number of significant hits associated with the interaction of PCBs with WBCs that were not previously associated with PCB exposure. These new hits are involved in platelet-related pathways in females and cancer-related pathways in males. The platelet-derived growth factor has been linked with cancer previously^{40,41}. Furthermore the combination of an immune challenge and PCB exposure may be a prerequisite for the development of lymphoma⁴². Immunosuppression and inflammation can be one of the adverse effects of exposure to environmental PCBs⁴³ which eventually can lead to tumorigenesis⁴⁴. This demonstrates the need for a detailed evaluation of the impact that different white blood cell types may have on gene expression responses to environmental exposures like PCBs. Therefore, we consider it of importance to take the influence of WBC distribution into account when assessing the biological effects induced by exposure to compounds like PCBs.

We found sex-specific differences in response to PCBs on the gene expression level. Small but significant differences in the exposure level of certain PCB congeners between men and women introduce some difficulties in the comparison of biological responses to PCBs. However, the strong signals that were found in the female population were still present after removing the high/low exposed individuals, which were responsible for the statistical difference in exposure as compared to the male exposure levels. This outcome together with the small variation in exposure level indicates that the large differences in response were likely due to biological differences between the sexes, and not to false positive findings. In addition, keeping the entire female population in the analysis has identified biological processes of relevance. Comparison of profiles of lymphoma future cases between males and females showed almost no overlap between the sexes on the pathway level (Supplementary Material Table S4). The disparity in responses to exposure between females and males were more likely due to a dissimilar biological response than to the small difference in exposure levels.

We assessed the direction of the expression of the genes involved in cancer-related pathways from males (Table 2). Seven out of eight genes that are involved in all of these pathways showed the same direction associated with PCB exposure as other studies comparing controls and certain cancer diseases including head, neck cancer, colorectal cancer, lung carcinoma, basal cell carcinoma and uterine carcinoma. These genes were GNG7⁴⁵, AXIN2⁴⁶, HIF1A⁴⁷, FGF2⁴⁸, PPAP2B⁴⁹ and LAMC1⁵⁰ and GNAO1⁵¹. GNG7, AXIN2, HIF1A, FGF2 and PPAP2B are found upregulated in our study in the same way as it is found in association with lymphoma disease according to literature^{52–57}.

We observed that the interaction between PCB exposure and distribution of WBC had a different effect in future lymphoma cases and controls (Table 3). This finding suggests that the interaction term might play a role in the development of disease and therefore this information might be relevant for risk assessment and prevention of disease. However, we did not find strong evidence of a relationship between genes identified by LMM and time to diagnosis.

Population	Cell type	PCB	Controls	Overlap	LY Future cases
Females	B cells	118	0	0	1055
Females	B cells	138	0	0	650
Females	B cells	153	0	0	563
Females	B cells	156	2	0	48
Females	B cells	170	2	0	14
Females	B cells	180	2	0	127
Females	CD8T	118	0	0	0
Females	CD8T	138	0	0	1
Females	CD8T	153	0	0	5
Females	CD8T	156	0	0	4
Females	CD8T	170	0	0	4
Females	CD8T	180	0	0	3
Females	CD4T	118	0	0	3121
Females	CD4T	138	0	0	2755
Females	CD4T	153	0	0	3004
Females	CD4T	156	0	0	3586
Females	CD4T	170	0	0	4326
Females	CD4T	180	0	0	5315
Females	NK	118	0	0	2
Females	NK	138	0	0	1
Females	NK	153	0	0	1
Females	NK	156	0	0	73
Females	NK	170	0	0	21
Females	NK	180	0	0	1
Females	Mono	118	0	0	723
Females	Mono	138	0	0	7445
Females	Mono	153	1	0	8373
Females	Mono	156	1	1	10120
Females	Mono	170	2	0	9352
Females	Mono	180	2	0	7944
Males	B cells	118	0	0	238
Males	B cells	138	0	0	195
Males	B cells	153	0	0	61
Males	B cells	156	0	0	28
Males	B cells	170	0	0	24
Males	B cells	180	0	0	14
Males	CD8T	118	1	0	0
Males	CD8T	138	1	0	0
Males	CD8T	153	1	0	0
Males	CD8T	156	1	0	0
Males	CD8T	170	1	0	0
Males	CD8T	180	1	0	0
Males	CD4T	118	1	0	0
Males	CD4T	138	0	0	0
Males	CD4T	153	0	0	0
Males	CD4T	156	0	0	0
Males	CD4T	170	26	0	0
Males	CD4T	180	34	0	0
Males	NK	118	0	0	0
Males	NK	138	0	0	0
Males	NK	153	1099	0	0
Males	NK	156	1227	0	1
Males	NK	170	3070	0	0
Males	NK	180	2621	0	0
Males	Mono	118	1	0	1
Males	Mono	138	0	0	2
Continued					

Population	Cell type	PCB	Controls	Overlap	LY Future cases
Males	Mono	153	0	0	0
Males	Mono	156	21	0	0
Males	Mono	170	284	0	5
Males	Mono	180	182	0	1

Table 3. Number of transcripts (FDR < 0.05) after LMM analysis after stratifying into controls and future cases for females and males using the interaction PCBs-WBCs.

Natural killer cells are components from the innate immune system and they are the only type of cells that lack pro-tumorigenic function^{44,58}. Moreover, tumor initiation by external factors influences the innate immune system to initiate cancer development^{59,60}. In healthy male subjects, we found that the combined effect of PCBs and natural killers or monocytes (innate response) induce a larger response than the interaction between PCBs and CD4T, CD8T or B-cells (adaptive immune system). In future male cases, the response was the other way around, suggesting that healthy subjects might respond to PCBs by activating the innate immune system and subjects that are prone to development of lymphoma show an adaptive immune response.

Most of the genes identified in the lymphoma profile comparison (in the range of 77% to 100% with a median of 99.8%) are different from the genes identified in our study in association with PCBs, suggesting a different mechanism of action than indicated by the strongest signals in the lymphoma-specific pathways. These genes are involved in 150 unique pathways related to drug metabolism, cell cycle and immune response (Supplementary Material Table S7). Future cases had not being diagnosed with cancer at the moment of sampling. However, we cannot exclude that they are in an early stage of the disease.

Conclusion

Results from LMM analysis showed sex-specific responses to PCBs. Males and females showed a significant transcriptional response to the interaction between PCB exposures and the distribution of WBC, indicating in particular the induction of cancer pathways among males and the involvement of the immune system among females. The relevant inflammatory responses in cancer development indicate the necessity of including the different WB cell types in the study in the context of risk assessment of environmental exposures such as PCBs.

Methods

The EnviroGenomarkers study population. Lifestyle factors and anthropometric measurements were registered at the time of sampling (demographics in Table S1 Supplementary material) and blood samples were frozen and stored for later analysis^{61,62}. The quality of the biobank samples was previously evaluated, concluding amenability for OMICS analyses⁶³. A follow-up of all recruited participants for a period of 2 to 17 years (5.96 ± 2.85) identified subjects that did not have a B-cell lymphoma at the moment of blood sampling but developed that in later life. These subjects are defined in this study as “future cases”, while not having a disease at the time of inclusion and sampling. These future cases were matched with controls based on sex, age, center of inclusion, fasting status and date of blood collection.

The EnviroGenomarkers project was approved by the Regional Ethical Review Board of the Umea Division of Medical Research and the Florence Health Unit Local Ethical Committee and all methods were carried out in accordance with the approved guidelines. All participants gave written informed consent.

Laboratory analysis and pre-processing. Samples from EPIC-Italy were stored in cryostraws that were opened by cutting them with RNase-free tools on a stainless steel plate imbedded in a box of dry ice to prevent thawing during handling. Then the frozen buffy coat was pushed out with a thin stainless steel plunger directly into 1.2 mL of RNAlater (QIAGEN) solution. NSHDS samples from their cryovials were retrieved in a frozen state by making a small opening at the bottom of each vial using a hot plunger and pushing the sample out with another plunger. This material was immediately thawed in 1.2 mL of RNAlater (QIAGEN). RNA was isolated on the same day with the RiboPure™ Blood kit (Ambion, Austin, TX, USA) using the manufacturer's miRNA isolation protocol.

Agilent 4 × 44 K human whole genome microarray analyses were conducted using our standard methodologies^{63,64}. In brief, each RNA sample was reverse transcribed into cDNA and labeled with cyanine 3 prior to hybridization. Subsequently, slides were washed and scanned using the Agilent Technologies G2565CA DNA Microarray Scanner. The technical performance and quality of the microarrays was established by visual evaluation of the scan images before and after within- and between-array normalization (using LOESS and A-quantile, respectively). In order to ensure that the samples from the different biobanks were suitable for -omics analysis, the quality was extensively evaluated⁶³. Probes that did not contain at least 75% of the maximum possible number of pixels were filtered out and imputation of the missing values was performed using the *k* nearest neighbors approach (*k* = 15, Euclidian metric). In total, 512 samples remained after quality assessment of the gene expression microarrays, removal of outliers and subjects with high incidence of missing values.

DNA was extracted from buffy coats. Samples were thawed on ice and DNA was isolated using the QIAamp Blood Mini Kit (QIAGEN), evaluating it spectrophotometrically and by agarose gel electrophoresis. Genome-wide analysis of DNA methylation using Infinium HumanMethylation450 BeadChips (Illumina, San Diego, CA, USA), which contain 485,764 probes (>99% with CpG dinucleotides), was conducted following the manufacturer's recommendations⁶³. Methylation module and an adaptation of the HumMeth27QCReport was used⁶⁵. Within- and between-array normalization (linear LOESS and A-quantile, respectively) using Gene ARMADA⁶⁶ and imputation of missing values using the *k* nearest neighbors approach (*k* = 15, Euclidian metric) was conducted.

PCBs 118, 138, 153, 156, 170, and 180 are abundant congeners that can describe the total exposure to a great extent. PCB congeners in serum samples were liquid-liquid extracted and analyzed with either HP 6890GC connected to Waters Autospec Ultima high resolution mass spectrometer (HRMS) or an Agilent 7000B Triple Quadrupole GC-MS/MS system⁶⁷. Results of the reference material generated by both methods were collected and we concluded that they are comparable.

Study of dependency between variables. The Variance Inflation Factor (R package “HH”)⁶⁸ assesses the level of multicollinearity between effects. Variables that can be linearly predicted from others are not suitable to be analyzed in the same model. To evaluate this, we performed a test for PCB118, PCB138, PCB153, PCB156, PCB170, PCB180, cohort, sex, age, BMI, smoking status (smokers, former smokers and nonsmokers) and the distribution of WBC including natural killer cells, B-cells, monocytes, granulocytes and cytotoxic T-cells such as CD8T, CD4T. The distribution of WBC had been calculated from the DNA methylation data (Illumina Infinium HumanMethylation450 BeadChip platform) of the same subjects (measured in the context of the EnviroGenomarkers project), using a published algorithm from Houseman *et al.*⁶⁷. This method quantifies the mixed composition of leukocytes using DNA methylation signatures. The cell mixture reconstruction method infers changes in the distribution of WBC using signatures from purified leukocyte samples and it has been validated using different datasets⁶⁹.

The variance inflation factor test returned values below 4 for all the variables except for PCB exposure levels and granulocytes, meaning that the remaining confounders (cohort, sex, type of cancer, age, BMI, smoking status and distribution of WBC except for granulocytes) could be analyzed independently for an individual PCB in each LMM. The different PCBs were considered separately in the statistical analysis also due to the different mechanisms of action among them and between dioxin/non dioxin-like⁷⁰.

LMM analysis. LMMs were fitted in order to find associations between the level of PCB exposures and the level of gene expressions after pre-processing, thereby correcting for potential confounders and nuisance random variation. The general mixed model formula for subject *i* is represented by:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 FE_i + uA_i + \epsilon_i \quad (1)$$

where Y_i represents the level of gene expression for each subject *i*, α the intercept of the model, X_i the independent variable of interest, FE_i the vector of fixed effects for subject *i*, β_1 and β_2 regression coefficient of the variables, uA_i the random effect due to nuisance variance and ϵ_i is the residual error or difference between observed and calculated value by the model.

We applied this LMM in order to evaluate the total effect of the independent variable, e.g. the individual PCB exposures, on the gene expression. We controlled for the potential confounders cohort, sex, type of future cancer, age, BMI, smoking status as well as the individual WBC distribution values as fixed effects. The date of microarray scanning was added as a random effect. We corrected for false discovery rate using the Benjamini and Hochberg's threshold at 5%.

Stratification of the analysis. The population was divided into females and males for reasons mentioned earlier.

Power calculation for the different subpopulations was conducted using the R package “pwr” (“pwr.f2.test” function)⁷¹ to estimate for each sample size the number of predictors in order to reach enough statistical power and level of significance 0.05.

Further stratification based on cohort and smoking status has been considered for females and males. We observed differences among groups in terms of number of hits but due to the large reduction of sample size, it was decided to correct for the variables “cohort” and “smoking status” in the whole study population.

Mixed models implementing exposure and distribution of WBC. Several approaches using LMMs were explored in order to study the impact of PCBs on gene expression. The first comprises the analysis of individual PCBs, accounting for WBC distribution as confounders together with the other variables. This approach was applied to the whole population as well as to females and males separately.

In view of the potential influence of the presence of different blood cell types on PCB-induced gene expression profiles, a second approach implementing WBC distribution was applied. Here, the variable of interest is presented by the interaction term between the individual PCBs and the different cell types:

$$Y_i = \alpha + \beta_1 PCB_i + \beta_2 WBC_i + \beta_3 PCB_i \times WBC_i + \beta_4 FE_i + uA_i + \epsilon_i \quad (2)$$

Initial biological interpretation. In order to assess the additional biological information provided by analyzing the interaction of PCBs and WBCs, we compared the list of significant hits for males and females separately with a list of genes known to be associated with PCB exposure by exploiting the Comparative Toxicogenomics Database (<http://ctdbase.org/>)⁷².

Pathway analyses. The lists of gene expressions significantly associated with PCB exposure were used to perform pathway analysis by means of ConsensusPathDB⁷³. The analysis was executed using the default parameters. Each input list of genes was compared against a background list created specifically for this analysis containing the total list of genes after the preprocessing filtering.

Lymphoma profile comparison. In order to relate transcriptomic responses to PCB exposure to the future disease status, from the ArrayExpress database (<https://www.ebi.ac.uk/arrayexpress/>), 34 microarray data sets based on blood samples from T-cell lymphoma patients were downloaded, of which 14 corresponded to females and 20 to males. Due to difficulties in finding more suitable databases for this study, T-cell lymphoma has to be used as a proxy for all lymphoma subcases. Additionally, another 100 microarray data sets from blood samples of healthy volunteers (50% females and 50% males) were used as controls⁴. Fold changes and level of significance were calculated using the “Limma” R package⁷⁴. In view of the large number of differentially expressed genes obtained (two thirds of the total genes: cutoff fold change |1.5|, adjusted p-value 0.05) and in order to highlight specific lymphoma-related genes, only significant genes from the top 5% and the bottom 5% in fold change were selected for creating the lymphoma signature.

Time to diagnosis. The time from sampling to diagnosis of disease in future cases might give some information about the development of disease. We therefore calculated the correlation between the expression after batch correction using LMM of significant genes and the time of diagnosis of lymphoma for females and males. We also studied the correlation between the level of WBCs and time of diagnosis.

Survival models in combination with elastic net were run in order to assess the predictive power of our list of transcripts to classify future lymphoma disease. All the unique transcripts that were significant were selected from all the statistical models from Table 3 (FDR 0.05), separately for females (12,938 transcripts in total) and males (3,767 transcripts in total). PCBs levels and the distribution of WBCs were added as additional features. Then, we conducted survival analyses using the R function “glmnet”, with “family” as “cox” from the R packages “survival” and “glmnet”. The survival analyses were modelled using status (future disease or future healthy) and time to diagnosis as input.

ROC curves were also generated using the R packages “caret” and “ROCR”. The same features as the survival analyses from above were used to apply elastic net, dividing the dataset into train dataset (70% of the original dataset) and test dataset (30% of the original dataset).

Data Availability

The data set from the EPIC-Italy cohort supporting the results of this article is available in the Array repository, accession ID E-MTAB-7273. The data set from the NSHDS cohort is available upon request. Requests for the individual-level data can be made to the Department of Biobank Research, Umeå University (<http://www.biobank.umu.se/biobank/nshds/>), and will be subject to ethical review and assessment by a panel of scientists. Individual-level data cannot be made publicly available due to legal restrictions imposed by the Swedish Data Protection Authority. All relevant aggregated data are presented in the article. Contact details: Ingvar Bergdahl (ingvar.bergdahl@umu.se) or Åsa Ågren (asa.agren@umu.se).

References

- Diamanti-Kandarakis, E. *et al.* Endocrine-disrupting chemicals: an Endocrine Society scientific statement. *Endocr Rev.* **30**(4), 293–342 (2009).
- Lauby-Secretan, B. *et al.* Carcinogenicity of polychlorinated biphenyls and polybrominated biphenyls. *Lancet Oncol.* **14**(4), 287–8 (2013).
- Albini, A. *et al.* Exogenous hormonal regulation in breast cancer cells by phytoestrogens and endocrine disruptors. *Curr Med Chem.* **21**(9), 1129–45 (2014).
- Espin-Perez, A. *et al.* Distinct genotype-dependent differences in transcriptome responses in humans exposed to environmental carcinogens. *Carcinogenesis.* **36**(10), 1154–61 (2015).
- De Coster, S. *et al.* Gender-specific transcriptomic response to environmental exposure in Flemish adults. *Environ Mol Mutagen.* **54**(7), 574–88 (2013).
- Hochstenbach, K. *et al.* Global gene expression analysis in cord blood reveals gender-specific differences in response to carcinogenic exposure in utero. *Cancer Epidemiol Biomarkers Prev.* **21**(10), 1756–67 (2012).
- Serdar, B. *et al.* Potential effects of polychlorinated biphenyls (PCBs) and selected organochlorine pesticides (OCPs) on immune cells and blood biochemistry measures: a cross-sectional assessment of the NHANES 2003–2004 data. *Environ Health.* **13**, 114 (2014).
- Hochstenbach, K. *et al.* Toxicogenomic profiles in relation to maternal immunotoxic exposure and immune functionality in newborns. *Toxicol Sci.* **129**(2), 315–24 (2012).
- Stolevik, S. B. *et al.* Prenatal exposure to polychlorinated biphenyls and dioxins from the maternal diet may be associated with immunosuppressive effects that persist into early childhood. *Food Chem Toxicol.* **51**, 165–72 (2013).
- De Krey, G. K. *et al.* Polychlorinated biphenyl-induced suppression of cytotoxic T lymphocyte activity: role of prostaglandin-E2. *Toxicol Lett.* **74**(3), 211–20 (1994).
- Arnold, D. L. *et al.* Toxicological consequences of Aroclor 1254 ingestion by female rhesus (*Macaca mulatta*) monkeys. Part 1B. Prebreeding phase: clinical and analytical laboratory findings. *Food Chem Toxicol.* **31**(11), 811–24 (1993).
- Arkoosh, M. R. *et al.* Suppression of B-cell mediated immunity in juvenile chinook salmon (*Oncorhynchus tshawytscha*) after exposure to either a polycyclic aromatic hydrocarbon or to polychlorinated biphenyls. *Immunopharmacol Immunotoxicol.* **16**(2), 293–314 (1994).
- Harper, N. *et al.* Immunosuppressive activity of polychlorinated biphenyl mixtures and congeners: nonadditive (antagonistic) interactions. *Fundam Appl Toxicol.* **27**(1), 131–9 (1995).
- Tryphonas, H. *et al.* Effects of Great Lakes fish consumption on the immune system of Sprague-Dawley rats investigated during a two-generation reproductive study. *Regul Toxicol Pharmacol.* **27**(1 Pt 2), S40–54 (1998).
- Pujolar, J. M. *et al.* Surviving in a toxic world: transcriptomics and gene expression profiling in response to environmental pollution in the critically endangered European eel. *BMC Genomics.* **13**, 507 (2012).
- Stecca, L. *et al.* Development of an *in vitro* test battery model based on liver and colon cancer cell lines to discriminate PCB mixtures by transcription factors gene expression analysis. *Toxicol In Vitro.* **34**, 204–11 (2016).
- Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet.* **11**(10), 733–9 (2010).
- Whitcomb, B. W. *et al.* Treatment of batch in the detection, calibration, and quantification of immunoassays in large-scale epidemiologic studies. *Epidemiology.* **21**(4), S44–50 (2010).

19. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*. **57**(1), 289–300 (1995).
20. Brauner, E. V. *et al.* Adipose tissue PCB levels and CYP1B1 and COMT genotypes in relation to breast cancer risk in postmenopausal Danish women. *Int J Environ Health Res*. **24**(3), 256–68 (2014).
21. Spector, J. T. *et al.* Plasma polychlorinated biphenyl concentrations and immune function in postmenopausal women. *Environ Res*. **131**, 174–80 (2014).
22. Chen, D. & Dou, Q. P. The Ubiquitin-Proteasome System as a Prospective Molecular Target for Cancer Treatment and Prevention. *Current Protein & Peptide Science*. **11**(6), 459–470 (2010).
23. Smith, A. *et al.* Lymphoma incidence, survival and prevalence 2004–2014: sub-type analyses from the UK's Haematological Malignancy Research Network. *Br J Cancer*. **112**(9), 1575–84 (2015).
24. Perkins, J. T. *et al.* Polychlorinated biphenyls and links to cardiovascular disease. *Environmental science and pollution research international*. **23**(3), 2160–72 (2016).
25. Papapanagiotou, A. *et al.* The Role of Platelets in Cardiovascular Disease: Molecular Mechanisms. *Current pharmaceutical design*. **22**(29), 4493–4505 (2016).
26. Willoughby, S., Holmes, A. & Loscalzo, J. Platelets and cardiovascular disease. *European journal of cardiovascular nursing: journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*. **1**(4), 273–88 (2002).
27. Yang, H. *et al.* Molecular Mechanisms of 2, 3', 4, 4', 5-Pentachlorobiphenyl-Induced Thyroid Dysfunction in FRTL-5 Cells. *PLoS one*. **10**(3), e0120133 (2015).
28. McFarland, V. A. & Clarke, J. U. Environmental occurrence, abundance, and potential toxicity of polychlorinated biphenyl congeners: considerations for a congener-specific analysis. *Environ Health Perspect*. **81**, 225–239 (1989).
29. Sartor, M. A. *et al.* Genomewide Analysis of Aryl Hydrocarbon Receptor Binding Targets Reveals an Extensive Array of Gene Clusters that Control Morphogenetic and Developmental Programs. *Environ Health Perspect*. **117**(7), 1139–1146 (2009).
30. Shin, K. J. *et al.* 2, 2', 4, 6, 6'-pentachlorobiphenyl induces apoptosis in human monocytic cells. *Toxicology and applied pharmacology*. **169**(1), 1–7 (2000).
31. van den Dungen, M. W. *et al.* Steroid hormone related effects of marine persistent organic pollutants in human H295R adrenocortical carcinoma cells. *Toxicology in Vitro*. **29**(4), 769–778 (2015).
32. Dutta, S. K. *et al.* Differential gene expression and a functional analysis of PCB-exposed children: Understanding disease and disorder development. *Environ Int*. **40**, 143–154 (2012).
33. Tharappel, J. C. *et al.* Regulation of cell proliferation, apoptosis, and transcription factor activities during the promotion of liver carcinogenesis by polychlorinated biphenyls. *Toxicology and applied pharmacology*. **179**(3), 172–184 (2002).
34. Miyazaki, W. *et al.* Identification of the functional domain of thyroid hormone receptor responsible for polychlorinated biphenyl-mediated suppression of its action *in vitro*. *Environ Health Perspect*. **116**(9), 1231–1236 (2008).
35. Raggi, F. *et al.* Divergent Effects of Dioxin- or Non-Dioxin-Like Polychlorinated Biphenyls on the Apoptosis of Primary Cell Culture from the Mouse Pituitary Gland. *PLoS one*. **11**(1) (2016).
36. Al-Anati, L., Hogberg, J. & Stenius, U. Non-dioxin-like PCBs interact with benzo[a]pyrene-induced p53-responses and inhibit apoptosis. *Toxicology and applied pharmacology*. **249**(2), 166–177 (2010).
37. Peters, A. K. *et al.* Determination of *in vitro* relative potency (REP) values for mono-ortho polychlorinated biphenyls after purification with active charcoal. *Toxicology letters*. **165**(3), 230–241 (2006).
38. Van den Berg, M. *et al.* The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicological Sciences*. **93**(2), 223–241 (2006).
39. Leng, L. *et al.* Polychlorinated biphenyls and breast cancer: A congener-specific meta-analysis. *Environ Int*. **88**, 133–141 (2015).
40. Heldin, C. H. Targeting the PDGF signaling pathway in tumor treatment. *Cell Communication and Signaling*. **11** (2013).
41. Demoulin, J. B. & Essaghir, A. PDGF receptor signaling networks in normal and cancer cells. *Cytokine & Growth Factor Reviews*. **25**(3), 273–283 (2014).
42. Strauss, H. S. & Heiger-Bernays, W. Methodological limitations may prevent the observation of non-Hodgkin's lymphoma in bioassays of polychlorinated biphenyls. *Toxicol Pathol*. **40**(7), 995–1003 (2012).
43. Dietert, R. R. Role of developmental immunotoxicity and immune dysfunction in chronic disease and cancer. *Reprod Toxicol*. **31**(3), 319–26 (2011).
44. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell*. **140**(6), 883–99 (2010).
45. Demokan, S. *et al.* Identification of guanine nucleotide-binding protein gamma-7 as an epigenetically silenced gene in head and neck cancer by gene expression profiling. *International journal of oncology*. **42**(4), 1427–36 (2013).
46. Yochum, G. S. AXIN2: Tumor Suppressor, Oncogene or Both in Colorectal Cancer? *J Cancer Sci Ther*, **2012** **4**, 22–23 (2012).
47. Koshikawa, N. *et al.* Constitutive upregulation of hypoxia-inducible factor-1alpha mRNA occurring in highly metastatic lung carcinoma cells leads to vascular endothelial growth factor overexpression upon hypoxic exposure. *Oncogene*. **22**(43), 6717–24 (2003).
48. Li, D. *et al.* A novel decoy receptor fusion protein for FGF-2 potently inhibits tumour growth. *British journal of cancer*. **111**(1), 68–77 (2014).
49. Asplund, A. *et al.* Expression profiling of microdissected cell populations selected from basal cells in normal epidermis and basal cell carcinoma. *The British journal of dermatology*. **158**(3), 527–38 (2008).
50. Kashima, H. *et al.* Laminin C1 expression by uterine carcinoma cells is associated with tumor progression. *Gynecologic oncology*. **139**(2), 338–44 (2015).
51. Pei, X. *et al.* The down-regulation of GNAO1 and its promoting role in hepatocellular carcinoma. *Bioscience reports*. **33**(5) (2013).
52. Giefing, M. *et al.* Identification of candidate tumour suppressor gene loci for Hodgkin and Reed-Sternberg cells by characterisation of homozygous deletions in classical Hodgkin lymphoma cell lines. *British Journal of Haematology*. **142**(6), 916–924 (2008).
53. Tiemessen, M.M. *et al.* The Nuclear Effector of Wnt-Signaling, Tcf1, Functions as a T-Cell-Specific Tumor Suppressor for Development of Lymphomas. *PLoS Biology*. **10**(11) (2012).
54. Qiao, Q. *et al.* NF-kappaB mediates aberrant activation of HIF-1 in malignant lymphoma. *Experimental hematology*. **38**(12), 1199–208 (2010).
55. Krejci, P. *et al.* FGF-2 expression and its action in human leukemia and lymphoma cell lines. *Leukemia*. **17**(4), 818–20 (2003).
56. Chapman, J. *et al.* Gene expression analysis of plasmablastic lymphoma identifies downregulation of B-cell receptor signaling and additional unique transcriptional programs. *Leukemia*. **29**(11), 2270–3 (2015).
57. Seifert, M. *et al.* Cellular origin and pathophysiology of chronic lymphocytic leukemia. *The Journal of experimental medicine*. **209**(12), 2183–98 (2012).
58. Marcus, A. *et al.* Recognition of tumors by the innate immune system and natural killer cells. *Adv Immunol*. **122**, 91–128 (2014).
59. Hagerling, C., Casbon, A. J. & Werb, Z. Balancing the innate immune system in tumor development. *Trends Cell Biol*. **25**(4), 214–20 (2015).
60. Sharma, P. & Allison, J. P. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell*. **161**(2), 205–14 (2015).
61. Hallmans, G. *et al.* Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort - evaluation of risk factors and their interactions. *Scand J Public Health Suppl*. **61**, 18–24 (2003).

62. Bingham, S. & Riboli, E. Diet and cancer - The European prospective investigation into cancer and nutrition. *Nature Reviews Cancer*. **4**(3), 206–215 (2004).
63. Hebels, D. G. *et al.* Performance in omics analyses of blood samples in long-term storage: opportunities for the exploitation of existing biobanks in environmental health research. *Environ Health Perspect*. **121**(4), 480–7 (2013).
64. Chadeau-Hyam, M. *et al.* Prediagnostic transcriptomic markers of Chronic lymphocytic leukemia reveal perturbations 10 years before diagnosis. *Ann Oncol*. **25**(5), 1065–72 (2014).
65. Mancuso, F. M. *et al.* HumMeth27QCReport: an R package for quality control and primary analysis of Illumina Infinium methylation data. *BMC Res Notes*. **4**, 546 (2011).
66. Chatziioannou, A. *et al.* Gene ARMADA: an integrated multi-analysis platform for microarray data implemented in MATLAB. *BMC Bioinformatics*. **10**, 354 (2009).
67. Koponen, J. *et al.* Determination of selected perfluorinated alkyl acids and persistent organic pollutants from a small volume human serum sample relevant for epidemiological studies. *Journal of chromatography. A*. **1309**, 48–55 (2013).
68. Heiberger, R. M. HH: Statistical Analysis and Data Display: Heiberger and Holland. R package version 3.1–24 (2015).
69. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. **13**, 86 (2012).
70. Grimm, F. A. *et al.* Metabolism and metabolites of polychlorinated biphenyls. *Crit Rev Toxicol*. **45**(3), 245–72 (2015).
71. Champely, S., *et al.* Helios pwr: Basic Functions for Power Analysis. R package version 1.1–3 (2015).
72. Davis, A. P. *et al.* The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res*. **45**(4), 972–78 (2017).
73. Kamburov, A. *et al.* The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Research*. **41**(D1), D793–D800 (2013).
74. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43**(7), e47 (2015).

Acknowledgements

This work was supported by the EnviroGenomarkers and EXPOsOMICS European projects (Grants Number 226756 and 308610).

Author Contributions

A.E.P. wrote the main manuscript text with the help of T.K., J.K., S.K. and M.B., A.E.P. prepared figures and tables. A.E.P. performed the statistical analysis and M.C. contributed to the development of the statistical models. D.G.A.J.H. generated gene expression raw data. H.K. and P.R. provided exposure data. I.A.B. and D.P. coordinated the bio-banks. P.G., F.S., A.J. and all other authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-37449-y>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019