



## The exposome in practice: Design of the EXPOsOMICS project



P. Vineis<sup>a,\*</sup>, M. Chadeau-Hyam<sup>a</sup>, H. Gmuender<sup>b</sup>, J. Gulliver<sup>a</sup>, Z. Herceg<sup>c</sup>, J. Kleinjans<sup>d</sup>, M. Kogevinas<sup>e</sup>, S. Kyrtopoulos<sup>f</sup>, M. Nieuwenhuijsen<sup>e</sup>, D.H. Phillips<sup>g</sup>, N. Probst-Hensch<sup>h</sup>, A. Scalbert<sup>c</sup>, R. Vermeulen<sup>i</sup>, C.P. Wild<sup>c</sup>, The EXPOsOMICS Consortium<sup>1</sup>

<sup>a</sup> MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, UK

<sup>b</sup> Genedata AG, Basel, Switzerland

<sup>c</sup> International Agency for Research on Cancer (IARC), Lyon, France

<sup>d</sup> Maastricht University, Department of Health Risk Analysis and Toxicology, Maastricht, Netherlands

<sup>e</sup> Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain

<sup>f</sup> National Hellenic Research Foundation, Institute of Biology, Medicinal Chemistry and Biotechnology, Division of Organic and Medicinal Chemistry, Athens, Greece

<sup>g</sup> King's College London, Analytical and Environmental Sciences Division, MRC-PHE Centre for Environment and Health, London, UK

<sup>h</sup> Swiss Tropical School of Public Health, Department of Epidemiology and Public Health, Basel, Switzerland and University of Basel, Basel, Switzerland

<sup>i</sup> Utrecht University, Institute for Risk Assessment Sciences, Environmental Epidemiology Division, Utrecht, Netherlands

### ARTICLE INFO

#### Article history:

Received 5 July 2016

Received in revised form 12 August 2016

Accepted 16 August 2016

### ABSTRACT

EXPOsOMICS is a European Union funded project that aims to develop a novel approach to the assessment of exposure to high priority environmental pollutants, by characterizing the external and the internal components of the exposome. It focuses on air and water contaminants during critical periods of life. To this end, the project centres on 1) exposure assessment at the personal and population levels within existing European short and long-term population studies, exploiting available tools and methods which have been developed for personal exposure monitoring (PEM); and 2) multiple “omic” technologies for the analysis of biological samples (internal markers of external exposures). The search for the relationships between external exposures and global profiles of molecular features in the same individuals constitutes a novel advancement towards the development of “next generation exposure assessment” for environmental chemicals and their mixtures. The linkage with disease risks opens the way to what are defined here as ‘exposome-wide association studies’ (EWAS).

© 2016 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Abbreviations:** PEM, personal exposure monitoring; GIS, geographic information system; EWAS, exposome-wide association studies; STS, experimental short-term studies; MCO, mother-child cohorts; ALTS, adult long-term studies; LUR, land-use regression; DBP, disinfection by-products; OP, oxidative potential; UFP, ultrafine particles; PM, particulate matter.

\* Corresponding author at: MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, W2 1PG, London UK.

E-mail address: [p.vineis@imperial.ac.uk](mailto:p.vineis@imperial.ac.uk) (P. Vineis).

<sup>1</sup> The EXPOsOMICS Consortium includes (in alphabetic order): Andre F.S. Amaral, Toby Athersuch, Sabrina Bertinetti, Marc Chadeau-Hyam, Leda Chatzi, Theo De Kok, Michaela Dijmarescu, Almudena Espin Perez, Mario Fernandez, Claudia Galassi, Akram Ghantous, Hans Gmuender, John Gulliver, John Henderson, Zdenko Herceg, Gerard Hoek, Medea Imboden, Pooja Jain, Debbie Jarvis, Frank Kelly, Pekka Keski-Rahkonen, Jos Kleinjans, Manolis Kogevinas, Julian Krauskopf, Soterios Kyrtopoulos, David Morley, Nahid Mostafavi Montazeri, Alessio Naccarati, Tim Nawrot, Mark Nieuwenhuijsen, Georgiadis Panos, David Phillips, Michelle Plusquin, George Preston, Nicole Probst-Hensch, Andrea Ranzi, Stephen Rappaport, Laia Font Ribeira, Lorenzo Richiardi, Susan M. Ring, Oliver Robinson, Alberto Rodriguez, Augustin Scalbert, Terrence Simmons, Martyn T. Smith, Jordi Sunyer, Sonia Tarallo, Veronique Terrasse, Ming Tsai, Erik van Nunen, Karin van Veldhoven, Roel C.H. Vermeulen, Cristina M. Villanueva, Paolo Vineis, Jelle Vlaanderen, Martine Vrijheid, Christopher P. Wild, Timo Wittenberger.

<http://dx.doi.org/10.1016/j.ijheh.2016.08.001>

1438–4639/© 2016 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Box 1: A conceptual model of life-course disease risk.**

Population studies of chronic diseases have traditionally recruited middle-aged subjects. However, there is strong evidence that (a) the risk of disease is influenced by early exposures, including *in utero*; (b) life-stages include critical (bearing long-term effects) and sensitive periods (susceptible to large effects) (Ben-Shlomo and Kuh, 2002). The idea of a sequence of *critical* and *sensitive* periods leads to the concept of “*chain of risk*”, i.e. the interplay of early exposures and late exposures. To use this concept in practice implies having access to multiple life-stages in exposure assessment and epidemiological studies, and repeated measurements of biomarkers at different time windows. This approach requires an inter-generational epidemiological study design and novel statistical analyses.

## 1. Introduction and study design

It is generally accepted that the majority of important chronic diseases are likely to result from the combination of environmental exposures to chemical and physical stressors and human genetics. There is also evidence that the effects are location-specific and influenced by climatic, lifestyle and socioeconomic characteristics. Although information on both environmental and genetic causes of disease is growing as a result of large-scale epidemiological research, exposure data (including diet, lifestyle, environmental and occupational factors) is often fragmentary (in time and depth), non-standardized, at crude resolution and often does not include estimates at the level of the individual. The information on environmental factors is often incomplete or inaccurate and the subsequent estimation of overall risks associated with these factors is severely hindered. As a result, important associations can go undetected. This limitation has recently been framed within the context of the exposome, the environmental counterpart of the genome. The concept of the exposome refers to the totality of environmental exposures from conception onwards, and has been described in detail elsewhere including its external and internal components (Wild, 2005; Rappaport and Smith, 2010; Anon, 2016; Vineis et al., 2009).

The context of EXPOsOMICS is the rapidly developing field of exposure assessment, including the use of omic technologies, according to the concept of the exposome. Historically, a key event was the publication of the report *Toxicity Testing in the 21st Century: A Vision and a Strategy* by the US National Research Council in 2007. The primary goals of the report were, among others: “(1) to provide broad coverage of chemicals, chemical mixtures, outcomes, and life stages, (2) to reduce the cost and time of testing, (3) to develop a more robust scientific basis for assessing health effects of environmental agents” (<http://www.nap.edu/catalog/11970/toxicity-testing-in-the-21st-century-a-vision-and-a>). EXPOsOMICS aims at partially covering points 1 (mixtures, outcomes, and life stages) and 3.

By comprehensively addressing the integration of the external and the internal exposomes at the individual level, EXPOsOMICS provides a holistic and consolidated approach to exposure science. Building upon several EU-funded research projects with rich sets of health data, exposure data, biomarker measurements and publicly available data sources, this multidisciplinary project:

- Pools and integrates information from short-term, experimental human studies and long-term epidemiological cohorts/consortia – including adults, children and newborns – to enable focused investigations to refine environmental exposure assessment based on the concept of *life-course epidemiology* (Box 1 and Fig. 1).

- Characterizes the exposome, by (a) measuring the external component of the exposome at different critical life stages by employing novel tools and drawing on experience gained in existing EU initiatives (PEM, databases coupled with GIS, remote sensing), with a focus on air and water pollution; and (b) measuring biomarkers of the internal exposome (xenobiotics and metabolites), using omic technologies (adductome, metabolome, transcriptome, epigenome, proteome).

- Integrates external and internal exposure measures to comprehensively model and assess exposure to air pollution and water contamination in large population cohorts, through novel statistical modeling.

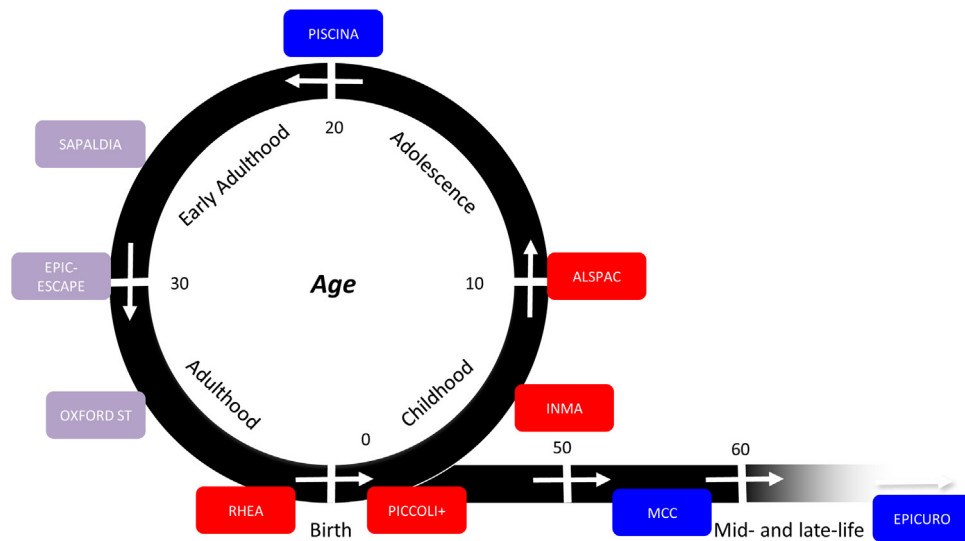
Together these approaches will lead to formulating a new concept of integrated exposure assessment at the individual level, reducing uncertainty, and assessing how these refinements influence disease risk estimates for combined, multiple exposures and selected diseases.

The scientific questions to be addressed by the project are:

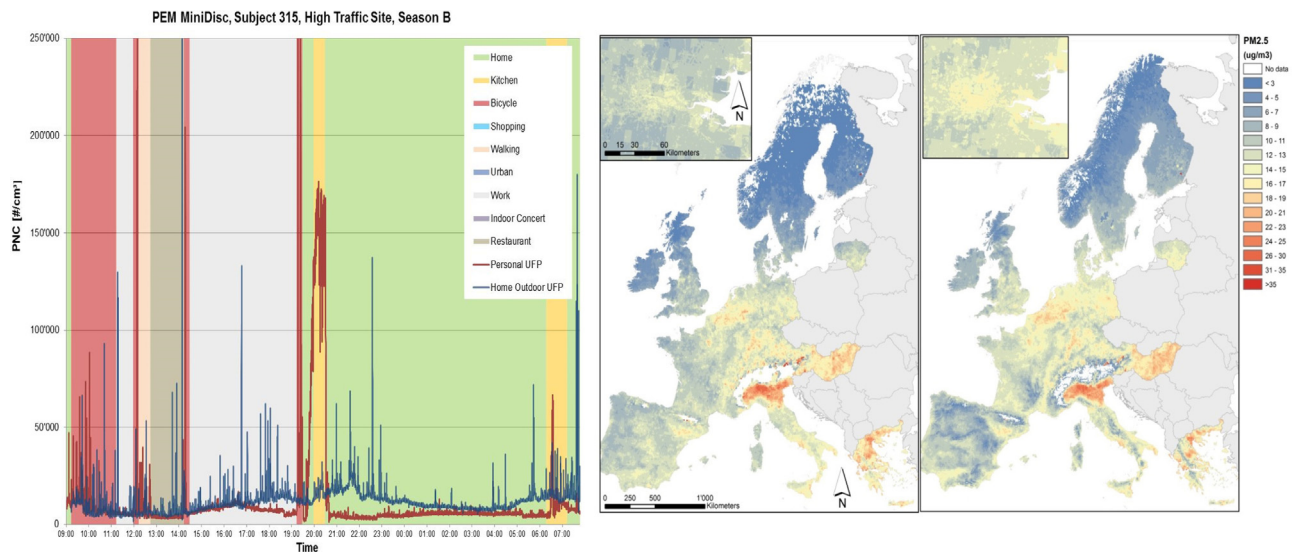
- (1) Is it possible to refine exposure assessment to air pollution and water contaminants using a combination of personal exposure monitoring and omic technologies?
- (2) Will that refinement lead to more accurate estimates of the association with selected diseases, by reducing measurement error?
- (3) Do new approaches allow the investigation of the effects of mixtures in addition to single components?
- (4) Do they improve the investigation of dose-response relationships?
- (5) Is it possible to strengthen causal reasoning by using the “meet-in-the-middle” concept, i.e. investigate the temporal sequence of exposure, biological pathway perturbation and disease onset?
- (6) Is it possible to use the exposome approach to study the life-course epidemiology of environmental diseases?

## 2. External exposome (Fig. 2)

Existing population-based studies on environmental exposures, such as the ESCAPE project (<http://www.escapeproject.eu/>), are providing essential information on the disease risks associated with exposure to air pollution, but they are limited in their exposure assessment methodologies and because of the uncertain biological plausibility of associations that are detected. Well-conceived, fully controlled short-term intervention studies, such as the Oxford Street Randomized Trial (included in the present project, ref. (McCreanor and Cullinan et al., 2007 Dec 6)), and TAPAS have shown that acute changes can occur in lung and heart function at low or very low levels of exposure to air pollutants. Analogous studies have been conducted on water contaminants, where genotoxic effects have been observed (see below). The identification of *long-term* effects, however, has been problematic, due to the lack of the same degree of accuracy in exposure assessment as obtained in *short-term* experimental studies. The first innovation of EXPOsOMICS is in bringing together *both types of investigations* and in linking Personal Exposure Monitoring (PEM) (high precision, used mostly in short-term panel studies) with up-to-date land use regression modeling and satellite-based exposure assessment. GPS-based techniques, smartphones, accelerometers and PEM to detect pollutants can provide accurate and instant estimates of changes in human exposure and in physical activity (de Nazelle et al., 2011; Dons et al., 2011; Dippold, 2006). In addition, satellite modeling (e.g. aerosol optical depth) is used for several purposes, including extension of exposure models to areas not covered by other sources of information, and retrospective assessment. EXPOsOMICS utilises and further



**Fig. 1.** The conceptual framework of Exposomics (life-long integration of cohorts).



**Personal monitoring for UFP in Basel  
(courtesy of N Probst-Hensch)**

**“Hybrid LUR” models including satellite data  
(AOD) and data from chemical transport  
models developed using ESCAPE (A) and  
AIRBASE (B) sites (courtesy J Gulliver)**

**Fig. 2.** External exposome: from personal monitoring to satellite integration.

develops these technological advances in a combined manner for the assessment of personal exposures, and to expand the assessment to exposures that have not been geo-spatially addressed in Europe to date (e.g. ultrafine particulates). By integrating technologies that provide estimates of exposure to multiple chemicals over relatively prolonged time-periods, at the level of the individual subject as well as populations, and at different life stages, our project systematically explores and further develops the concept of the exposome.

### 3. Internal exposome

Biomarkers of exposure that were developed during the past couple of decades (e.g. DNA and protein adducts), while repre-

senting a significant step towards the definition of more accurate measures of personal exposure, still suffer from the disadvantage of addressing single chemicals and covering therefore only a small fraction of the exposome. However, adaptation of biomarker technologies to provide a more global estimate of the internal component of the exposome is emerging and is further developed within EXPOSOMICS. For example, high-resolution LC/MS (Liquid Chromatography/Mass Spectrometry) metabolomics now permits the detection and characterisation of large numbers of small molecules, typically with molecular mass less than 2000 Da, in biological fluids. This includes any metabolites that can be affected by environmental exposures, and that can be related to inflammation, oxidative stress and various metabolic pathways. Analogous adductomic technologies will be pioneered to detect

### Box 2: Basic components of the EXPOsOMICS project. Phase 1.

1. We have selected and integrated subjects, samples and data from three types of existing studies: Experimental Short-Term Studies (STS), Mother-Child Cohorts (MCO) and Adult Long-Term Studies (ALTS), between them reflecting all life stages from conception to old age.

2. We have measured the external exposome component for air and water contaminants by performing extensive, repeated Personal Exposure Monitoring (PEM). Fresh blood samples have been collected from all the individuals undergoing PEM, i.e. individuals in STS and those from a representative subsamples from MCO and ALTS, selected on the basis of contrasting exposures as estimated by traditional exposure assessment methods. In these samples, plus already stored samples of the recalled subjects from MCO and ALTS, EXPOsOMICS has conducted untargeted omic analyses. The aim is to look for new biomarkers of exposure to chemicals or mixtures and evaluate intra-individual variation of the internal exposome.

### Phase 2.

3. OMIC profiles were also measured in approximately 2000 stored samples from MCO and ALTS, using untargeted analytical methods, with the aim of evaluating them as predictive of risk by examining their association with health effects, and also generating new hypotheses on disease etiology.

4. We combine external and internal exposome data, Land Use Regression models (LUR) and satellite data to calibrate air pollution exposure estimates (e.g. PM<sub>2.5</sub>, ultra-fine particles, black carbon) obtained using traditional methods in MCO/ALTS and use these refined estimates for risk assessment and burden of disease evaluations.

protein adducts in an untargeted fashion, to provide a global picture of biomarkers of individual exposure either to electrophiles or chemicals that are metabolically activated to electrophiles. Because they can directly modify DNA and important proteins, reactive electrophiles are important constituents of the exposome. Because of their greater abundance and residence times in human blood, adducts on the circulating proteins hemoglobin and human serum albumin (HSA) are preferable to those of DNA and glutathione for characterizing adductomes (Rappaport et al., 2011).

In addition to global data on internal markers of exposure, provided by these technologies, the internal component of the exposome includes further biological profile data generated by high-density biological analysis technologies. Transcriptomic, epigenomic and proteomic profiles of biological samples provide a detailed picture of the evolving state of cells under the influence of environmental chemicals, thus revealing early mechanistic links with potential health effects (Chadeau-Hyam et al., 2011; Lan et al., 2004; Smith et al., 2005).

## 4. Population studies

The long-term cohorts used in EXPOsOMICS have been selected on the basis of the following criteria: access to stored biological samples, residential histories (for GIS based models), information on confounders and ability to recall participants. Large studies have been selected that cover wide geographic areas and have a wealth of information including biomarkers, geo-referencing, follow-up data, and extensive behavioural and environmental variables. As it is hypothesized that many internal exposome signatures are influenced by external exposures beyond air and water pollutants, study populations have been selected with available and accurate data on lifestyle, anthropometry, nutritional and environmental factors.

The basic components of the project are described in Box 2.

Through the integration of a total of 14 studies of different types, located in different regions, the project has generated a substantial amount of new data of continent-wide relevance. Phase 1 involves refined exposure assessment based on the external (Personal Exposure Monitoring) and the internal (omics) components of the exposome, as well as (for comparison) exposure assessment using traditional methods. This is performed in subjects participating in short-term experimental studies (STS) and in sub-sets of representative subjects from population cohorts. The latter include mother-child cohorts (MCO) and adult long-term studies (ALTS). Recalled subsets have been chosen at the extremes of exposure (as estimated by traditional methods). For recalled subjects we have conducted omic analyses on both freshly collected and stored samples, thus providing repeated measurements of the internal exposome. In Phase 2, omics are tested for their association with health effects using additional untargeted analyses in samples from ALTS and MCO, according to a case-cohort study design. In addition, the refined exposure estimates derived from the external and internal components of the exposome, which take into account individual variation, are compared to estimates derived from traditional, less accurate exposure assessment methods, thus leading to a calibration of estimates of exposure in long-term studies. Subsequently the association of calibrated exposure estimates with selected diseases is assessed.

Advantages of the chosen study design are that it makes it possible to:

- Link short-term experimental studies, having controlled exposure conditions, with long-term studies in the same geographic areas;
- Address the effects of common exposures at different life stages;
- Address individual variability (in space and in time) through repeated Personal Exposure Monitoring and omic measurements in repeat blood samples;
- Incorporate individual variability and biomarkers of exposure into exposure estimates and risk assessment for selected diseases, thus reducing uncertainty in risk measures.

## 5. Experimental studies, short term effects (STS)

Short-term experimental studies on air pollution (Oxford Street Randomized Trial, ref. (McCreanor and Cullinan et al., 2007); and TAPAS) are utilized. In these studies volunteers are exposed to high or low levels of air pollution within different urban contexts. During each session, a number of outcome measurements (lung and heart function) have been made while simultaneously measuring PM<sub>10-2.5</sub>, PM<sub>2.5-0.1</sub> and PM<sub>0.1</sub>. In addition to these measurements, in the EXPOsOMICS project adductomics, metabolomics, transcriptomics, and epigenomics are applied to evaluate the internal exposome on blood samples collected before and after the exposure periods, to search for biomarkers associating with acute exposures (PEM) and acute health effects.

Exposure to water contaminants and their short-term effects has been evaluated in the ongoing PISCINA experimental study of swimmers in pools, where individuals have high exposure to water disinfection by-products (DBPs) through skin absorption and inhalation. Studies on swimmers have been found to be a good model for high-exposure to DBPs in drinking water in the general population. Existing PISCINA samples (ref. (Kogevinas et al., 2010; Richardson et al., 2010)) have been complemented, in this project, with additional samples collected from swimmers in pools with different levels of water DBPs. The external and internal exposome have been measured in these individuals before and after swimming and complemented, for purposes of phenotypic anchoring



(i.e. of correlating with phenotypic effects) (Kogevinas et al., 2010), with measurements of micronuclei in blood and urine, bacterial mutagenicity (of urine samples), and inflammation and oxidative stress markers.

## 6. Observational life-course studies: mother-child cohorts (MCO) and adult long-term studies (ALTS)

In Phase 1, measurements of the internal and external exposome have been made in representative subjects of (a) five birth cohorts (RHEA, ENVIRONAGE, INMA, ALSPAC and PICCOLI+), (b) the EPIC, ECRHS and SAPALDIA adult cohorts, exploiting data generated within the large ESCAPE consortium for air pollution, and (c) a study on colon cancer (MCC-Spain) used for the investigation of water pollution.

These data are utilized for in-depth studies aimed at the development of improved exposure estimates based on the external and internal exposome concept in a chronic exposure context. Some of the above cohorts have stored samples of venous blood collected at different times of follow-up and most of the birth cohorts have cord blood. We reconstruct the *in utero* exposome based on cord blood samples integrating different potential critical exposure windows of air pollution exposure. Epidemiological and public health significance of the *in utero* exposome will be investigated based on associations with child health and development including birth weight, growth, obesity and other outcomes.

## 7. Exposures

### 7.1. Air pollution

Epidemiological studies have traditionally assigned exposure based on monitoring results of a central monitoring site in the city of residence (Jerrett et al., 2005). In the last decade, methods for more individual exposure estimates have been applied, including land use regression models (LUR) and dispersion models (Jerrett et al., 2005; Hoek et al., 2008). A number of studies have shown differences between long-term average personal and ambient air pollution exposure (Van Roosbroeck et al., 2005). In the latter studies, integrated personal samplers were used that provided a single sample over 1–2 days. Hence no assessment of the contribution of different microenvironments to total exposure (and difference) was possible.

Recent advances in air pollution monitoring technology mean that personal monitoring can be undertaken with real-time monitors, capable of logging minute-to-minute variability in exposures. Sensors are portable and sensitive. Coupling the new air pollution monitors with GPS and accelerometry from smartphones, and information on personal activity patterns, the potential now exists to differentiate between exposure during journeys and fixed-site locations and to assess the contribution of different micro-environments to the magnitude and variability of environmental exposures. However, no product existed at the outset of EXPOsOMICS that integrated these technologies into a single personal monitoring system that can be deployed with lay users (e.g. cohorts), a problem that has been addressed by the current project (see below).

To date, the majority of evidence about the health effects of air pollution from epidemiology and controlled exposure studies is based on particle mass as the measure of exposure. Consequently, regulatory agencies have adopted mass-based ambient air quality standards. Yet particulate matter (PM) is a complex mixture, and particles of different size and composition are likely to have different toxic effects. The EU-funded ESCAPE study has comprehensively characterized sources of outdoor air pollution and developed ambi-

ent LUR models for PM<sub>10</sub>, PM<sub>2.5</sub> and NO<sub>2</sub>. Models are currently in development for elemental composition (XRF; X-ray fluorescence), EC/OC (elemental carbon/organic carbon) and PAHs (polycyclic aromatic hydrocarbons). There is, however, a need to develop models for ultrafine particles for which the long-term health effects have been poorly studied because of difficulties in exposure assessment. This is now possible using an innovative mobile monitoring design that has been shown to be reliable and cost-effective in recent studies (Klompaker et al., 2015; Montagne et al., 2015).

One of the properties of particles likely to reflect toxicity is their oxidative potential (OP). By analyzing the spatial and temporal variability of the OP of particulate matter collected on filters, the determinants of that variation are characterized, and new, spatially resolved air pollution models for OP have been developed. The air pollution models alone, however, only provide information on ambient outdoor pollutant concentrations. Recent advances in GIS (e.g. route modeling) (Gulliver and Briggs, 2005) and micro-environmental models (e.g. indoor-to-outdoor), have led to the development of more detailed, personal exposure models which can be fed by rich data sources on detailed population time-activity patterns.

To advance the state of the art EXPOsOMICS provides an innovative framework for air pollution external exposome, *via* the following steps:

- 1) The scientific partners have integrated instruments from equipment manufacturers (e.g. DiSCmini for UFP and BGI pump for PM<sub>2.5</sub>) to develop a novel integrated personal monitoring system comprising UFP and PM<sub>2.5</sub> personal air monitors with smartphone technology aimed at characterization of micro-environments, activity patterns, and inhalation rates. One of the main technological innovations is the collection of various measurements and the ability to access them via a single source (i.e. smartphone).
- 2) Deploy the new personal monitoring system among cohort members in a subset of five study areas, covering different sites (e.g. city centre, suburban, industrial, and rural), to collect the largest series of detailed personal exposure measurements of UFP, PM<sub>2.5</sub> and activity data in Europe to date (with simultaneously collected blood samples for omics).
- 3) Develop models in longitudinal studies such as ESCAPE (<http://www.escapeproject.eu/>), enriched for all the cohorts in the study, and develop methods to transfer these models of air pollution concentrations back in time.
- 4) Undertake an UFP air pollution “mobile monitoring campaign” in the study areas where we also perform PEM to develop and validate the new land-use regression models for UFP.
- 5) Apply the OP depletion analysis technique to extend PM metrics to look at OP. PM<sub>2.5</sub> filters collected during fixed-site outdoor monitoring in selected ESCAPE areas have been analysed to detect the spatial and temporal variability in PM<sub>2.5</sub>-related OP and subsequently create new land-use regression models for OP. We compare and assess the spatio-temporal differences in exposure estimates between OP with those from traditional particulate metrics (e.g. PM<sub>10</sub>, PM<sub>2.5</sub>).
- 6) Compare PEMs in 3) above with those from residential address locations to quantify the correlation between the two. We further quantify the spatial and temporal micro-environmental contributions to total exposures.
- 7) Assess the potential for exposure variability (or misclassification) in exposures, e.g. for UFP and PM<sub>2.5</sub> where we have both LUR models and measured exposures. Compare traditional (i.e. residential address) and new methods to inform the *omic* studies (see below) and health risk assessment.
- 8) Develop a new Europe-wide air pollution model for PM<sub>2.5</sub> and methods to apply exposure models for the other pollutants (OP,

UFP) to other cohorts and across Europe for health risk assessment. In order to develop exposure models for the pollutants being studied in the external exposome, which can be applied to health risk assessment, the transferability of the new LUR models and hybrid models to other countries/regions and time periods is investigated. Develop models which incorporate satellite data either as variables to enhance existing approaches or as a means of calibrating/validating models in areas where routine monitoring data are sparse or do not exist.

- 9) Provide model estimates for single and multiple exposures (i.e. mixtures) to feed into risk models (e.g. partial least-square regression, ridge regression, Bayesian mixture methods) (Chadeau-Hyam et al., 2013a; Agier et al., 2016a) to study the contribution of single compounds and combination of compounds to adverse health effects in children and adults.

## 7.2. Water contamination

EXPOSOMICS makes use of existing short-term experimental studies (STS) and of long-term population studies (MCO and ALTS). For the short-term scale (PISCINA study) direct measurements of specific chemicals and mixtures in water have mainly been used, while for the life-course studies a combination of models with measurements for validation purposes, smart phone technology, extensive use of databases, remote sensing and land use maps are used. European-based estimates have been obtained further through systematic use of regional and national databases for disinfection by-products (DBPs).

### 7.2.1. Personal and micro-environmental exposure measurements (Fig. 2)

In the PISCINA swimming pool study, external exposure measurements include determination of an expanded range of disinfection by-products (DBP) (trihalomethanes, haloacetic acids, MX, chloramines, haloacetonitriles), overcoming traditional approaches that measure only trihalomethanes. These chemicals have been measured in air and water and/or in biological samples such as exhaled breath (e.g. trihalomethanes) and urine (haloacetic acids) from study subjects. The external exposome is complemented with omic analyses and supported (for phenotypic anchoring) by *in vitro* assays such as the *Salmonella* (Ames) mutagenicity test (Maron and Ames, 1983) and mammalian cell chronic and acute cytotoxicity and the SCGE (comet) assay (van Leeuwen et al., 2006). Biological samples (blood, urine, exhaled breath condensate) have been obtained immediately before and after swimming in the pool. Repeated biological samples have been collected after swimming to cover different formation/elimination kinetics of biomarkers, thus contributing to an improved quantitative evaluation of internal exposures.

### 7.2.2. Long-term exposure assessment

In the mother-child cohorts, external exposure measurement includes determinations of a range of DBP chemicals in drinking water (trihalomethanes, haloacetic acids, MX, haloacetonitriles). Measurements for some of these chemicals are available from the EU-funded HiWate project ([https://www.researchgate.net/publication/24037941\\_Health\\_Impacts\\_of\\_Long-Term\\_Exposure\\_to\\_Disinfection\\_By-Products\\_in\\_Drinking\\_Water\\_in\\_Europe\\_HiWate](https://www.researchgate.net/publication/24037941_Health_Impacts_of_Long-Term_Exposure_to_Disinfection_By-Products_in_Drinking_Water_in_Europe_HiWate)). In the colorectal cancer study (MCC), exposure modeling of DBPs is based on the evaluation of lifetime residential history together with the collection of historical information on DBPs in the relevant regions and water toxicity testing from the short-term studies.

### 7.2.3. European-wide water-borne exposure models for health risk assessment

Exposure to specific DBPs is derived for European populations from routinely collected information in each country. These data, not centrally available in the EU, come from focus contact points in each country (expanding work completed in INTARESE, and HiWate). In addition a theoretical framework for future routine evaluation of water contaminant information in Europe is proposed.

## 7.3. Internal exposome: omics

By permitting the simultaneous analysis of large numbers of potential targets without recourse to prior hypotheses, omic technologies provide unique opportunities for the discovery of a new generation of biomarkers of exposure and disease risk, significantly enriched by mechanistic information (Chadeau-Hyam et al., 2011; Lan et al., 2004; Smith et al., 2005; Perera et al., 2009; Zhang et al., 2016a; Zhang et al., 2016b; Bictash et al., 2010; Nicholson and Rantalainen et al., 2011; Wang-Sattler and Yu et al., 2008; Holmes et al., 2008; Calderón-Garcidueñas et al., 2004). The EXPOSOMICS project both advances omic technologies that are related to identification and quantification of environmental exposure (metabolomics, adductomics) and applies multiplex omic technologies that measure downstream effects of environmental exposures (epigenome, transcripts, miRNA, proteins). Omic technologies are deployed in an untargeted manner in both the short-term and the long-term studies, to a total of approximately 3000 samples (Table 1). The subjects for untargeted omics in Phase 1 were selected on the basis of extreme exposures. Publicly available toxicogenomics databases have been explored to identify candidate omic-based exposure biomarkers. Databases are the diXA data infrastructure (FP7 project: diXA) and the US Comparative Toxicogenomics Data Base (<http://ctdbase.org/>). These contain both *in vitro* and *in vivo* toxicogenomic data and, to a degree, omic results from exposed humans, and have links to important chemical toxicity databases such as the US NTP and the OECD eChemPortal. We will attempt to replicate the signals of Phase 1 in selected case-cohort studies for a total of approximately 2000 subjects. One of the main products of the consortium will be endophenotypes, i.e. intermediate phenotypes common to several diseases.

Untargeted omic analyses in general are conducted in subjects for whom the external exposome has been measured and, where possible, all omic platforms are used in the same subjects to enable cross-omic analyses.

### 7.3.1. Metabolomics (MS technology) (Fig. 3)

Plasma or serum samples have been analysed by high resolution mass spectrometry (MS) coupled to UPLC. Metabolic features characterizing exposed groups are identified by multivariate statistics with appropriate correction for the False Discovery Rate (FDR), and efforts to identify the discriminating metabolites are made. A database of known biomarkers from the main environmental contaminants in air and water as measured in various populations has been compiled from the scientific literature into the Exposome-Explorer database (<http://exposome-explorer.iarc.fr/>).

### 7.3.2. Adductomics

Application of adductomics to the exposome concept involves an untargeted investigation of the internal exposome based on the measurement of complete categories of features (e.g. protein adducts). The untargeted approach has been realised by focusing on a specific locus of HSA, and using MS of hydrolysis products to profile covalent modifications over a range of masses.

**Table 1**  
Characteristics of the cohorts.

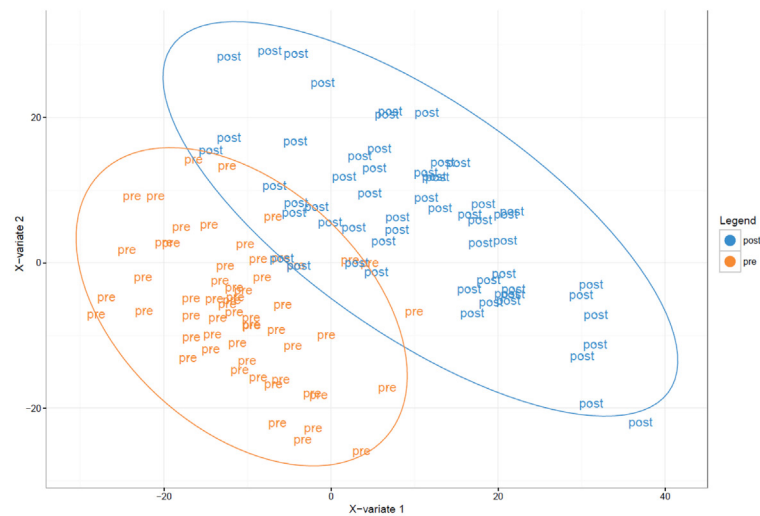
Type of study	Cohort	Study Characteristics
<b>Experimental Studies</b>	<b>Oxford Street</b>	The study is a randomized cross-over trial <b>on the acute effects of ambient exposure to traffic</b> and exacerbations of cardiovascular and chronic pulmonary disease. 59 adults were recruited for the EXPOsOMICS project: 20 healthy adults, 20 COPD (Chronic Obstructive Pulmonary Disease) and 19 IHD (Ischaemic Heart Disease). They were invited to walk in Oxford Street (high exposure) or in Hyde Park (low exposure), London. A series of clinical, physiological and inflammatory responses are compared between the two exposures and to those measured in healthy controls undergoing similar exposures, as well as various omics measurements (transcriptomics, metabolomics, adductomics) obtained before, 4 h after and 24 h after walking.
	<b>TAPAS Barcelona</b>	The <b>TAPAS 2</b> Barcelona study follows a crossover study design involving 30 healthy, non-smoking adults in the age range of 18–60. Volunteers were exposed from 8am to 10am to either heavily polluted air (busy street in morning rush hour, traffic site) or to low dense traffic air pollution (control site) in Barcelona, Spain. Six subjects were studied simultaneously on each study day: three of them performed moderate physical activity, riding an ergometer (intermittent 15 min cycling and 15 min break) while the other three volunteers stayed at rest. Participants that performed physical activity had to cycle at such a pace that their heart rate was 50–70% of their maximum rate.
	<b>PISCINA2</b>	The experimental PISCINA2 study – consisting of 60 healthy non-smoking volunteers – assessed a) exposure to chlorination by products in swimmers, b) short-term respiratory health effects and c) genetic damage after swimming. Blood, urine, and exhaled breath condensate (EBC) were collected from the volunteers before and immediately after swimming in a chlorinated pool for 40 min. Transcriptomics, metabolomics, adductomics and concentrations of selected inflammatory markers were measured in blood samples obtained before and 2 h after the experiment.
<b>Mother-Child Cohorts</b>	<b>RHEA</b> <b>Mother-Child</b> <b>Study in Crete</b>	The <b>mother-child “Rhea”</b> study in Crete is a prospective cohort examining a population sample of pregnant women and their children, at the prefecture of Heraklion (n = 1500). The study aims are to evaluate a) nutritional, environmental, biological and psychosocial exposures in the prenatal period and in early childhood, b) the association of these exposures with the development of the foetus and the child, c) mother's health during and after pregnancy, and d) genetic susceptibility and the interactions between genetic and environmental factors affecting child health. A set of 100 children from the RHEA-cohort is included in the EXPOsOMICS children studies, for which data on cord blood DNA-methylation, metabolomics, proteomics and albumin adducts are available.
	<b>INMA</b> <b>Environment and</b> <b>Childhood</b>	The <b>INMA – Infancia y Medio Ambiente</b> (Environment and Childhood) project follows up a population sample of around 3100 pregnant mothers and newborns. New and existing cohorts of pregnant women have been incorporated from seven different Spanish regions. Pregnant women are assessed at 12, 20 and 32 weeks of gestation to collect information about environmental exposures and foetal growth. Children are assessed at birth, at the age of 1 year and at the age of 4 years. The cohorts have been designed to evaluate the impact of environmental exposures and diet on children's health. A set of 100 children from the INMA centre Sabadell is included in the EXPOsOMICS children studies, for which data on cord blood metabolomics, proteomics and albumin adducts are available.
	<b>ALSPAC</b> <b>The Avon</b> <b>Longitudinal Study</b> <b>of Parents and</b> <b>Children</b>	<b>ALSPAC</b> recruited more than 14,000 pregnant women with estimated dates of delivery between April 1991 and December 1992. Outcomes include prenatal and birth health events, growth, neurodevelopment, behavioural functioning, lung function and cardiovascular measurements. Nitrogen oxide levels and PM <sub>10</sub> estimates based on home location are available for the <i>in utero</i> and early life period. A set of 2300 children from the ALSPAC study is included in the EXPOsOMICS children studies, for which data on metabolomics at 7 years of age is available. DNA methylation for 650 children is available in cord blood and at 7 and 15 years of age. Also information about growth trajectories, asthma/lung function and neurocognitive tests are available.
	<b>Piccoli+</b>	<b>Piccoli+</b> is a <b>multicentric Italian birth cohort</b> that recruited 3000 new-borns and their mothers in 5 centres: Turin, Trieste, Florence, Viareggio and Rome (www.piccolipi.it). Participants receive a follow-up interview 6, 12 and 24 months after delivery and a medical examination when the children turn 4. Growth trajectories and neurocognitive test-results are available. A set of 99 children from the Turin centre is included in the EXPOsOMICS children studies, for which data on cord blood DNA methylation, metabolomics, proteomics and albumin adducts are available.
	<b>EnvironAge</b> <b>ENVIRonmental</b> <b>influence ON</b> <b>AGEing in early life</b>	The <b>ENVIRonAGE (ENVIRonmental influence ON AGEing in early life)</b> cohort includes 1300 mother-infant pairs and further recruitment is ongoing. Data include mothers' lifestyle and socio-economic status, gestational history, measurements including the new-borns' blood pressure (all healthy with gestational age 37–42 weeks), bio banked placental tissue and cord blood including RNA/DNA, toxic metals in cord blood and placenta, and <i>in utero</i> and early life exposure to fine particulates and NO <sub>2</sub> using a spatial temporal interpolation method. A set of 200 children from the ENVIRonAGE-cohort is included in the EXPOsOMICS children studies, for which data on cord blood DNA methylation, transcriptomics, metabolomics, proteomics and albumin adducts are available.
<b>Adult Cohorts and</b> <b>Case-Control Studies</b>	<b>SAPALDIA</b>	The <b>SAPALDIA Cohort</b> was recruited at baseline in 1991 as random samples from population registries of 8 different communities. The aim is to investigate the various determinants – including environmental and genetic – of asthma in adults. Exposure modeling included area specific outdoor models of the key pollutants (UFP, PM <sub>2.5</sub> , PM <sub>10</sub> , NO <sub>2</sub> and PM <sub>2.5</sub> chemical constituents including soot, trace metals and inorganic ions). Blood and DNA samples were collected into a biobank at the first and second follow-up and the cohort has genome-wide data on over 4000 participants. Sapaldia provides 200 asthma cases and their matched controls to EXPOsOMICS. Matching is by smoking (never/time since quitting), gender, age groups, season of recruitment, year of recruitment in the cohort.
	<b>European</b> <b>Community</b> <b>Respiratory Health</b> <b>Survey</b> <b>(Norwich and</b> <b>Barcelona)</b>	The <b>European Community Respiratory Health Survey (ECRHS)</b> included sites in East Anglia and in Barcelona. In 1991/2 random samples of adults aged 20–44 were invited to complete a short questionnaire asking about symptoms of asthma (n = 3000 in each centre). Random subsamples (plus additional cases with symptoms suggestive of asthma) underwent extensive clinical investigation including extended questionnaires on allergic and lung symptoms, use of health services and indoor/occupational exposures. Spirometry (baseline and 2 follow-ups), and information on atopy (baseline and two follow-ups), self-reported respiratory symptoms, and allergic diseases were collected. Participants in Norwich, East Anglia participated in PEM (see below). 40 asthma cases and 40 matched control participated in EXPOsOMICS from Barcelona, Spain. Matching by smoking (never/time since quitting), gender, age groups, season of recruitment, year of recruitment in the cohort.

Table 1 (Continued)

Type of study	Cohort	Study Characteristics
Personal Exposure Monitoring (PEM)	<b>EPIC CVD</b>	<b>EPIC is a multicenter prospective cohort study</b> , which recruited 518,408 volunteers from 23 centres in 10 countries (Sweden, Denmark, Norway, the Netherlands, United Kingdom, France, Germany, Spain, Italy, and Greece) between 1992 and 2000. The study population included volunteers aged mostly 25–70 years at the time of recruitment. Cancer incidence, mortality data and incident CVD events were obtained at the regional level. Cases of cardiovascular disease and controls were identified within existing cohorts of EPIC Italy (Torino n = 180 and Varese n = 208) and with archived blood samples available. In Italy, incident cases were identified via clinical records and validated by a cardiologist. Other criteria for selection: only never smokers and ex-smokers since at least 1 year, and an equal number of men and women. Matching by smoking (never/time since quitting), gender, age groups, season of recruitment, year of recruitment in the cohort.
	<b>Multi Cancer Control Study Spain (MCC)</b>	<b>MCC-Spain is a multi-case-control study</b> conducted in 11 Spanish provinces with the aim to evaluate environmental and genetic factors associated with prevalent tumours, using a common set of population based controls. Incident cases and population-based matched controls 20–85 years old have been recruited in Barcelona, Madrid, León, Murcia, Gipuzkoa, Navarra, Asturias, Granada, Huelva, Valencia and Cantabria provinces between 2007 and 2013. The cancer sites evaluated are <b>colorectal</b> , breast, prostate, gastric and lymphocytic chronic leukaemia. The study population participating in the EXPOsOMICS project includes 200 cases of colorectal cancer and 200 controls.
		PEM have been performed in subsamples of existing cohorts in five study Areas (Italy (IT n = 42), the Netherlands (NL = 44), Spain (SP = 40), Switzerland (CH = 48), and the United Kingdom (UK = 31)). In IT, NL, CH and UK adults were recruited; in SP children aged 8–12 years old. All subjects carried a backpack containing air pollution sensors and batteries, and a belt containing a smartphone (and also separate GPS and accelerometer in some areas) for 24 h to measure individual exposure to PM <sub>2.5</sub> and UFP. The study took place over a year period during three different seasons (Summer, Winter and Spring/Autumn). After each PEM session, a blood sample was collected for OMICS Analysis in all adults. To interpret the air pollution and biological measurements, a set of questionnaires was administered to the participants.



Exposure	Smallest p-value
CHCl <sub>3</sub>	5.50E-18
BDCM	1.80E-17
DBCM	5.17E-15
CHBr <sub>3</sub>	2.27E-12
TTHM	4.63E-18
BrTHM	2.96E-14



**Fig. 3.** Internal Exposome – PISCINA study – Metabolomic analysis (PLS-DA and top hits) – Exposure to disinfection by-products and their short-term effects in swimmers – Untargeted metabolomics of blood samples. N = 3761 metabolomic features detected (courtesy K Van Veldhoven).

### 7.3.3. Intermediate omic biomarkers

Additional omic technologies for the analysis of biological molecules have been used to measure the more downstream effects of environmental exposures. These include the effect on the transcriptome, the epigenome and the proteome. Measurements have been conducted using:

- Transcriptomics (Agilent platform) (number of signals per sample: 44 k)
- Epigenomics:
  - Global DNA methylation (Illumina platform) (number of signals per sample: 450 k)
  - microRNA analysis (Agilent platform) (number of signals per sample > 1000)
  - Targeted proteomics (Luminex Multianalyte Profiling platform for inflammation-related proteins) (number of signals per sample ~30).



## 8. Data analyses

### 8.1. Data processing and integration: data analysis

This project has created a dedicated data infrastructure by building on existing data repositories to avoid redundancy (at the Genedata SME). We have defined query methods, workflows for quality control, data harmonization and normalization in order to make all the data available for integrated analysis. Data integration includes five types of data: (a) data from existing databases (e.g. OMI, diXA, CTDB); (b) external exposure data from experimental/observational population studies; (c) untargeted omic data; (d) health information from cohorts, in different age groups; and (e) other exposures from cohorts (confounders, effect modifiers).

The complexity of the project design and data collected poses new methodological and analytical data challenges. Main such challenges are: developing data analytics to address multiple exposures (both external and internal) in exposome-phenome association studies; identification of statistical tools to address the hierarchical structure within external and internal exposure data (e.g. cross-omics analyses); and implementing analytical tools to address the dynamics of the exposome-phenome associations over life. A first effort has been made by comparing the performances of linear regression based statistical methods in assessing exposome-phenome associations. In an extensive simulation study using realistic empirical exposome data we have shown that correlation between exposures is indeed a major challenge for exposome research, and that current statistical methods are limited in their ability to efficiently differentiate true predictors from correlated covariates. However, some currently available methods like GUESS and DSA do provide a marginally better balance between sensitivity and FDP, although they did not outperform the other multivariate methods across all scenarios and properties examined. Future work in Exposomics is focused on extending these methods to be able to adopt the aforementioned computational complexity and to be able to incorporate external information from previous research, network ontologies, and biological pathways. Some details can be found in methodological papers that have emerged from the collaboration between EXPOSOMICS and HELIX (Agier et al., 2016b; Chadeau-Hyam et al., 2013b).

### 8.2. Exposure calibration

The multidisciplinary expertise of the EXPOSOMICS group enables the collection, in several populations, of modeled exposures based on refined models and cutting edge technologies, as well as PEM. Based on samples in which both types of data are available, we define calibration coefficients to optimize the prediction of true exposure (from PEM studies) from combinations of variables in the modeled exposure matrices. Statistical methods include classical measurement error models and their Bayesian alternatives providing calibrated exposure estimates for air and water pollutants. The resulting estimates are used within EXPOSOMICS a) for the characterization of the internal response to the external component of the exposome and b) to study the association between the calibrated exposures and disease endpoints.

### 8.3. Development, validation and investigation of internal markers of external exposures

EXPOSOMICS has provided a large range of –omic profiles in populations where PEM and calibrated exposures are also available. Such analyses rely on the adaptation, in the exposome concept, of methods employed in GWAS. Mainly we use as a benchmark univariate approaches (linear or generalized linear models) coupled to ad-hoc multiple testing correction and FDR control techniques.

Bayesian variable selection methods, typically seeking the best combination of markers (omic signals) to predict the outcome (exposure levels) are used here for the first time. In addition, these analyses provide an estimate of the variance in individual responses at similar exposure levels. Finally, in order to identify potential common patterns in the internal signature of exposures across platforms, the correlations between profiles from each platform are analysed with respect to the external exposures. These ‘cross-omic’ analyses potentially identify features of the biological pathways linking exposure to disease risk, and rely on well-established methods such as network and clustering methods.

### 8.4. Risk analyses

The calibration algorithms and estimates of measurement error and inter-individual differences in biological responses are used to perform updated risk analyses in the short- and long-term effect studies, using standard parametric and non-parametric risk functions and advanced risk analysis procedures that account for the contributing effects of often-correlated exposures (e.g. partial least square regression, ridge regression, and Bayesian mixture models). In the short-term studies (such as the Oxford Street Randomized Trial) EXPOSOMICS relates the external and internal exposome markers and profiles to acute outcomes such as pulmonary function measures, other patho-physiological outcomes, heart function and asthma episodes. For PISCINA, correlations between external and internal exposome markers are investigated together with genotoxicity outcomes. In the long-term studies the calibrated exposure estimates and internal exposure markers are applied to study the association between subfractions of air pollution (PMs, UFP, and oxidative stress potential), water contaminants and health outcomes in different areas of Europe. This project uses case-cohort studies within the pooled cohorts, in two ways. In Phase 2 risk estimates for selected diseases are performed, i.e. asthma (in adults and children), CVD and colon cancer in adults, and neurodevelopment in children, for a total of approximately 1100 cases and 1100 controls. For these subjects the consortium has complete data on targeted analyses, PEM-calibrated exposure assessment, and well-phenotyped disease outcomes. We have performed untargeted omic analyses in Phase 2, to allow us to generate new hypotheses on the etiology of the selected health outcomes, but also to test whether the omic candidates selected from Phase 1 (associated with exposures) predict the risk of disease. Subsequently, these analyses are extended to risk analyses of groups of 500–1000 cases for major diseases that have been associated with air pollution or water contamination, and the corresponding controls. In this enlarged risk estimation exercise, exposure calibration based on Phases 1 and 2 is extended to exposure assessment for a total of several thousands of subjects in Europe. Lastly, the inclusion of adult and children cohorts using the same external and internal exposome methodologies allows the investigation of age-specific responses to environmental exposures at critical life stages.

## 9. Conclusions and lessons learned

EXPOSOMICS is one of the first sizable projects on the exposome. Though the original concept of the exposome implies that the study of all life-course exposures needs to be considered, this goal is currently clearly not achievable by a single investigation. We chose therefore to focus on air pollution and water contaminants as two common sets of exposures, and to explore the life-course by linking population studies recruited at different ages. There are several challenges in the study we are undertaking, which include the uncertainty in finding meaningful omic signals from low or very low environmental exposures; the ability to investigate mix-

tures (i.e. leading to effects that are different from the sum of single exposures); the ability to investigate dose-response relationships with sufficient power; the challenge of extrapolating Phase 2 findings (in nested case-control studies) to Phase 3, i.e. in larger cohort studies, after correction for measurement errors. There are also limitations in our project, namely the limited sample size and the lack of repeated samples in cohorts; this is an incentive to establish connections with similar projects and plan pooled or meta-analyses. The main lessons learned are: (a) it is difficult to harmonize several small studies in the attempt of linking age groups, and a single large cohort encompassing all ages, with long follow-up and with repeated samples would be better suited. In the current project we are taking a statistical approach that links together the different studies, treating them as phases of a life-long experience, but this approach has its own limitations (particularly power and comparability). (b) Exposure measurement technologies are rapidly evolving and such evolution needs to be monitored in the years ahead to allow a rapid incorporation into epidemiological studies. In spite of cost reduction it is likely that for a long time incorporation of the new sensors will be possible only in hundreds or thousands of subjects, thus making the approach we have used (calibration) a realistic and concrete way ahead. (c) The same applies to omic technologies, with the additional need to evaluate their reliability and applicability to samples stored from large cohorts.

The data produced by EXPOSOMICS will be made publicly available.

More than a decade after the original publication of the exposome concept, followed by years of discussions on definitions and challenges of implementing the concept into research, the time has now come to move the concept from theory to practice. EXPOSOMICS serves as a proof-of-principle of the exposome concept and through its implementation it will inform future exposome studies.

### Conflict of interest

None declared

### Acknowledgment

This work has been supported by the Exposomics EC FP7 grant (Grant agreement no: 308610) to PV.

### References

- Agier, L., et al., 2016a. *Environ. Health Perspect.*, May 24 Epub ahead of print.
- Agier, L., et al., 2016b. *Environ. Health Perspect.*, May 24 in press. <http://exposomealliance.org>.
- Ben-Shlomo, Y., Kuh, D., 2002. *Int. J. Epidemiol.* 31 (April (2)), 285–293.
- Bictash, M., et al., 2010. *J. Clin. Epidemiol.* 63 (September (9)), 970–979.
- Calderón-Garcidueñas, L., et al., 2004. *Toxicol. Pathol.* 32 (November–December(6)), 650–658.
- Chadeau-Hyam, M., et al., 2011. *Biomarkers* 16 (February (1)), 83–88.
- Chadeau-Hyam, M., et al., 2013a. *Environ. Mol. Mutagen.* 54 (August (7)), 542–557. <http://dx.doi.org/10.1002/em.21797>, Epub 2013 Aug 5. Review.
- Chadeau-Hyam, M., et al., 2013b. *Environ. Mol. Mutagen.* 54 (August (7)), 542–557.
- de Nazelle, A., et al., 2011. ISEE Proceedings, Sept.
- Dippold, M., 2006. *Personal dead reckoning with accelerometers*. Universität Bremen Mobile Research Center.
- Dons, E., et al., 2011. *Atmos. Environ.* 45 (21), 3594–3602.
- Gulliver, J., Briggs, D.J., 2005. *Environ. Res.* 97 (1), 10–25.
- Hoek, G., et al., 2008. *Atmos. Environ.* 42, 7561–7578.
- Holmes, E., et al., 2008. *Nature* 453, 396–400.
- Jerrett, M., et al., 2005. *J. Expo. Anal. Environ. Epidemiol.* 15 (2), 185–204.
- Klompaker, J.O., et al., 2015. *Sci. Total Environ.* 508 (March (1)), 266–275. <http://dx.doi.org/10.1016/j.scitotenv.2014.11.088>, Epub 2014 Dec 5.
- Kogevinas, M., et al., 2010. *Environ. Health Perspect.* 118 (11), 1531–1537.
- Lan, Q., et al., 2004. *Science* 306 (December (5702)), 1774–1776.
- Maron, D.M., Ames, B.N., 1983. *Mutat. Res.* 113, 173–215.
- McCreanor, J., Cullinan, P., et al., 2007. *N. Engl. J. Med.* 357 (December (3)), 2348–2358.
- Montagne, D.R., et al., 2015. *Environ. Sci. Technol.* 49 (14 (July)), 8712–8720. <http://dx.doi.org/10.1021/es505791g>, Epub 2015 Jul 8.
- Plewa, M.J., et al., 2008. Washington, D.C.: American Chemical Society, 36–50.
- Nicholson, G., Rantalainen, M., et al., 2011. *Mol. Syst. Biol.* 7, 525.
- Perera, F., et al., 2009. *PLoS One* 4 (2), e4488.
- Rappaport, S.M., Smith, M.T., 2010. *Science* 330 (6003), 460–461.
- Rappaport, S.M., et al., 2011. *Toxicol. Lett.*, Apr 8.
- Richardson, S.D., et al., 2010. *Environ. Health Perspect.* 118 (11), 1523–1530.
- Smith, M.T., et al., 2005. *Chem. Biol. Interact.* 153–154 (May (30)), 123–127.
- Van Roosbroeck, S., et al., 2005. *Atmos. Environ.* 41, 3381–3394.
- Vineis, P., et al., 2009. *Environ. Health* 30 (8), 54.
- van Leeuwen, D.M., et al., 2006. (August (1–2)), 12–22.
- Wang-Sattler, R., Yu, Y., et al., 2008. *PLoS One* 3 (12).
- Wild, C.P., 2005. *Cancer Epidemiol. Biomarkers Prev.* 14 (August (8)), 1847–1850.
- Wild, C.P., 2012. *Int. J. Epidemiol.* (January (31)), PubMed PMID: 22296988.
- Zhang, Y., et al., 2016a. *Ecotoxicology* 18 (6), 722–728.
- Zhang, Y., et al., 2016b. *J. Hazard. Mater.* 190 (1–3), 515–519.