Plain language summary

Environmental epidemiological research is conducted to study the different factors that cause (the distribution of) diseases within the population and is performed worldwide. Wherein the exposure to different types of air pollution is a global concern. In most cases, performing epidemiological research results in sets of genes or so called biomarkers, special molecules which can indicate or predict specific health outcomes. In environmental epidemiology these signals (genes or biomarkers) are mild and complex to get hard evidence. While performing research, researchers are looking for conformation of their data; does my data match the results from other researchers? However, like a farmer, who would be expected to select only the ripest and healthiest fruits. Causing an observer who sees only the selected fruit may thus wrongly conclude that most, or even all, of the tree's fruit is in a likewise good condition. The same could, and probably should not, happen in research. By 'cherry picking' the results that confirm your obtained data, could lead to conformation biases and give a misrepresentation of reality. Artificial intelligence (AI)-driven tools could provide a powerful tool in this type of data rectification. These artificial intelligence programmes can examine, observe and predict possible relations between experimental derived data. In epidemiology a variety of AI tools are implemented in different types of research. However for this study we used an AI tool which is already implemented for pharmaceutical research. The Euretos Knowledge Platform (EKP) is used in cancer research and other disease and drug research. In the pharmaceutical research field, the programme has been recognized on one of the leading companies and used by world leading pharma, biotech and academic institutions such as AstraZenica and Johnsen & Johnsen (Janssen). The main objective of this study was to see if the EKP can improve the environmental epidemiological research and provide us with new insights in this field of biology.

We used the EKP to look into relations between the exposure of air pollution, the effected genes and relations to different health outcomes. We defined the genes associated to different types of air pollution (e.g. ozone, nitrogen oxides particulate matter (fine or ultrafine particles)) combined with the 5 different health outcomes (Asthma, COPD, Diabetes Mellitus Type 2, acute myocardial infarction and CVA). For this study we used the EKP to organize relation maps, to give a better representation of the relations between the different genes. Figures, like Fig. 3, help us to give insights on the different relations.

In the end, the programme showed different relations between the air polluted related genes and the different health outcomes. Next to this, the EKP could perform useful pathway analysis. An type of analysis that helps researchers gain insight into gene lists. This method identifies biological pathways that are enriched in a gene list more than would be expected by chance. However, the programme was programmed to work for drugs and disease research. A scientific field in which more information is available. The lack of curated (organized) epidemiological data, limited the programme to work optimal. More available data and some minimal changes could make the EKP feasible for widespread use in different fields of research.

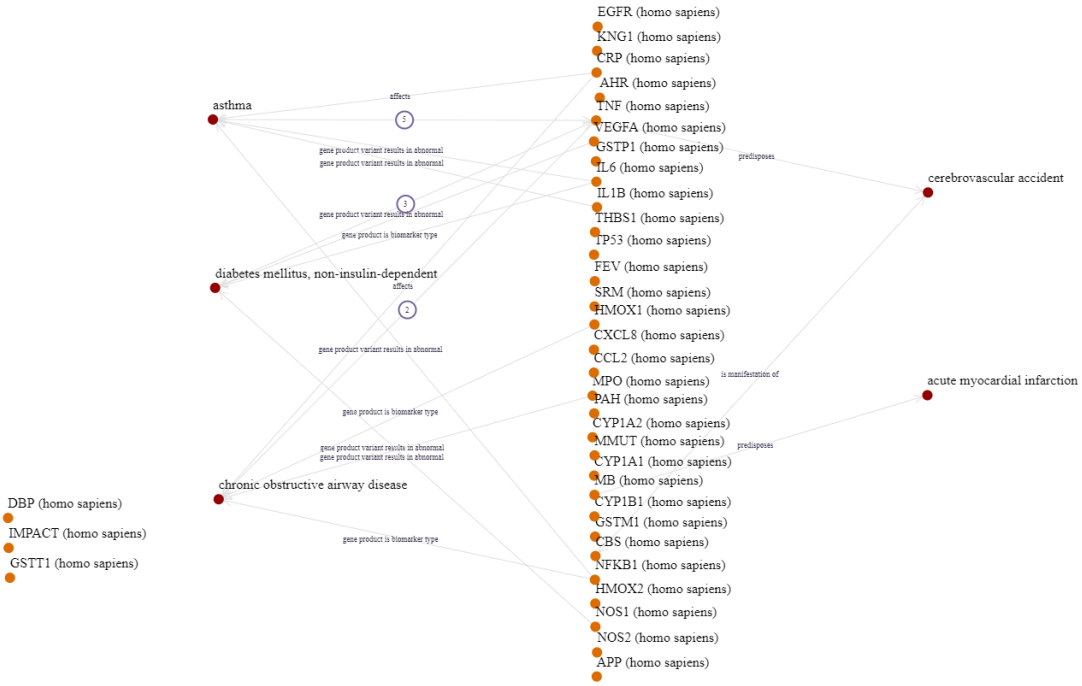

Figure 3. Relation map represents the relations between the obtained top 10 related genes to all the different pollutants, enriched with the five different health outcomes. The links among the genes, and those between the diseases are removed. a) All relation types are shown. b) All relation types are shown, except the relation 'is associated with', to give a more accurate representation of the relations.