**Understanding Genetic Data Information**

**Checks to see if they can be linked to other files:**

**E‑GEOD‑7543** – Human Gene Expression Across Pollution Exposure Groups

Type of data: Individual-level gene expression microarray data (whole blood samples) from individuals in high vs low pollution regions

Location info: Based on regions in the Czech Republic (Ostrava vs Prague). This data was grouped broadly and not with precise local Authority codes or continuous time series environmental exposure.

Can This data be merged with: This data is not suitable for merging; it does not have individual level area code and yearly alignment with environmental PM2.5 dataset

**E‑MTAB‑3630** and **E‑MTAB‑3629 -** UK Studies Related to PM2.5 and cardiovascular disease.

Type of Data: 3629 transcription profiling and 3630 microRNA profiling, transcriptomic datasets from human cohorts exposed to UK PM2.5 related to CVD outcomes

Location Info: UK based Data.

Can This data be merged with: Unlikely

**Summary of the 3 genetic datasets.**

1. The 3 datasets you explored (E-GEOD-7543, E-MTAB-3630, E-MTAB-3629) are lab-based.
2. They do not have Geographical regions like Local Authority and Area Code
3. They do not cover Time or Year

Based on the 3 datasets, they would be useful for studying how cells respond to PM2.5 or PM10. But the 3 datasets cannot be used for merging with the environmental datasets.

**Next Plan of Action.**

To merge environmental and genetic data meaningfully, i need real-world, population-level genetic data that includes:

* Location information to match with pollution exposure data like LA, AC
* Date or birth cohort info to align with air pollution time series

[Combined Effect of Air Pollution and Genetic Risk on Incident Cardiovascular Diseases - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC11681391/)