PARTICIPANT INFORMATION SHEET

This document shows important information for the use of 'MethylDetectR'. Robert Hillary and Riccardo Marioni at the University of Edinburgh are leading this research. Before you decide to take part, it is important you understand how data will be used in relation to 'MethylDetectR', the risks of uploading your data and the data security measures which have been put in place to mitigate these risks. Please take time to read the following information carefully.

WHAT IS THE PURPOSE OF THE STUDY?

'MethylDetectR' is a series of two linked web-based applications. 'MethylDetectR' allow users to quickly and conveniently obtain estimated values for human traits based off of DNA methylation data. 'MethylDetectR' uses information on whether chemical tags have or have not been added to people's DNA - which can tell us about health profiles and lifestyle behaviours. At present, the traits which we estimate from methylation data include age [1], alcohol consumption, body fat percentage, body mass index, HDL cholesterol levels, smoking status and waist: hip ratio. Additional traits will be included when they are developed by our group and others through state-of-the-art machine learning methods. Our application is the first interactive, online platform which automatically generates estimators for human traits from DNA methylation data and permits interactive methylation-based health profiling.

Importantly, the users can opt to use the first of the applications which takes methylation data and automatically processes it to generate the estimated values or methylation-derived scores for human traits. This application is named 'MethylDetectR – Calculate Your Scores'. Methylation data (as .rds objects) may be reported as beta values or M values, can have individuals input as either rows or columns and can have missing CpG sites or missing values. The application can convert the data to beta values, transpose the data into the correct format and deal with missing values and CpG sites – a log of processing steps that have been applied to the data is generated alongside the download of methylation-based estimators of traits. The user can alternatively carry out this for themselves if they are comfortable with coding, instructions will be included and the method is carried out as per the method described in McCartney and Hillary *et al.* (2018) [2].

Whether the user has generated the methylation-based scores for traits through the first of the applications or locally, they can upload a .csv file to the main 'MethylDetectR' app. In this app, users can interactively view how methylation-based scores for a given trait for the individual(s) in the input dataset compare to the remainder of the input dataset. Percentile ranks for all individuals may be downloaded. Users can also restrict comparison to different age ranges and sex. Additionally, users may opt to upload binary case vs. control data for traits and interest and view how methylation-based scores vary according to cases and controls.

DO I HAVE TO TAKE PART?

No – it is entirely up to you. If you do decide to take part, please consider the information on this Participant Information Sheet and read the Informed Consent Form. Data are not stored following your use of the app, it is simply an interactive research tool to help translate important findings in relation to methylation-based health profiling. It is up to you whether you wish to upload your methylation data and/or methylation-based scores for human traits. Appropriate safeguards have been put in place to mitigate risks associated with data security and protection concerns – these include the applications being hosted on a secure and patched server, extra steps being taken to prevent attacks, minimisation of required data and also the use of pseudonymised IDs. We recommend that when you upload your data, you use anonymised or at least pseudonymised data where possible.

WHAT WILL HAPPEN IF I DECIDE TO TAKE PART?

Data will be uploaded to the applications but not stored. They will be purged after you close the app and are not stored in a global environment. The app is simply for the benefit of the user and for efforts towards translating findings in epigenetic epidemiology for other researchers and the wider public.

WHAT ARE THE POSSIBLE BENEFITS OF TAKING PART?

Should you use the first application, you will obtain an automated calculation of methylation-based scores for a variety of traits for the individual(s) in your input dataset. If you use either application or the main 'MethylDetectR' function, you will be able to interactively view how scores vary across your sample and according to case vs. control status following an optional upload of binary phenotype data. Use of 'MethylDetectR' also helps to support the translation of methylation-based health profiling and epigenetic epidemiology research at the University of Edinburgh.

ARE THERE ANY RISKS ASSOCIATED WITH TAKING PART?

This project involves the upload of biological data to a web-based application. A number of measures are in place to ensure best practice in relation to data security and data protection. These are outlined in the next section. However, there is always some risk of 'motivated intruders' trying to utilise such information and also the risk of attacks on personal data. As a limitation, methylation-based estimators for human traits are not wholly accurate, they work well at a population level but may not faithfully recapitulate information about one's metrics on an individual level. This also limits possible identification of individuals; however, general information regarding the overall health status of an individual may be inferred from methylation data. As stated, the risks of this are mitigated as much as possible, but you may decide for yourself based on the information provided that the benefit of using our app does not sufficiently outweigh the risks of uploading biological data. Further information or clarification can be obtained through the relevant contact details at the bottom of this sheet, should you have any additional queries or concerns. These are most welcomed.

DATA PROTECTION AND CONFIDENTIALITY

All information uploaded should be anonymised or pseudonymised, if possible. Generation Scotland data is pseudonymised but this is not viewable to the user. The upload of age data is optional and if the user chooses not to do so, the age of individuals is predicted from methylation data which helps to reduce the possible identification of individuals. The upload of sex and/or case vs. control information are also optional. The applications are hosted on a University of Edinburgh server that is patched and secure. Data are not stored after use once the app is closed and are not saved to a global environment in R which is the programming language used to design the application. Results are available for download for the user, should they so wish.

WHO CAN I CONTACT?

If you have any further questions, complaints about the study or require assistance, please contact the lead researchers, Robert Hillary (robert.hillary@ed.ac.uk) or Riccardo Marioni (riccardo.marioni@ed.ac.uk)

For general information about how we use your data, please visit the following webpage:

https://www.ed.ac.uk/records-management/privacy-notice-research

References

1. Zhang Q, Vallerga CL, Walker RM, Lin T, et al: **Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing.** *Genome Medicine* 2019, **11:**54.
2. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, et al: **Epigenetic prediction of complex traits and death.** *Genome Biol* 2018, **19:**136.