CORRESPONDENCE

In conclusion, western blot and protein array analyses can indeed be useful tools when selecting specific antibodies for other applications. The use of these methods is encouraged both for antibody providers and users, and antibodies with signs of cross-reactivity in these applications should be treated with caution. However, the formal validation of an antibody for a specific application must be performed in an application- and context-specific manner as suggested by the working group[1].

Mathias Uhlen

Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden and NNF Center for Biosustainability, Danish Technology University, Copenhagen, Denmark.
e-mail: mathias.uhlen@scilifelab.se

1. Uhlen, M. *et al. Nat. Methods* **13**, 823–827 (2016).
2. Uhlen, M. *et al. Science* **347**, 1260419 (2015).
3. Älgenäs, C. *Biotechnol. J.* **9**, 435–445 (2014).
4. Hjelm, B. *et al. Protein Sci.* **20**, 1824–1835 (2011).
5. Lindskog, C. *et al. Discov. Med.* **9**, 565–578 (2010).
6. Sjöberg, R. *et al. N. Biotechnol.* **33**, 582–592 (2016).

Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses

To the Editor: Recently, there has been growing interest in statistical algorithms designed for tackling intra-sample cellular heterogeneity (ISCH) in epigenome-wide association studies (EWAS)[1]. Such algorithms can be broadly classified as reference-based (if they use reference DNA methylation (DNAm) profiles of representative cell types)[2] or reference-free (if they do not require such reference profiles)[3–6]. Reference-free methods can be further subdivided into those that use the phenotype of interest in the inference process (this includes algorithms such as Surrogate Variable Analysis (SVA)[4,7] and RefFreeEWAS[3]) and those that do not (for example, EWASher[5] and Remove Unwanted Variation (RUV)[6]). Comparisons of these different inference paradigms are of paramount interest to inform the EWAS community on how best to approach the ISCH problem.

A recent study by Rahmani *et al.*[8] presented a reference-free algorithm called ReFACTor and suggested that it leads to improved estimates of cell-type composition and power compared to other competing algorithms. However, the approach on which ReFACTor is based could incorrectly remove the biological signal of interest if the latter is stronger than the variation associated with cell-type composition. We confirmed this by applying ReFACTor to additional data sets. Below we discuss key issues to which any future methodological comparative study should pay particular attention to ensure robust and meaningful conclusions that can then be used to guide the EWAS community.

In principle, an advantage of a reference-free method such as ReFACTor is that it is applicable to any tissue type. It is important, therefore, to assess performance in tissue types other than blood, because assumptions valid in one tissue type may not be valid in others. For instance, ReFACTor relies on the assumption that the top components of variation are associated with changes in cell-type composition, effectively using these components to construct variables that account for variations in cell type. Although this assumption may be valid for EWAS conducted in whole blood[9], the generality of it to other tissue types remains to be shown. In essence, ReFACTor has some similarity to RUV[6] in that both select factors that capture confounding variation. However, blind application of ReFACTor could lead to a substantial loss of power if the inferred sparse factors are misidentified as those carrying biological signal. Although these problems represent an intrinsic limitation of any reference-free method, it will be particularly acute for methods such as ReFACTor or EWASher[5], which do not use phenotype information from the outset. We used normal mammary epithelial and breast cancer cell line data to define a gold-standard set of true positive features and a breast cancer tissue EWAS for the evaluation of several methods. SVA[4] had a much better control of power, outperforming ReFACTor by as much as 70% (**Table 1**, **Supplementary Data 1** and **2** and **Supplementary Software 1** and **2**). Although specificity is harder to estimate, the improved power of SVA over ReFACTor was at the expense of only a 10–20% lower specificity (**Table 1**). ReFACTor's loss of power in our cancer tissue EWAS was due to the top components of variation having a stronger correlation with disease status than with cell-type composition (**Supplementary Fig. 1**). Only lower-ranked components correlated with adipose cell content, which is the major source of cell-type variation in breast tissue (**Supplementary Fig. 1**). This problem could in principle be circumvented by applying ReFACTor to the normal samples only, as suggested by Rahmani *et al.*[8], but it remains to be tested on more data sets. Hence, application of a method like ReFACTor demands that one must carefully consider the tissue and biological context.

A second key issue concerns the evaluation of a reference-free method in terms of modeling cell-type composition. In the case of ReFACTor, estimated components were added successively to a linear model, leading to an improvement in the fraction of variance explained (summarized with R^2 values). To avoid the problem of overfitting, we used a nested-models likelihood ratio test (LRT) (or adjusted R^2 values). We found little justification for the successive addition of components (**Supplementary Methods**, **Supplementary Software 1–4**, **Supplementary Data 3** and **Supplementary Fig. 2**). Alternatively, one could attempt to estimate the number of significant components of variation. In our hands, entering such estimates into ReFACTor led to a drop of as much as 20% in R^2 values, resulting in reduced modeling performance compared to that of reference-based methods (**Supplementary Figs. 3** and **4**). This indicates that application of ReFACTor with all estimated components could lead to overfitting. We confirmed this further using training–test set partitions (**Supplementary Fig. 5**).

Another issue is the use of a single or a limited number of data sets with matched FACS data to benchmark a novel method against existing algorithms. In our experience, the complexity and unknown nature of the sources of variation in EWAS data requires many data sets to reach unbiased conclusions. To demonstrate this, we performed a cell-composition analysis for an independent whole-blood data set as well as an extensive analysis

Table 1 | Relative sensitivity (SE) and specificity (SP) of ReFACTor and SVA

	Unadjusted	SVA	ReFACTor ($k = 6$, ncp = 6)	ReFACTor ($k = 6$, ncp = 15)	ReFACTor ($k = 10$, ncp = 10)	ReFACTor ($k = 10$, ncp = 15)
SE ($P < 0.05$)	0.90	0.83	0.09	0.02	0.02	0.02
	($n = 20{,}876$)	($n = 19{,}356$)	($n = 2{,}066$)	($n = 412$)	($n = 410$)	($n = 410$)
SE (FDR < 0.05)	0.89	0.81	0.04	~0	~0	~0
	($n = 20{,}667$)	($n = 18{,}743$)	($n = 835$)	($n = 23$)	($n = 13$)	($n = 13$)
SP ($P < 0.05$)	0.53	0.70	0.62	0.92	0.95	0.95
	($n = 16{,}057$)	($n = 10{,}274$)	($n = 12{,}793$)	($n = 2{,}603$)	($n = 1{,}582$)	($n = 1{,}582$)
SP (FDR < 0.05)	0.58	0.75	0.84	0.99	~1	~1
	($n = 14{,}146$)	($n = 8{,}436$)	($n = 5{,}571$)	($n = 115$)	($n = 11$)	($n = 11$)

ReFACTor for two different choices for k ($k = 6$ or 10) and for number of components (ncp = k or ncp = 15, estimated using RMT[7]) (**Supplementary Methods**) was compared to SVA and an unadjusted analysis. SE and SP were estimated using a set of $n = 23{,}258$ true positives and 34,078 true negatives, respectively, and are shown at an unadjusted $P < 0.05$ and FDR corrected < 0.05.

encompassing five different *in silico* mixture experiments, drawing on 1,573 purified blood cell types from more than 6 different studies (**Supplementary Table 1**). These analyses demonstrate the strength of Houseman's reference-based method compared to ReFACTor (**Supplementary Figs. 2** and **6–9**). Further issues, including inappropriate choice of gold standards in real data, are discussed in **Supplementary Methods**.

In summary, we suggest that future studies proposing novel methods ought to (i) provide comprehensive comparisons to existing algorithms, (ii) use biological scenarios and data sets that allow objective comparisons and (iii) include tissues other than blood, when applicable. We provide some recommendations in **Supplementary Methods** and **Supplementary Table 2**. Briefly, we recommend reference-based methods for scenarios where the composition of tissues is relatively well known and reference-free methods such as SVA or RefFreeEWAS when reference DNAm profiles are not available. We point out that our recommendations are based on currently available data sets and approaches, which may change as the field continues to evolve.

Data availability statement. All data analyzed are publicly available. Accession numbers for all data sets analyzed are listed in **Supplementary Methods**.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

Shijie C Zheng[1,2]**, Stephan Beck**[3]**, Andrew E Jaffe**[4,5]**, Devin C Koestler**[6]**, Kasper D Hansen**[7–9]**, Andres E Houseman**[10]**, Rafael A Irizarry**[11,12] **& Andrew E Teschendorff**[1,13,14]

[1]CAS Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]Medical Genomics, UCL Cancer Institute, University College London, London, UK. [4]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [5]Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland, USA. [6]Department of Biostatistics, University of Kansas Medical Center, Kansas City, Kansas, USA. [7]Center for Epigenetics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [8]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [9]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [10]School of Biological and Population Health Sciences, College of Public Health and Human Sciences, Oregon State University, Corvallis, Oregon, USA. [11]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [12]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [13]Statistical Cancer Genomics, Paul O'Gorman Building, UCL Cancer Institute, University College London, London, UK. [14]Department of Women's Cancer, University College London, London, UK.
e-mail: a.teschendorff@ucl.ac.uk

1. Jaffe, A.E. & Irizarry, R.A. *Genome Biol.* **15**, R31 (2014).
2. Houseman, E.A. *et al. BMC Bioinformatics* **13**, 86 (2012).
3. Houseman, E.A., Molitor, J. & Marsit, C.J. *Bioinformatics* **30**, 1431–1439 (2014).
4. Leek, J.T. & Storey, J.D. *PLoS Genet.* **3**, e161 (2007).
5. Zou, J., Lippert, C., Heckerman, D., Aryee, M. & Listgarten, J. *Nat. Methods* **11**, 309–311 (2014).
6. Gagnon-Bartsch, J.A. & Speed, T.P. *Biostatistics* **13**, 539–552 (2012).
7. Teschendorff, A.E., Zhuang, J. & Widschwendter, M. *Bioinformatics* **27**, 1496–1505 (2011).
8. Rahmani, E. *et al. Nat. Methods* **13**, 443–445 (2016).
9. Liu, Y. *et al. Nat. Biotechnol.* **31**, 142–147 (2013).