

Epigenome-wide association studies for breast cancer risk and risk factors

Annelie Johansson and James M. Flanagan*

Epigenetics Unit, Division of Cancer, Department of Surgery and Cancer, Imperial College London, UK.

ABSTRACT

There have been six epigenome-wide association studies (EWAS) for breast cancer risk using blood DNA from prospective cohorts published thus far, and the only consistent finding is a global loss of methylation observed in breast cancer cases compared with controls, with no individual CpG sites passing validation across studies. In contrast, a more successful approach has been the identification of EWAS signatures of cancer risk factors such as smoking, body mass index, age and alcohol use with numerous validated CpG sites. These signatures may be used as a molecular test to quantify cancer risk associated with these factors. It is clear from the larger EWAS of risk exposures that similar-sized large collaborative studies may be needed to robustly identify DNA methylation signatures of breast cancer risk.

KEYWORDS: DNA methylation, breast cancer risk, epigenome-wide association study, EWAS, epigenetics

INTRODUCTION

Current breast cancer screening methods are effective; however, they could be improved by targeting those at highest risk and the idea of developing a risk-stratified screening strategy is gaining support [1]. Multiple lifestyle and environmental factors influence breast cancer aetiology, including age, hormonal and reproductive factors, body mass index (BMI), physical activity,

alcohol intake, smoking, benign breast disease, high mammographic-density and family history. The addition of the combined “polygenic risk scores” integrating genetic risk markers is likely to make a modest improvement to breast cancer risk models. However, there is considerable room for improvement by identifying further independent risk factors [2]. In this review we consider the evidence for epigenetic risk markers for breast cancer.

Recent research has explored the possibility that epigenetics may play a role in determining cancer risk. The field of epigenetic epidemiology has rapidly advanced and numerous studies have discovered epigenetic markers of breast cancer risk in white blood cell (WBC) DNA [3-15]. A recent review has summarised 45 studies that have reported epigenetic traits associated with breast cancer risk [16]. While replication of candidate genes and top hits from epigenome-wide studies have failed to identify any strong validated biomarkers of breast cancer risk, the conclusion thus far is that there appears to be an epigenome-wide loss of DNA methylation in breast cancer cases many years prior to diagnosis [10, 13]. The underlying mechanisms of how epigenetic patterns are altered and how this is related to cancer risk are unclear. The main hypothesis proposes that cancer risk exposures, lifetime and environmental events, can alter the epigenome and stably modify an individual's cancer risk. The best example of this thus far is the epigenetic signatures of smoking [11, 12, 17] that partly mediate lung cancer risk [18]. For breast cancer, a similar model is yet to be described.

*Corresponding author: j.flanagan@imperial.ac.uk

What is epigenetics?

Epigenetic modifications, such as DNA methylation or histone modifications are key determinants of chromatin structure, genomic stability and gene expression. DNA methylation is the addition of methyl groups to the cytosine in a cytosine-guanine dinucleotide (CpG) site and generally leads to gene silencing when gene promoters become methylated, by blocking the binding of transcription factors. In contrast, DNA methylation in the middle of genes (gene body methylation) appears to have a positive correlation with gene expression, potentially by preventing aberrant transcription [19, 20]. This is an important distinction for interpreting the mechanisms of associations in EWAS results, as many reported CpG associations occur in gene body CpG sites. Histone proteins form the nucleosome around which the DNA is wrapped and packaged into the nucleus and a whole range of modifications of these proteins including methylation, phosphorylation and acetylation among others dictate their function. These epigenetic modifications are maintained during cell division and when perturbed, play a key role in cancer development [21]. A proportion of the epigenome is tissue specific and is the reason why different cell types have different phenotypes (e.g. brain cells, liver cells, breast epithelial cells and white blood cells) despite having identical DNA sequences.

If epigenetic patterns are tissue specific why investigate WBCs in relation to cancer risk in other parts of the body? The argument often used is that cancer causing exposures might affect only breast epithelial cells or stem and progenitor cells and would only be detected in this cell type, in which case looking in a surrogate tissue such as blood DNA is irrelevant to the disease. Firstly, it is possible that exposures would affect blood cells but not epithelial cells, being the immune/inflammatory system which may then indirectly act on the breast epithelial cells to modify risk [22]. However, most evidence suggests that cancer risk exposures can independently influence methylation of both cell types which supports the counter argument that WBC DNA methylation provides an accurate surrogate for breast epithelial cells [23]. Indeed, epigenetic biomarkers have been identified in WBC DNA that are associated with numerous

cancer risk factors, including genetic polymorphisms [24], aging [25-27], hormones [28], alcohol [7, 29], weight or BMI [7, 30, 31], age-at-menarche [32] and smoking [11, 12, 33]. Furthermore, some exposure methylation biomarkers are stable over time [7], and therefore, represent excellent candidates for biomarkers of cancer risk.

While it is now possible to consider histone modifications in an epigenome-wide study for risk [34] the majority of studies have focused on DNA methylation. There are two main approaches that have been used thus far for epigenome-wide DNA methylation studies related to cancer risk; identifying DNA methylation biomarkers of cancer risk directly in an agnostic case-control study design, and alternatively identifying DNA methylation biomarkers of known cancer risk factors. It is important to only focus on studies that have used prospective cohorts and incident breast cancer cases with blood samples taken prior to diagnosis when investigating biomarkers of cancer risk. Studies that have used breast cancer cases recruited at or after diagnosis are confounded by “reverse causation”, the possibility that the cancer itself is causing the epigenetic changes detected in the blood DNA and therefore should not be used to draw conclusions about cancer risk.

Prospective cohort studies for breast cancer risk-EWAS

To date, there have been six published EWAS using prospective cohorts to investigate breast cancer risk with a case-control study design, two using the low-resolution Illumina 27k DNA methylation array [3, 15] and four using the 450k array [10, 11, 13, 35]. Using the Sister Study cohort Xu and colleagues identified 250 CpG sites (FDR <0.05) that were significantly different between cases and controls [15]. The majority (>75%) of these probes were hypomethylated in cases compared with controls. Anjum and colleagues derived a methylation signature associated with BRCA1 carriers that is also modestly associated with incident non-familial breast cancer cases in the UKCTOCS cohort (area under the curve (AUC) = 0.57 (95% confidence interval: 0.50-0.64); $p = 0.03$) [3]. In the first 450k EWAS for breast cancer risk Shenker and colleagues also derived a hypomethylation signature associated with

breast cancer risk in the EPIC-Italy cohort; however, no individual probes were associated with breast cancer risk at the genome-wide level of significance ($p < 1e-7$) [11]. Given that these first studies failed to validate any individual CpG sites across studies, two more recent studies used a different approach by averaging methylation across the whole 450k array to investigate whole-genome methylation levels. Severi and colleagues reported an average loss of methylation in breast cancer cases compared with controls (odds ratio (OR) per 1 standard deviation (SD) change in methylation = 0.69 (0.50-0.95); $p = 0.02$) [10]. Combining data from three cohorts van Veldhoven and colleagues reported a meta-analysis which showed significant heterogeneity across populations with the EPIC-Italy cohort, validating this hypomethylation association with breast cancer risk (OR per 1 SD = 0.61 (0.46-0.80), $p = 0.0003$), while the NOWAC cohort did not (OR per 1 SD = 1.03 (0.82-1.30), $p = 0.81$) [13]. One potential explanation for this heterogeneity that was proposed was the significant difference in follow up times, with the two studies with the longer follow up time (MCCS and EPIC-Italy) showing the association, while the shorter follow up time of 5 years in the NOWAC cohort did not. Another explanation may be significant differences in cohort characteristics and the distribution of known breast cancer risk factors. Lastly, the most recent EWAS for breast cancer risk by Ambatipudi and colleagues used the EPIC cohort (subjects recruited from Italy, Spain, UK, The Netherlands, Greece and Germany) and reported accelerated methylomic aging associated with breast cancer risk in post-menopausal women (OR = 1.07 (1.02-1.11); $p = 0.003$), but not in pre-menopausal women (OR = 1.00 (0.95-1.06); $p = 0.94$) [35]. This study also reported a genome-wide hypermethylation in CpG islands associated with breast cancer risk (OR per 1 SD = 1.20 (1.03-1.40); $p = 0.02$) in contrast to previous reports.

In summary, the results thus far for EWAS of breast cancer risk have been inconsistent and further work is needed to clarify these results. There are numerous limitations of the studies published thus far, including relatively small study sizes with between 100-500 case-control pairs in each, potential population biases in all cohorts, the need to adjust for important measured

or unmeasured environmental and lifestyle factors that influence methylation variation and lastly the limitation of the 450k array used in only targeting 1.8% of the CpG sites in the genome. Using whole genome bisulphite sequencing (WGBS) would address this limitation, but is currently cost prohibitive on individual subjects [13]. Clearly, as with exposure-related EWAS discussed below, large collaborations and consortia approaches will be needed to address these limitations.

EWAS for cancer risk factors

An alternative approach that has been more successful has been the identification of EWAS signatures of cancer risk factors such as smoking (reviewed in [36]), BMI [7, 30, 31, 37-39], age [25-27] and alcohol use [7, 29, 40]. This approach proposes that epigenetic changes in response to these exposures may provide a biological mechanism for increased risk and may also provide a biomarker to quantify the level of risk associated with the exposure.

Age is the biggest risk factor for cancer. Epigenetic signatures of age have been identified and it has been shown that the epigenetic clock is accelerated in cancer tissues [25, 26]. Two recent studies have reported accelerated methylomic aging associated with breast cancer risk [35, 41] and all-cause and cancer-specific mortality [42]. Advanced epigenetic age is also associated with known risk factors such as menopausal status [43], BMI [44, 45], air pollution [46] and smoking [47]. Unlike many other exposures the changes in DNA methylation associated with aging results in increased variance with age, and both hyper- and hypo-methylation [48].

Other risk exposures such as BMI, smoking and alcohol consumption are associated largely with hypomethylation [7, 11, 47, 49]. A number of consistently identified genes associated with smoking have been identified; hypomethylation at loci located in the gene bodies of *AHRR* and *F2RL2* and the intergenic loci 6p21.33 and 2q37.1 [11, 33]. Furthermore, methylation at *AHRR* has been linked to lung cancer risk and methylation at *F2RL3* to cardiovascular diseases (CVD), lung cancer and all-cause mortality [18, 47, 50, 51], and the 2q37.1 locus has been further associated with breast cancer risk [11]. Combining CpG sites

into a methylation signature, instead of focusing on individual CpG sites, could provide a robust biomarker for the risk exposure. For example Shenker and colleagues combined four loci (*AHRR*, 6p21 and two loci at 2q37) into a methylation index, which could distinguish non-smokers with high accuracy from former smokers (AUC = 0.82) [12]. A recent study further supported this strong predictive ability for current smokers using buccal cells [52]. However, individual loci could also potentially provide enough information alone; methylation at a single CpG site in *F2RL3* in combination with pack-years (lifetime cumulative smoking intensity) could predict lung cancer incidence and mortality with high accuracy [51], and one CpG site in *AHRR*, cg05575921 for which the authors have developed a commercial kit, could discriminate smokers from non-smokers with a very high accuracy (AUC = 0.99) [53]. Interestingly, in the case of smoking it has been shown that the DNA methylation changes are reversible but may take up to 40 years to return to the levels seen in individuals who have never smoked [17]. We have proposed that this indicates that epigenetic signatures represent a biological memory of past exposures and that the signature is retained in the exposed stem and progenitor cells [54]. This matches well with the long term stability of the methylation levels at these smoking markers in serial blood samples taken six years apart [7]. Indeed, smoking associated DNA methylation changes have been shown to mediate lung cancer risk [18, 51, 55] and potentially breast cancer risk [11]. This long lasting epigenetic response to the exposure provides a biological mechanism for the long term increased risk of cancer associated with former smoking.

Increased BMI has been associated with specific CpG sites; however, findings have been inconsistent across studies. Most studies have identified methylation at *HIF3A* associated with BMI, in combination with other CpG sites that often are associated with inflammation, immunology and cellular responses [56]. Furthermore, CpG sites identified in EWAS of BMI have been linked to type-2-diabetes [57], cardiovascular disease (CVD) [30], and cancers [31]. The largest EWAS for BMI to date has identified 187 CpG sites associated with BMI and showed that many significant CpG sites

in blood show the same pattern in adipose tissue, providing more evidence for the use of blood DNA as a surrogate of other tissues for exposure-induced risk assessment [30]. Relevant to breast cancer, one significant CpG site associated with BMI in the *PHGDH* gene is involved in cell proliferation and over expressed in ~70% of oestrogen receptor (ER)-negative breast cancers [37]. Indeed, many BMI associated DNA methylation sites can be observed in breast tissue [58], providing a potential hypothesis for a mechanistic role for BMI in carcinogenesis linked by DNA methylation alterations.

The mechanisms linking alcohol consumption to methylation changes and liver disease, colorectal cancer and breast cancer risk has been previously described [49]. Significant CpG sites associated with alcohol consumption in breast cancer include hypermethylation of the *ER-alpha* and *E-cadherin* genes and decreased promoter methylation in general [49, 59]. Recently a large collaboration identified a methylation signature of 144 CpG sites that can discriminate heavy drinkers from non-drinkers with high accuracy across several cohorts (AUC > 0.90) [29]; however this has not been further linked with any specific cancer risk. There are many plausible molecular mechanisms underlying the association between alcohol consumption and breast cancer risk, which may be due to lower folate levels in heavy drinkers that can affect DNA methylation by the decrease in DNA methyltransferase (DNMT)-activity from ethanol intake, or through an increase in oestrogen levels in blood [49]. Further work linking DNA methylation signatures of alcohol consumption to cancer risk is therefore needed.

These studies provide an excellent paradigm for other cancer risk factors that are known to have long lasting effects such as age-at-menarche, age-at-first pregnancy, oral contraceptive (OC) use, hormone replacement therapy (HRT) use, folate levels or even breast density. Conducting EWAS and developing exposure-specific methylation signatures as a molecular quantification of these risk factors, in contrast to unreliable self-reported questionnaire data, is likely to be a fruitful avenue for future research. There is already evidence that each of these risk factors may lead to epigenetic changes. For example an older age-at-menarche has been linked to global hypomethylation in

blood using the Luminometric Methylation Assay (LUMA), but no association was yet found using the 450k array [32]. In an asthma study, DNA methylation measured in blood was linked with specific genetic polymorphisms in *GATA3* that was also associated with OC use in a young population [24]. Global hypomethylation in WBCs was also seen in longer-term users of oral contraceptives [60]. HRT-users have lower methylation in blood compared to non-users as observed in a genome-wide study conducted in monozygotic twin pairs [61]. A younger age at first full term pregnancy and number of pregnancies have been associated with lower DNA methylation at the *ER-alpha* gene in breast tissue [62], and global methylation in blood using the [³H]-methyl acceptance assay show lower methylation for nulliparous women and a late age at first pregnancy [63]. All these risk factors for breast cancer have one thing in common - they are known to affect oestrogen exposure, either in varying the length of time a woman is exposed to high levels of oestrogens (e.g. age at menarche and age at menopause), or regulating oestrogen levels directly (e.g. BMI and parity). It is well established that oestrogen has a role in the development of ER+ breast cancer; however, the mechanism behind this is not fully understood and one hypothesis could be that oestrogen itself stably alters DNA methylation and thereby results in long-term changes in gene expression. It would be interesting therefore, to conduct EWAS for long-term oestrogen exposure.

Finally, there are many exposures and risk factors that are difficult to estimate from questionnaire data and that can change dramatically over a life course, such as diet/food intake, exposure to specific air pollutants, drug/therapy use and physical exercise. This provides a significant limitation to studying their association with cancer risk, and if epigenetic biomarkers of these exposures can be identified then this limitation could be addressed. A good example for this that is worth exploring in future studies could be folate levels which is an important co-factor for DNA methylation involved in one-carbon metabolism and is strongly associated with methylation variation in blood DNA [13, 64]. Decreased serum folate levels have also been associated with increased breast cancer risk in case-control studies, but not in prospective

cohort studies, and perhaps reconsidering this association stratified by baseline DNA methylation levels may provide further clarity [65, 66].

Mechanisms and hypotheses

How different exposures actually cause epigenetic changes that are detectable in blood DNA is largely unexplored. There are two main hypotheses: transcriptional changes and DNA damage and repair. The transcriptional hypothesis argues that the exposures cause cell-type-specific gene expression changes that are “locked in” by the epigenetic reprogramming. Epigenetic changes are mitotically heritable and therefore any changes in stem or progenitor cells will persist in terminally differentiated cell types. Furthermore, if the exposures are systemic, then any changes detected in blood DNA are likely to reflect the changes in other cell types. This supports the argument that blood DNA acts as a surrogate for inaccessible tissue types. The DNA damage and repair hypothesis suggests that the exposures cause DNA damage and the underlying genetic changes or DNA repair of that damage dictate the epigenetic consequences. Examples of this include treatment with the DNA-damaging agent platinum which leads to epigenetic changes in blood DNA [67] which show how DNA damage repair leads to epigenetic changes [68, 69]. Some exposures may act via both mechanisms, such as tobacco smoking which alters transcription of the aryl hydrocarbon receptor repressor by activating the Ahr pathway [11] and can also form DNA adducts that can lead to mutations [70]. Another example is oestrogen, where oestrogen metabolites can cause mutations [71] and also alter transcriptional regulation via the oestrogen receptor [72]. It remains important to consider the biological mechanisms causing the epigenetic changes when interpreting the EWAS results.

CONCLUSION

In conclusion, the results thus far show that the individual CpG sites tested are not robustly associated with breast cancer risk in WBC DNA potentially due to the fact that the effect sizes are too small to be able to detect with statistical significance with the sizes of studies used thus far ($n = 100-500$ case-control pairs). Some of the more recent large collaborative EWAS have

identified and replicated significant numbers of CpG sites associated with the phenotype of interest, for example the $n = 13,317$ subjects used to identify the 144 CpG sites associated with alcohol use [29] or the $n = 15,907$ subjects used to identify the 2,623 CpG sites associated with cigarette smoking [73]. It is clear from these larger exposure studies that similar sized collaborative studies may be needed to robustly identify signatures of breast cancer risk. Finally, there are several unanswered questions regarding epigenetic risk for breast cancer that require further work. It remains important to consider whether epigenetic risk is independent of other risk models such as genetic risk or epidemiological risk models if it is to contribute to clinical risk assessment. Whether epigenetic risk in general, like smoking associated methylation changes, are reversible and whether the effectiveness of prevention (therapeutic or lifestyle) can be monitored with epigenetics is also important to establish.

ACKNOWLEDGMENTS

This work was funded by Breast Cancer Now. The authors acknowledge funding from Imperial Cancer Research UK Centre, the Imperial College Experimental Cancer Medicine Centre (ECMC) and the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) based at Imperial College Healthcare NHS Trust and Imperial College London.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

REFERENCES

1. Pashayan, N. and Pharoah, P. 2012, *Personalized Medicine*, 9, 451-455.
2. Garcia-Closas, M., Gunsoy, N. B. and Chatterjee, N. 2014, *Journal of the National Cancer Institute*, 106, dju305.
3. Anjum, S., Fourkala, E. O., Zikan, M., Wong, A., Gentry-Maharaj, A., Jones, A., Hardy, R., Cibula, D., Kuh, D., Jacobs, I. J., Teschendorff, A. E., Menon, U. and Widschwendter, M. 2014, *Genome Medicine*, 6, 47.
4. Brennan, K. and Flanagan, J. M. 2012, *Cancer Prevention Research*, 5, 1345-1357.
5. Brennan, K. and Flanagan, J. M. 2012, *Methods in Molecular Biology*, 863, 439-465.
6. Brennan, K., Garcia-Closas, M., Orr, N., Fletcher, O., Jones, M., Ashworth, A., Swerdlow, A., Thorne, H., Investigators, K. C., Riboli, E., Vineis, P., Dorronsoro, M., Clavel-Chapelon, F., Panico, S., Onland-Moret, N. C., Trichopoulos, D., Kaaks, R., Khaw, K. T., Brown, R. and Flanagan, J. M. 2012, *Cancer Research*, 72, 2304-2313.
7. Flanagan, J. M., Brook, M. N., Orr, N., Tomczyk, K., Coulson, P., Fletcher, O., Jones, M. E., Schoemaker, M. J., Ashworth, A., Swerdlow, A., Brown, R. and Garcia-Closas, M. 2015, *Cancer Epidemiology, Biomarkers & Prevention*, 24, 221-229.
8. Flanagan, J. M., Munoz-Alegre, M., Henderson, S., Tang, T., Sun, P., Johnson, N., Fletcher, O., Dos Santos Silva, I., Peto, J., Boshoff, C., Narod, S. and Petronis, A. 2009, *Human Molecular Genetics*, 18, 1332-1342.
9. Iwamoto, T., Yamamoto, N., Taguchi, T., Tamaki, Y. and Noguchi, S. 2011, *Breast Cancer Research and Treatment*, 129, 69-77.
10. Severi, G., Southey, M. C., English, D. R., Jung, C. H., Lonie, A., McLean, C., Tsimiklis, H., Hopper, J. L., Giles, G. G. and Baglietto, L. 2014, *Breast Cancer Research and Treatment*, 148, 665-673.
11. Shenker, N. S., Polidoro, S., van Veldhoven, K., Sacerdote, C., Ricceri, F., Birrell, M. A., Belvisi, M. G., Brown, R., Vineis, P. and Flanagan, J. M. 2013, *Human Molecular Genetics*, 22, 843-851.
12. Shenker, N. S., Ueland, P. M., Polidoro, S., van Veldhoven, K., Ricceri, F., Brown, R., Flanagan, J. M. and Vineis, P. 2013, *Epidemiology*, 24, 712-716.
13. van Veldhoven, K., Polidoro, S., Baglietto, L., Severi, G., Sacerdote, C., Panico, S., Mattiello, A., Palli, D., Masala, G., Krogh, V., Agnoli, C., Tumino, R., Frasca, G., Flower, K., Curry, E., Orr, N., Tomczyk, K., Jones, M. E., Ashworth, A., Swerdlow, A., Chadeau-Hyam, M., Lund, E., Garcia-Closas, M., Sandanger, T. M., Flanagan, J. M. and Vineis, P. 2015, *Clinical Epigenetics*, 7, 67.
14. Wong, E. M., Southey, M. C., Fox, S. B., Brown, M. A., Dowty, J. G., Jenkins, M. A., Giles, G. G., Hopper, J. L. and Dobrovic, A. 2011, *Cancer Prevention Research*, 4, 23-33.

15. Xu, Z., Bolick, S. C., DeRoo, L. A., Weinberg, C. R., Sandler, D. P. and Taylor, J. A. 2013, *Journal of the National Cancer Institute*, 105, 694-700.
16. Tang, Q., Cheng, J., Cao, X., Surowy, H. and Burwinkel, B. 2016, *Clinical Epigenetics*, 8, 115.
17. Guida, F., Sandanger, T. M., Castagne, R., Campanella, G., Polidoro, S., Palli, D., Krogh, V., Tumino, R., Sacerdote, C., Panico, S., Severi, G., Kyrtopoulos, S. A., Georgiadis, P., Vermeulen, R. C., Lund, E., Vineis, P. and Chadeau-Hyam, M. 2015, *Human Molecular Genetics*, 24, 2349-2359.
18. Fasanelli, F., Baglietto, L., Ponzi, E., Guida, F., Campanella, G., Johansson, M., Grankvist, K., Johansson, M., Assumma, M. B., Naccarati, A., Chadeau-Hyam, M., Ala, U., Faltus, C., Kaaks, R., Risch, A., De Stavola, B., Hodge, A., Giles, G. G., Southey, M. C., Relton, C. L., Haycock, P. C., Lund, E., Polidoro, S., Sandanger, T. M., Severi, G. and Vineis, P. 2015, *Nature Communications*, 6, 10192.
19. Lou, S., Lee, H. M., Qin, H., Li, J. W., Gao, Z., Liu, X., Chan, L. L., Kl Lam, V., So, W. Y., Wang, Y., Lok, S., Wang, J., Ma, R. C., Tsui, S. K., Chan, J. C., Chan, T. F. and Yip, K. Y. 2014, *Genome Biology*, 15, 408.
20. Teissandier, A. and Bourc'his, D. 2017, *The EMBO Journal*, pii: e201796812.
21. Esteller, M. 2008, *The New England Journal of Medicine*, 358, 1148-1159.
22. Huo, C. W., Chew, G., Hill, P., Huang, D., Ingman, W., Hodson, L., Brown, K. A., Magenau, A., Allam, A. H., McGhee, E., Timpson, P., Henderson, M. A., Thompson, E. W. and Britt, K. 2015, *Breast Cancer Research*, 17, 79.
23. Garcia-Closas, M., Gail, M. H., Kelsey, K. T. and Ziegler, R. G. 2013, *Journal of the National Cancer Institute*, 105, 678-680.
24. Guthikonda, K., Zhang, H., Nolan, V. G., Soto-Ramirez, N., Ziyab, A. H., Ewart, S., Arshad, H. S., Patil, V., Holloway, J. W., Lockett, G. A. and Karmaus, W. 2014, *Clinical Epigenetics*, 6, 17.
25. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sada, S., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T. and Zhang, K. 2013, *Molecular Cell*, 49, 359-367.
26. Horvath, S. 2013, *Genome Biology*, 14, R115.
27. Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T. P., Beyan, H., Whittaker, P., McCann, O. T., Finer, S., Valdes, A. M., Leslie, R. D., Deloukas, P. and Spector, T. D. 2010, *Genome Research*, 20, 434-439.
28. Li, S., Hursting, S. D., Davis, B. J., McLachlan, J. A. and Barrett, J. C. 2003, *Annals of the New York Academy of Sciences*, 983, 161-169.
29. Liu, C., Marioni, R. E., Hedman, A. K., Pfeiffer, L., Tsai, P. C., Reynolds, L. M., Just, A. C., Duan, Q., Boer, C. G., Tanaka, T., Elks, C. E., Aslibekyan, S., Brody, J. A., Kuhnel, B., Herder, C., Almlil, L. M., Zhi, D., Wang, Y., Huan, T., Yao, C., Mendelson, M. M., Joehanes, R., Liang, L., Love, S. A., Guan, W., Shah, S., McRae, A. F., Kretschmer, A., Prokisch, H., Strauch, K., Peters, A., Visscher, P. M., Wray, N. R., Guo, X., Wiggins, K. L., Smith, A. K., Binder, E. B., Ressler, K. J., Irvin, M. R., Absher, D. M., Hernandez, D., Ferrucci, L., Bandinelli, S., Lohman, K., Ding, J., Trevisi, L., Gustafsson, S., Sandling, J. H., Stolk, L., Uitterlinden, A. G., Yet, I., Castillo-Fernandez, J. E., Spector, T. D., Schwartz, J. D., Vokonas, P., Lind, L., Li, Y., Fornage, M., Arnett, D. K., Wareham, N. J., Sotoodehnia, N., Ong, K. K., van Meurs, J. B., Conneely, K. N., Baccarelli, A. A., Deary, I. J., Bell, J. T., North, K. E., Liu, Y., Waldenberger, M., London, S. J., Ingelsson, E. and Levy, D. 2016, *Molecular Psychiatry*, [Epub ahead of print].
30. Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W. R., Kunze, S., Tsai, P. C., Ried, J. S., Zhang, W., Yang, Y., Tan, S., Fiorito, G., Franke, L., Guarrera, S., Kasela, S., Kriebel, J., Richmond, R. C., Adamo, M., Afzal, U., Ala-Korpela, M., Albeti, B., Ammerpohl, O., Apperley, J. F., Beekman, M., Bertazzi, P. A., Black, S. L., Blancher, C., Bonder, M. J., Brosch, M., Carstensen-Kirberg, M., de Craen, A.J., de Lusignan, S., Dehghan, A., Elkalaawy, M., Fischer, K., Franco, O.H., Gaunt, T.R., Hampe, J., Hashemi, M., Isaacs, A., Jenkinson, A., Jha, S., Kato, N.,

- Krogh, V., Laffan, M., Meisinger, C., Meitinger, T., Mok, Z. Y., Motta, V., Ng, H. K., Nikolakopoulou, Z., Nteliopoulos, G., Panico, S., Pervjakova, N., Prokisch, H., Rathmann, W., Roden, M., Rota, F., Rozario, M. A., Sandling, J. K., Schafmayer, C., Schramm, K., Siebert, R., Slagboom, P. E., Soininen, P., Stolk, L., Strauch, K., Tai, E. S., Tarantini, L., Thorand, B., Tigchelaar, E. F., Tumino, R., Uitterlinden, A. G., van Duijn, C., van Meurs, J. B., Vineis, P., Wickremasinghe, A. R., Wijmenga, C., Yang, T. P., Yuan, W., Zhernakova, A., Batterham, R. L., Smith, G. D., Deloukas, P., Heijmans, B. T., Herder, C., Hofman, A., Lindgren, C. M., Milani, L., van der Harst, P., Peters, A., Illig, T., Relton, C. L., Waldenberger, M., Jarvelin, M. R., Bollati, V., Soong, R., Spector, T. D., Scott, J., McCarthy, M. I., Elliott, P., Bell, J. T., Matullo, G., Gieger, C., Kooner, J. S., Grallert, H. and Chambers, J. C. 2017, *Nature*, 541, 81-86.
31. Wilson, L. E., Harlid, S., Xu, Z., Sandler, D. P. and Taylor, J. A. 2017, *International Journal of Obesity*, 41, 194-199.
32. Demetriou, C. A., Chen, J., Polidoro, S., van Veldhoven, K., Cuenin, C., Campanella, G., Brennan, K., Clavel-Chapelon, F., Dossus, L., Kvaskoff, M., Drogan, D., Boeing, H., Kaaks, R., Risch, A., Trichopoulos, D., Lagiou, P., Masala, G., Sieri, S., Tumino, R., Panico, S., Quiros, J. R., Sanchez Perez, M. J., Amiano, P., Huerta Castano, J. M., Ardanaz, E., Onland-Moret, C., Peeters, P., Khaw, K. T., Wareham, N., Key, T. J., Travis, R. C., Romieu, I., Gallo, V., Gunter, M., Herceg, Z., Kyriacou, K., Riboli, E., Flanagan, J. M. and Vineis, P. 2013, *PLoS One*, 8, e79391.
33. Breitling, L. P., Yang, R., Korn, B., Burwinkel, B. and Brenner, H. 2011, *American Journal of Human Genetics*, 88, 450-457.
34. Sun, W., Poschmann, J., Cruz-Herrera Del Rosario, R., Parikshak, N. N., Hajan, H. S., Kumar, V., Ramasamy, R., Belgard, T. G., Elanggovan, B., Wong, C. C., Mill, J., Geschwind, D. H. and Prabhakar, S. 2016, *Cell*, 167, 1385-1397.
35. Ambatipudi, S., Horvath, S., Perrier, F., Cuenin, C., Hernandez-Vargas, H., Le Calvez-Kelm, F., Durand, G., Byrnes, G., Ferrari, P., Bouaoun, L., Sklias, A., Chajes, V., Overvad, K., Severi, G., Baglietto, L., Clavel-Chapelon, F., Kaaks, R., Barrdahl, M., Boeing, H., Trichopoulos, A., Lagiou, P., Naska, A., Masala, G., Agnoli, C., Polidoro, S., Tumino, R., Panico, S., Dolle, M., Peeters, P. H., Onland-Moret, N. C., Sandanger, T. M., Nost, T. H., Weiderpass, E., Quiros, J. R., Agudo, A., Rodriguez-Barranco, M., Huerta Castano, J. M., Barricarte, A., Fernandez, A. M., Travis, R. C., Vineis, P., Muller, D. C., Riboli, E., Gunter, M., Romieu, I. and Herceg, Z. 2017, *European Journal of Cancer*, 75, 299-307.
36. Gao, X., Jia, M., Zhang, Y., Breitling, L. P. and Brenner, H. 2015, *Clinical Epigenetics*, 7, 113.
37. Aslibekyan, S., Demerath, E. W., Mendelson, M., Zhi, D., Guan, W., Liang, L., Sha, J., Pankow, J. S., Liu, C., Irvin, M. R., Fornage, M., Hidalgo, B., Lin, L. A., Thibeault, K. S., Bressler, J., Tsai, M. Y., Grove, M. L., Hopkins, P. N., Boerwinkle, E., Borecki, I. B., Ordovas, J. M., Levy, D., Tiwari, H. K., Absher, D. M. and Arnett, D. K. 2015, *Obesity*, 23, 1493-1501.
38. Demerath, E. W., Guan, W., Grove, M. L., Aslibekyan, S., Mendelson, M., Zhou, Y. H., Hedman, A. K., Sandling, J. K., Li, L. A., Irvin, M. R., Zhi, D., Deloukas, P., Liang, L., Liu, C., Bressler, J., Spector, T. D., North, K., Li, Y., Absher, D. M., Levy, D., Arnett, D. K., Fornage, M., Pankow, J. S. and Boerwinkle, E. 2015, *Human Molecular Genetics*, 24, 4464-4479.
39. Wang, B., Gao, W., Li, J., Yu, C., Cao, W., Lv, J., Pang, Z., Cong, L., Wang, H., Wu, X., Liang, L., Wu, T. and Li, L. 2016, *Lancet*, 388(Suppl.1), S21.
40. Boitano, A. E., Wang, J., Romeo, R., Bouchez, L. C., Parker, A. E., Sutton, S. E., Walker, J. R., Flaveny, C. A., Perdew, G. H., Denison, M. S., Schultz, P. G. and Cooke, M. P. 2010, *Science*, 329, 1345-1348.
41. Durso, D. F., Bacalini, M. G., Sala, C., Pirazzini, C., Marasco, E., Bonafe, M., do Valle, I. F., Gentilini, D., Castellani, G., Faria, A. M. C., Franceschi, C., Garagnani, P. and Nardini, C. 2017, *Oncotarget*, 8, 23237-23245.

42. Perna, L., Zhang, Y., Mons, U., Holleccek, B., Saum, K. U. and Brenner, H. 2016, *Clinical Epigenetics*, 8, 64.
43. Levine, M. E., Lu, A. T., Chen, B. H., Hernandez, D. G., Singleton, A. B., Ferrucci, L., Bandinelli, S., Salfati, E., Manson, J. E., Quach, A., Kusters, C. D., Kuh, D., Wong, A., Teschendorff, A. E., Widschwendter, M., Ritz, B. R., Absher, D., Assimes, T. L. and Horvath, S. 2016, *Proceedings of the National Academy of Sciences of the United States of America*, 113, 9327-9332.
44. Nevalainen, T., Kananen, L., Marttila, S., Jylhava, J., Mononen, N., Kahonen, M., Raitakari, O. T., Hervonen, A., Jylha, M., Lehtimaki, T. and Hurme, M. 2017, *Clinical Epigenetics*, 9, 20.
45. Quach, A., Levine, M. E., Tanaka, T., Lu, A. T., Chen, B. H., Ferrucci, L., Ritz, B., Bandinelli, S., Neuhauser, M. L., Beasley, J. M., Snetselaar, L., Wallace, R. B., Tsao, P. S., Absher, D., Assimes, T. L., Stewart, J. D., Li, Y., Hou, L., Baccarelli, A. A., Whitsel, E. A. and Horvath, S. 2017, *Aging*, 9, 419-446.
46. Ward-Caviness, C. K., Nwanaji-Enwerem, J. C., Wolf, K., Wahl, S., Colicino, E., Trevisi, L., Kloog, I., Just, A. C., Vokonas, P., Cyrus, J., Gieger, C., Schwartz, J., Baccarelli, A. A., Schneider, A. and Peters, A. 2016, *Oncotarget*, 7, 74510-74525.
47. Gao, X., Zhang, Y., Breitling, L. P. and Brenner, H. 2016, *Oncotarget*, 7, 46878-46889.
48. Zheng, S. C., Widschwendter, M. and Teschendorff, A. E. 2016, *Epigenomics*, 8, 705-719.
49. Varela-Rey, M., Woodhoo, A., Martinez-Chantar, M. L., Mato, J. M. and Lu, S. C. 2013, *Alcohol Research : Current Reviews*, 35, 25-35.
50. Reynolds, L. M., Wan, M., Ding, J., Taylor, J. R., Lohman, K., Su, D., Bennett, B. D., Porter, D. K., Gimple, R., Pittman, G. S., Wang, X., Howard, T. D., Siscovick, D., Psaty, B. M., Shea, S., Burke, G. L., Jacobs, D. R. Jr., Rich, S. S., Hixson, J. E., Stein, J. H., Stunnenberg, H., Barr, R. G., Kaufman, J. D., Post, W. S., Hoeschele, I., Herrington, D. M., Bell, D. A. and Liu, Y. 2015, *Circulation. Cardiovascular Genetics*, 8, 707-716.
51. Zhang, Y., Schottker, B., Ordonez-Mena, J., Holleccek, B., Yang, R., Burwinkel, B., Butterbach, K. and Brenner, H. 2015, *International Journal of Cancer*, 137, 1739-1748.
52. Teschendorff, A. E., Yang, Z., Wong, A., Pipinikas, C. P., Jiao, Y., Jones, A., Anjum, S., Hardy, R., Salvesen, H. B., Thirlwell, C., Janes, S. M., Kuh, D. and Widschwendter, M. 2015, *JAMA Oncology*, 1, 476-485.
53. Philibert, R., Hollenbeck, N., Andersen, E., Osborn, T., Gerrard, M., Gibbons, F. X. and Wang, K. 2015, *Frontiers in Psychology*, 6, 656.
54. Vineis, P., Chatziioannou, A., Cunliffe, V. T., Flanagan, J. M., Hanson, M., Kirsch-Volders, M. and Kyrtopoulos, S. 2017, *FASEB Journal*, [Epub ahead of print].
55. Baglietto, L., Ponzi, E., Haycock, P., Hodge, A., Bianca Assumma, M., Jung, C. H., Chung, J., Fasanelli, F., Guida, F., Campanella, G., Chadeau-Hyam, M., Grankvist, K., Johansson, M., Ala, U., Provero, P., Wong, E. M., Joo, J., English, D. R., Kazmi, N., Lund, E., Faltus, C., Kaaks, R., Risch, A., Barrdahl, M., Sandanger, T. M., Southey, M. C., Giles, G. G., Johansson, M., Vineis, P., Polidoro, S., Relton, C. L. and Severi, G. 2017, *International Journal of Cancer*, 140, 50-61.
56. Dick, K. J., Nelson, C. P., Tsaprouni, L., Sandling, J. K., Aissi, D., Wahl, S., Meduri, E., Morange, P. E., Gagnon, F., Grallert, H., Waldenberger, M., Peters, A., Erdmann, J., Hengstenberg, C., Cambien, F., Goodall, A. H., Ouwehand, W. H., Schunkert, H., Thompson, J. R., Spector, T. D., Gieger, C., Tregouet, D. A., Deloukas, P. and Samani, N. J. 2014, *Lancet*, 383, 1990-1998.
57. Ronn, T., Volkov, P., Gillberg, L., Kokosar, M., Perfilyev, A., Jacobsen, A. L., Jorgensen, S. W., Brons, C., Jansson, P. A., Eriksson, K. F., Pedersen, O., Hansen, T., Groop, L., Stener-Victorin, E., Vaag, A., Nilsson, E. and Ling, C. 2015, *Human Molecular Genetics*, 24, 3792-3813.
58. Hair, B. Y., Xu, Z., Kirk, E. L., Harlid, S., Sandhu, R., Robinson, W. R., Wu, M. C., Olshan, A. F., Conway, K., Taylor, J. A. and Troester, M. A. 2015, *Breast Cancer Research and Treatment*, 151, 453-463.
59. Christensen, B. C., Kelsey, K. T., Zheng, S., Houseman, E. A., Marsit, C. J., Wrensch, M. R.,

- Wiemels, J. L., Nelson, H. H., Karagas, M. R., Kushi, L. H., Kwan, M. L. and Wiencke, J. K. 2010, *PLoS Genetics*, 6, e1001043.
60. Campesi, I., Sanna, M., Zinellu, A., Carru, C., Rubattu, L., Bulzomi, P., Seghieri, G., Tonolo, G., Palermo, M., Rosano, G., Marino, M. and Franconi, F. 2012, *Biology of Sex Differences*, 3, 4.
 61. Bahl, A., Pollanen, E., Ismail, K., Sipila, S., Mikkola, T. M., Berglund, E., Lindqvist, C. M., Syvanen, A. C., Rantanen, T., Kaprio, J., Kovanen, V. and Ollikainen, M. 2015, *Twin Research and Human Genetics*, 18, 647-661.
 62. Izadi, P., Noruzinia, M., Fereidooni, F., Mostakhdemine Hosseini, Z. and Kamali, F. 2014, *Molecular Biology Reports*, 41, 7607-7612.
 63. Terry, M. B., Ferris, J. S., Pilsner, R., Flom, J. D., Tehranifar, P., Santella, R. M., Gamble, M. V. and Susser, E. 2008, *Cancer Epidemiology, Biomarkers & Prevention*, 17, 2306-2310.
 64. Crider, K. S., Yang, T. P., Berry, R. J. and Bailey, L. B. 2012, *Advances in Nutrition*, 3, 21-38.
 65. Larsson, S. C., Giovannucci, E. and Wolk, A. 2007, *Journal of the National Cancer Institute*, 99, 64-76.
 66. Lewis, S. J., Harbord, R. M., Harris, R. and Smith, G. D. 2006, *Journal of the National Cancer Institute*, 98, 1607-1622.
 67. Flanagan, J. M., Wilson, A., Koo, C., Masrou, N., Gallon, J., Loomis, E., Flower, K., Wilhelm-Benartzi, C., Hergovich, A., Cunnea, P., Gabra, H., Braicu, E. I., Sehoul, J., Darb-Esfahani, S., Vanderstichele, A., Vergote, I., Kreuzinger, C., Cacsire Castillo-Tong, D., Wisman, G. B., Berns, E. M., Siddiqui, N., Paul, J. and Brown, R. 2017, *Clinical Cancer Research*, 23, 2213-2222.
 68. Cuzzo, C., Porcellini, A., Angrisano, T., Morano, A., Lee, B., Di Pardo, A., Messina, S., Iuliano, R., Fusco, A., Santillo, M. R., Muller, M. T., Chiariotti, L., Gottesman, M. E. and Avvedimento, E. V. 2007, *PLoS Genetics*, 3, e110.
 69. O'Hagan, H. M., Mohammad, H. P. and Baylin, S. B. 2008, *PLoS Genetics*, 4, e1000155.
 70. Pleasance, E. D., Stephens, P. J., O'Meara, S., McBride, D. J., Meynert, A., Jones, D., Lin, M. L., Beare, D., Lau, K. W., Greenman, C., Varela, I., Nik-Zainal, S., Davies, H. R., Ordóñez, G. R., Mudie, L. J., Latimer, C., Edkins, S., Stebbings, L., Chen, L., Jia, M., Leroy, C., Marshall, J., Menzies, A., Butler, A., Teague, J. W., Mangion, J., Sun, Y. A., McLaughlin, S. F., Peckham, H. E., Tsung, E. F., Costa, G. L., Lee, C. C., Minna, J. D., Gazdar, A., Birney, E., Rhodes, M. D., McKernan, K. J., Stratton, M. R., Futreal, P. A. and Campbell, P. J. 2010, *Nature*, 463, 184-190.
 71. Savage, K. I., Matchett, K. B., Barros, E. M., Cooper, K. M., Irwin, G. W., Gorski, J. J., Orr, K. S., Vohhodina, J., Kavanagh, J. N., Madden, A. F., Powell, A., Manti, L., McDade, S. S., Park, B. H., Prise, K. M., McIntosh, S. A., Salto-Tellez, M., Richard, D. J., Elliott, C. T. and Harkin, D. P. 2014, *Cancer Research*, 74, 2773-2784.
 72. Carroll, J. S. 2016, *European Journal of Endocrinology*, 175, R41-49.
 73. Joehanes, R., Just, A. C., Marioni, R. E., Pilling, L. C., Reynolds, L. M., Mandaviya, P. R., Guan, W., Xu, T., Elks, C. E., Aslibekyan, S., Moreno-Macias, H., Smith, J. A., Brody, J. A., Dhingra, R., Yousefi, P., Pankow, J. S., Kunze, S., Shah, S. H., McRae, A. F., Lohman, K., Sha, J., Absher, D. M., Ferrucci, L., Zhao, W., Demerath, E. W., Bressler, J., Grove, M. L., Huan, T., Liu, C., Mendelson, M. M., Yao, C., Kiel, D. P., Peters, A., Wang-Sattler, R., Visscher, P. M., Wray, N. R., Starr, J. M., Ding, J., Rodriguez, C. J., Wareham, N. J., Irvin, M. R., Zhi, D., Barrdahl, M., Vineis, P., Ambatipudi, S., Uitterlinden, A. G., Hofman, A., Schwartz, J., Colicino, E., Hou, L., Vokonas, P. S., Hernandez, D. G., Singleton, A. B., Bandinelli, S., Turner, S. T., Ware, E. B., Smith, A. K., Klengel, T., Binder, E. B., Psaty, B. M., Taylor, K. D., Gharib, S. A., Swenson, B. R., Liang, L., DeMeo, D. L., O'Connor, G. T., Herceg, Z., Ressler, K. J., Conneely, K. N., Sotoodehnia, N., Kardina, S. L., Melzer, D., Baccarelli, A. A., van Meurs, J. B., Romieu, I., Arnett, D. K., Ong, K. K., Liu, Y., Waldenberger, M., Deary, I. J., Fornage, M., Levy, D. and London, S. J. 2016, *Circulation Cardiovascular Genetics*, 9, 436-447.