Epigenome-based cancer risk prediction: rationale, opportunities and challenges

Author(s): Martin Widschwendter*,#, Allison Jones, Iona Evans, Daniel Reisel, Joakim Dillner, Karin Sundstrom, Ewout W. Steyerberg, Yvonne Vergouwe, Odette Wegwarth, Felix G. Rebitschek, Uwe Siebert, Gaby Sroczynski, Inez D. de Beaufort, Ineke Bolt, David Cibula, Michal Zikan, Line Bjørge, Nicoletta Colombo, Nadia Harbeck, Frank Dudbridge, Anne-Marie Tasse; Bartha M. Knoppers, Yann Joly, Andrew E. Teschendorff, Nora Pashayan.

Author affiliations:
Department of Women's Cancer, Institute for Women's Health, University College London, London, United Kingdom (MW, AJ, IE, DR, AET).
Department of Applied Health Research, Institute of Epidemiology and Healthcare, University College London, United Kingdom (NP).
Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden (JD, KS).
Karolinska University Laboratory, Karolinska University Hospital (JD).
Center for Medical Decision Sciences, Department of Public Health, Erasmus MC, Rotterdam, The Netherlands (ES, YV).
Department of Medical Statistics, LUMC, Leiden, The Netherlands (ES).
Max Planck Institute for Human Development, Harding Center for Risk Literacy, Berlin, Germany (OW, FR).
Institute of Public Health, Medical Decision Making and Health Technology Assessment, Department of Public Health, Health Services Research and HTA, UMIT-University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria (US, GS).
Oncotyrol: Center for Personalized Medicine, Innsbruck, Austria (GS).
Department of Medical Ethics and Philosophy of Medicine, Erasmus Medical Center, Rotterdam, The Netherlands (IDdB, IB).
Department of Gynaecological Oncology, Charles University - First Faculty of Medicine, Prague, Czech Republic (DC, MZ).
Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway and Centre for Cancer Biomarkers, Department of Clinical Science and, University of Bergen, Bergen (LB).
European Institute of Oncology and University Milan-Bicocca, Milan, Italy (NC)
Breast Center, Department of Gynaecology and Obstetrics, Ludwig-Maximilians Universität, University of Munich, Munich, Germany (NH).
Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom (FD).
Department of Health Sciences, University of Leicester, Leicester, United Kingdom (FD).

Centre of Genomics and Policy, McGill University, Montreal, Canada (BMK, YJ).
Public Population Project in Genomics and Society, McGill University and Genome Quebec Innovation Centre, Montreal, Canada (AMT).

*on behalf of the FORECEE(4C) Consortium (www.forecee.eu).
#Correspondence to MW (M.Widschwendter@ucl.ac.uk).

ABSTRACT (148 words)

The incidence of cancer is continuing to rise and risk-tailored early diagnostic or primary prevention strategies are urgently required. Risk predictive tests should (i) integrate both genetic and non-genetic factors captured by an omics-technology that is biologically stable and technically reproducible, (ii) derive a score from easily accessible biological samples that act as surrogate for the organ in question and (iii) allow the efficacy of risk reducing measures to be monitored. Substantial evidence has accumulated suggesting that the epigenome and in particular DNA methylation (DNAme) based tests may meet these requirements. However, developing and implementing DNAme based risk predictive tests pose considerable challenges. Cell-type specificity of DNAme and cell-type heterogeneity in easily accessible surrogate cells requires novel methods to account for confounding issues. Engagement of the scientific community with healthcare professionals, policy makers and the public is required to address the organisational, ethical, legal, social and economic challenges.

TEXT (6316 words)

Introduction

Cancer is a leading cause of mortality worldwide, accounting for 14.1 million new cases and 8.2 million deaths in 2012[1]. It has been estimated that global cancer burden will increase yearly to 20.3 million new cases and 13.2 million deaths by 2030 [2]. Environmental, behavioural and life style risk factors [3], genetic predisposition, and acquisition of random mutations can lead to cancer development [3-5]. Prevention and early detection remain the key interventions to reduce global cancer burden.

Almost all cancers occur against a background of individual risk factors including environmental, lifestyle, reproductive and heritable genetic factors. High penetrance genetic mutations are rare in the population and account for only a small proportion of cases. Nevertheless, most common cancers have a heritable component spread across thousands of common germline variants each conferring small risk increments[6]. Genome-wide association studies have revealed common variants that explain a small fraction of heritability (Table 1). The remainder of the heritability may eventually be found through ever larger association studies, but more immediately it may be effectuated through -omics intermediates (e.g. epigenomics) that have stronger, more direct effects on cancer occurrence [7].

Considering the heterogeneity of risk in the population, tailoring preventive and early detection interventions to an individual's risk level could improve the efficacy of population-based programmes in prevention and early detection of cancer [8]. In the prostate cancer setting, for example, targeting screening to men at higher than population average risk could reduce the proportion of men likely to be over-diagnosed and, consequently, over-treated [9,10].

Currently, several biomarker tests and complementary statistical models have been developed to predict cancer risk (Table 2). With notable exceptions, such as a model based on HPV-DNA testing to predict a precursor of cervical cancer (CIN2+)[11], most risk models include only epidemiological factors. The discriminative ability of these models in separating low from high risk subjects is modest, as expressed by the area under the Receiver Operating Characteristic curve (ROC AUC, a measure of discriminatory accuracy of a model and the probability that a test correctly identifies an individual who will develop the disease from a pair of whom one will be affected and one will remain unaffected; AUC values range from 0.5 which is a total lack of discrimination to 1.0 which is a perfect discrimination). In addition, current models do not typically differentiate in terms of

prognosis, which is vital for tailored screening and primary prevention (i.e. early detection or prevention of those cancers which would otherwise lead to death).

A predictive test should (i) **integrate genetic and non-genetic factors** captured by an omics-technology that is both biologically stable and technically reproducible (ii) **derive a risk prediction score, using easily accessible tissues,** that is relevant for cancer development or is able to capture risk-inducing factors and ideally (iii) has the added potential **of enabling monitoring of the efficacy of potential risk reducing measures**. The basis of this strategy is drawn from the discipline of cardiovascular medicine. Risk prediction and tailored chemoprevention for non-symptomatic individuals have been fundamental in the dramatic reduction in mortality from myocardial infarction and stroke [12]. The cardiovascular community has accepted the principle that freedom from symptoms does not equate to a guarantee of health and the use of 'surrogate end points' was central to their success. Both blood pressure and cholesterol concentration can (i) be easily assessed by non-invasive measures, (ii) act as surrogates for an individual's interaction with environmental factors (i.e. stress, nutrition, smoking, absence of physical exercise, etc.) and (iii) are key components of multivariable risk algorithms[13]. It is also well known that phenotypic variability between different populations takes place both at the genetic and epigenetic levels, indicating that epigenetic modification substantially contributes to natural human variation [14]. Correspondingly, we propose a novel population-based screening methodology that relies upon epigenetics as a surrogate marker for risk prediction.
We will discuss the potential of DNA methylation (DNAme) markers to predict the risk of developing specific cancers and highlight the importance of Epigenome-Wide Association Studies (EWAS). Since epigenetic changes are tissue specific one of the biggest challenges is to identify easily accessible surrogate cells and develop algorithms to assess cell heterogeneity. In addition, we will address the legal, ethical and economic challenges along with other aspects associated with the implementation of epigenetic tests into the clinical and public health arena.

Epigenetics in cancer development: Epigenetic traits can be mitotically and also meiotically (i.e. transgenerationally) inherited, but unlike genetics they are not conferred by the sequence of bases defining the genetic code. Epigenetics is rather defined by a collective of dynamic processes that fine tune and regulate gene expression. As such epigenetics can be considered the 'editor' of the genome, affording our cells their identity and providing genomic plasticity, particularly at key time points in early development[15], in the maintenance of adult select tissues and in response to lifetime environmental exposures. Three interacting components – DNA methylation (DNAme), histone modification and non-coding RNA – are

integral to epigenetic regulation and function in a tissue specific manner. Methylation of the C^5 position of cytosines within the CpG dinucleotide (DNAme) context is technically and biologically the most stable component of the epigenome and is modified both by inherent genetic components as well as all non-heritable factors which shape living organisms[16]. Nevertheless CpGs represent at best ~2% of the spatial genome and are notably concentrated within short stretches of DNA in gene promoters known as 'CpG islands'. In cancer tissue hypermethylation of CpG islands against a background of global hypomethylation, both associated with skewed genetic expression, are hallmarks of epigenetic modulation witnessed across a multitude of cancer types.

Over recent years it has become evident that epigenetic mis-programming constitutes a core component of cancer initiation and progression. A key involvement of epigenetic de-regulation in cancer development has been the observation that DNA regions, that under normal conditions are specifically marked and transiently silenced by Polycomb-Group (PCG) proteins in stem cells, become methylated and completely silenced in cancer[17-19]. This led to the proposal of an 'epigenetic stem cell model' of cancer whereby cells acquiring DNA methylation at (Polycomb-Group Target (PCGT) genes become erroneously de-differentiated and subsequently prone to somatic mutations. Such targeted DNAme can be mediated by a specific non-coding RNA (e.g. HOTAIR[20-22]) that interacts with Polycomb Repressive Complex 2 (PRC2). HOTAIR links the PRC2 component EZH2 (Enhancer of zeste homolog 2), a histone-lysine N-methyltransferase enzyme, to histone H3 in order to catalyse the addition of methyl groups to lysine 27 (H3K27) which eventually leads to DNAme in the corresponding region[23-25] (Figure 1). There is ample proof demonstrating that PCGT methylation is a prerequisite for cells to transform into cancer cells[23,24,26-28] and that PCGT methylation seems to accumulate in stem cells as a function of cell divisions which is strongly associated with cancer risk [29-31]. The proportion to which epigenetic and genetic alterations contribute to cancer formation has not been assessed, but it has become clear that only stem cells (which are epigenetically determined) have the capacity to survive oncogene-induced substantial DNA damage[32] (Figure 2A). Recent research demonstrates that epigenetic contribution to cancer progression is far more complex than originally appreciated. Studies have shown that PCGT methylation in cells other than those from which the cancer originates can influence cancer development. For example, *HAND2*, a gene located downstream of the progesterone pathway, is highly expressed during the luteal phase in the endometrial stroma and affects the attenuation of oestrogen-mediated paracrine proliferation signals from the stroma that target endometrial epithelial cells[33]. *HAND2* methylation and silencing in the normal endometrial stroma leads to functional oestrogen dominance that results in complex atypical hyperplasia of the endometrium[34] (Figure 2B) of which approximately 50% progress to a full blown cancer[35]. Observational evidence in other cancer entities support the view that

epigenetic changes in the morphological normal stroma contribute to cancer initiation and progression[36-40]. Although not yet experimentally proven it is tempting to speculate that epigenetic alterations in cell-nonautonomous contributors to cancer development (e.g. immune cells and organs which provide endocrine signals) play important roles. Early evidence indicating that PCGT methylation (i.e. *HOX* gene family methylation in normal endometrium) is strongly associated with the presence of a cancer in an anatomically distant organ (i.e. ovarian cancer)[41] provided preliminary proof of concept suggesting that DNAme analyses in more easily accessible cells could be used to predict the risk of developing cancer.

Besides methylation at PCGT, there are a variety of other examples describing how epigenetic alterations contribute to cancer development. For instance, a large number of trait-associated genetic variants have been shown to affect DNAme levels at different CpG sites including binding sites of a variety of transcription factors (such as *NFKB1* and *CTCF*) which are known to be involved in cancer formation [7]. The importance of this mechanism is further evidenced by the fact that those CpG sites which demonstrate aberrant DNAme in colorectal cancer are substantially enriched for those genetic variants which are discovered by genome-wide association studies comparing individuals with and without colorectal cancer [42]. The fact that methylated cytosines are substantially more prone to undergo spontaneous deamination [43] and mutations at CpG sites are frequently observed in cancer [44,45] provides another example of how aberrant DNAme contributes to cancer development.

Effects of cancer-predisposing factors on the epigenome:

The epigenome, specifically DNAme, is shaped by both heritable and non-heritable factors which are also known to have a substantial impact on cancer development (Figure 1) and therefore hold great promise as an objective surrogate for these factors.

<u>Genome-Epigenome interaction:</u>

The mechanisms by which inherited common sequence variations lead to cancer are largely uncharted, but may become manifest through their impact on the epigenome in three different ways:

- *Cell autonomous impact - genetic variants impact directly on the epigenome:* Allele-specific methylation may be associated with methylation quantitative trait loci (meQTL), single nucleotide polymorphisms (SNPs) that associate with the methylation status of specific sites or entire regions[7,46-48]. To date, numerous meQTL have been discovered utilising novel tools[49]. While efforts to relate meQTL to disease processes are still at an early stage, DNA methylation represents one plausible downstream effect of SNPs

on disease that may be directly measured to achieve greater accuracy in risk modelling. This is supported by the finding that regions at previously reported and, as-yet, unidentified cancer risk polymorphisms show aberrant DNA methylation [50].

- *Cell non-autonomous impact:* High-penetrance germline mutations (e.g. *BRCA* mutations) modulate endocrine factors[51-56] (e.g. higher oestrogen and progesterone production in the ovary) which then impact specifically on the epigenome of cells receptive to these signals, in the case of oestrogen, tubal or breast epithelial cells[57]. These changes are typically tissue type-dependent.
- *"Genetic environmental filter" impact:* The activity of enzymes involved in the metabolism of exogenous substances is largely determined by inherited genetic polymorphisms and will determine, in part, the effect of environmental exposures on the epigenetic makeup as evidenced by, for example, *CYP2A6* genotype and nicotine/cotinine clearances[58] and the linear relationship between serum cotinine, a major metabolite of nicotine, and DNA methylation[59].

Transgenerational inheritance:

The phenomenon of transmitting information from one generation to the next affecting the traits of offspring without altering the germline sequence of the nucleotides (i.e. epigenetically) has been repeatedly demonstrated[60,61]. For example, access to food [62] and exposure to smoking[63] early in life have repeatedly been demonstrated to impact on future generation's phenotypes. There is substantial evidence that DNAme of the POMC gene is transmitted via the paternal germline leading to an increased risk of developing obesity later in life[64] and that parental diet can affect cholesterol and lipid metabolism in offsprings[65]. It is well established that body mass index (BMI) is strongly associated with human cancer risk[66] and obesity itself also seems to trigger epigenetic alterations [67].

In-utero environment:

Many women who were exposed *in utero* to diethylstilbestrol (DES), a synthetic non-steroidal oestrogen provided to their mothers, have a substantially elevated risk of cervical intraepithelial neoplasia, breast cancer and clear cell vaginal cancer decades later [68]. DES upregulates HOTAIR [69] and leads to hypermethylation of *HOXA10* [70], a key gene involved in female genital tract development, in DES-exposed offspring. Together these findings suggest that DES' carcinogenic potential is mediated via epigenetic mechanisms. Effects of foetal exposure to other endocrine-disrupting chemicals including Bisphenol A have demonstrated carcinogenic effects to varying degrees[71] and are at least, in part, epigenetically transmitted[72].

Obesity:

Obesity is thought to have a substantial and direct impact on the epigenome[67]. The epigenetic effects are reflected in a program for shared drivers for cancer progression in organs such as the endometrium[73], liver[74], breast[75] and colon[76], the very same organs at increased risk of developing cancer in obese individuals[77]. Obesity, is likely to cause chronic low-grade inflammation[78], and potentially mediates its impact on DNA methylation via oxidative damage induced formation and re-localisation of epigenetic silencing complexes to stem cell PCGTs [79]. These processes are likely to differ among obese individuals with identical BMIs in accordance with their innate environmental response[67]. Habitual changes that lead to a reduction of obesity (i.e. caloric restriction) substantially slows the epigenetic clock[80,81] with a resultant decrease in cancer risk[82-84].

Smoking:

Exposure to cigarette smoke triggers striking epigenetic changes. Hypomethylation of genes involved in toxin response pathways such as *AHRR*, *CYP1A1*, and *CYP1B1*[85-89] has been observed across different tissues though most of the evidence comes from blood and buccal tissue. Since hypomethylation of these genes is not consistently observed in cancer[85] these epigenetic changes may not be causally involved in cancer progression. Smoking-triggered hypermethylation of genes bivalently marked in human stem cells (i.e. PCGT genes) is predominantly observed in epithelial (e.g. buccal) cells[85]. A smoking index constructed using these hypermethylated sites is highly efficient in discriminating between normal and cancerous tissues[85].

Microbiome and virome impact on the host epigenome:

Infections with certain bacteria or viruses have been identified as strong risk factors for specific human cancers[90] and alterations in microbiota may contribute to human carcinogenesis[91] . Mono- or polymicrobial factors can cause changes in the human host mediated through genetics, epithelial injury, immune system function and/or inflammation [91]. Microbiota have also been shown to affect oestrogen metabolism[92]. The microbiome appears to affect the epigenome through DNAme dependent pathways in the host[93]. For example, gut bacteria can provide epigenetically active metabolites essential for DNAme such as folate, butyrate and acetate, as well as enzymes and cofactors for epigenetic processes [92].

Chronic inflammation:

Some cancers develop due to chronic inflammatory insults[94]. Carcinogenesis associated with inflammatory bowel disease, reflux oesophagitis, pancreatitis or pelvic inflammatory disease converge at the level of the transcription factors nuclear factor-κB (NF-κB) and signal

transducer and activator of transcription 3 (STAT3) which lead to epigenetic reprogramming in epithelial cells of the affected organ[57,95-97]. Again, the majority of genes affected by inflammation-mediated reprogramming are PCGT genes [98,99].

Hormones and DNAme:

Absolute levels of hormones, dynamics over time (e.g. throughout the menstrual cycle) as well as relative levels across various hormones (e.g. oestrogen/progesterone balance) contribute to the cancer risk of hormone sensitive organs[52,53,100-103]. Steroid hormones are key regulators of genes involved in epigenetic programming (AID[104], DNMTs, EZH2, etc.). Dramatic changes in the systemic hormonal environment – as for example during the menopause – lead to substantial epigenetic changes, which are in part, cell type specific[105]. In addition, proxy indicators for endogenous prenatal testosterone exposure (i.e. the anogenital distance [106] or the ratio of digit length [107]) are associated with prostate cancer risk, consistent with the view that androgens also leave an epigenetic imprint which, after several decades, lead to a specific phenotype.

Age:

Age contributes to the cancer risk of a given tissue/organ in two ways: a cell-intrinsic, tissue-dependent, way that increases with the number of stem-cell divisions, and a cell-extrinsic way that increases in line with the cumulative exposure to environmental risk factors (e.g. smoking, obesity mediated inflammation, viral infections) [3,108-110]. Both components increase with chronological age, and are intricately linked; cumulative exposure to cancer risk factors is thought to accelerate the stem-cell division rate of tissues [108]. In addition, DNA methylomes at the two extremes of the human lifespan (i.e. new-borns and centenarians) are distinct in the same subset of cells [111]. Like somatic mutations and copy number variations (CNVs), DNAme alterations gradually accumulate with chronological age [112-114] and with exposure to cancer risk factors independently of age [115]. These factors are thought to reflect cell-intrinsic (e.g. stem-cell division) and cell-extrinsic (e.g. metabolically induced) factors contributing to the molecular damage of tissues. Thus, specific DNAme changes in the tissue of origin (or suitable surrogates), may be informative of cancer risk, as demonstrated in the context of cervical cancer[116]. Supporting this further, an epigenetic mitotic-like clock ("EpiTOC") [29], which correlates with the cumulative number of stem-cell divisions in the tissue of origin, is universally accelerated in cancer tissues and pre-neoplastic lesions, again offering promise for cancer-risk prediction [29,116]. In contrast, Horvath's epigenetic clock, a tissue-independent non-mitotic clock which measures chronological age [117] [29], appears to be less informative with respect to cancer risk [29,115].

Current evidence from EWAS:

There is substantial evidence for the existence of epigenetic field defects i.e. aberrant epigenetic signatures in normal tissue adjacent to the cancer[57,118-121]. Within EWAS a genome-wide set of quantifiable epigenetic marks (i.e. DNA methylation) in different individuals will be analysed with the aim of deriving associations between epigenetic variation and a particular identifiable phenotype/trait. Analogous to the genome-wide association studies (GWAS)[122] we propose that a minimum of 100,000 CpGs per individual are analysed in order to apply the term "epigenome-wide". When compared with GWAS, several additional challenges exist. Notwithstanding the correct choice of easy to access surrogate tissue, the modifiable character of epigenetic markers creates difficulties in discriminating between cause and consequence and must therefore be taken into account when considering the timing of the sample collection in relation to the manifestation of the disease. Unlike GWAS where variants at single nucleotide positions are associated with a specific trait, the basis of EWAS is to quantify methylation at CpGs across the genome in a given sample and rank these sites according to their different methylation levels between cases and controls. To date, both EWAS and studies looking at a predefined sets of CpGs have been performed.

Two principal categories of epigenetic risk predictors exist.

- Category 1 – DNAme markers of "extrinsic risk exposure": These are DNAme markers that reflect exposure to specific exogenous carcinogens. The magnitude of the impact on DNAme reflects the individual response and acts as a surrogate marker for the development of cancer in an individual. For example, there is dose-dependency of methylation levels of CpGs in the *AHRR* or *F2RL3* gene with smoking pack-years[85,123] which is a quantitative measure of active lifetime tobacco exposure. Demethylation at the *AHRR* or *F2RL3* CpG site (1st versus 4th quartile) was associated with a 16- and 11- fold increased risk for lung cancer respectively even after adjusting for a variety of factors including current smoking status and duration[124,125]. These findings have been validated by independent studies based on different cohorts[126]. Importantly, the top ten smoking–associated CpGs in blood surpassed the performance of the top ten lung-cancer-related CpGs in blood with regard to predicting lung cancer mortality[127]. So far there is no clear evidence that aberrant methylation of *AHRR* observed in the surrogate tissue (i.e. blood or buccal cells) of smokers who are predisposed to lung cancer development actually drives cancer development in the tissue at risk (i.e. lung epithelial cells); functional work on *AHRR* methylation in lung cell models will need to be carried out.

A recent EWAS demonstrated that BMI is associated with substantial DNAme changes in blood samples and that these associations are mainly a consequence of obesity, not the cause of it[67]. Obese individuals in the top quartile of the methylation risk score had a 10-fold increased risk of developing type 2 diabetes in the future compared with those in the

lowest quartile [67]. The observation that genes involved in oestrogen response (e.g. in p53 and NF-kB pathways) were enriched amongst the obesity-associated genes implies that an obesity-associated DNAme signature is capable of predicting the incidence of obesity-associated cancers, irrespective of the actual individual BMI at the time of assessment.
Epigenetic age acceleration (i.e. the deviation of epigenetic age from the actual chronological age) assessed in peripheral blood was associated with cancer incidence[82] and mortality[82,128] in general and, specifically, with postmenopausal breast[84] or lung[83] cancer susceptibility.

- Category 2 - DNAme markers of "intrinsic risk": Most known DNAme markers predicting cancer risk have been discovered based on case control or population-based nested case control settings and have not as yet been linked to extrinsic risk factors.

More than a decade ago, anecdotal reports[129,130] provided initial evidence that DNA methylation of the mismatch repair gene *MLH1* in normal cells is present in individuals with multiple cancers. Early reports indicated that loss of imprinting of *IGF2* in lymphocytes is predictive of colorectal cancer risk[131] but studies using DNAme in peripheral blood predating diagnosis could not confirm these findings[132].
The first large study (sample size larger than 1000 cases and controls) provided a direct link between DNAme of the oestrogen-receptor interacting *ZNF217* gene, serum oestrogen receptor alpha bioactivity and breast cancer risk[133]. These data and the majority of data referenced in this section (apart from those referenced in Table 3) have been generated based on the analysis of biological material (i.e. surrogate tissue) derived from prevalent (i.e. already existing) cases; this comes with several challenges as outlined in the following example: The first study analysing a larger number of CpGs - approximately 25,000 CpGs (i.e. Illumina's 27k methylation array) - was conducted in blood from ovarian cancer patients and non-cancer control women[134] and concluded that the timing of sample collection for DNAme analysis and adjustment for sample cell-type composition is essential for valid interpretation of results (see chapter "Tissue specificity of the epigenome" for more details). Another study using the same assay derived a DNAme signature from the peripheral blood of *BRCA1* mutation carriers, which was significantly enriched for PCGT hypermethylation and predicted breast cancer incidence and death independently of family history or other known risk factors[135].
To date, only a very limited number of studies have acknowledged the tissue specificity of the DNA methylome. The majority of ovarian cancers are derived from cells arising from the Fallopian Tube, the latter of which shares the same developmental origin as the endometrium. DNAme of *HOXA9*, a gene essential for differentiation of the Fallopian Tube,

is substantially increased in the normal endometrium of ovarian cancer patients, but not in the adjacent myometrium, the non-epithelial component of the uterus[41].

In the context of cervical cancer screening, the uterine cervix is one of the very few organs that allows for the assessment of normal cells years in advance of the onset of any cytological/histological changes. A DNAme signature derived from cytological normal samples which predate a diagnosis of cervical intraepithelial neoplasia grade 2 or 3 (CIN2+) by three years[116] discriminated cytologically normal cells from CIN2+ smears with a ROC AUC of 0.69-0.87 and a normal uterine cervix from an invasive cervical cancer with a AUC of 0.94[136].

Numerous additional studies (all carried out in whole blood samples or a subset of blood cells) have found evidence of different global[137] or gene specific DNAme in samples collected from testicular[137], ovarian[138,139], colorectal[140], breast[141,142], head and neck[143], melanoma[144,145] and renal[146] cancer patients and cancer-free controls.

An increasing number of studies have identified and/or validated DNAme markers with the help of population based cohorts predicting the development of breast[147-150], bladder[151,152] or hepatocellular cancer[153,154].

Cancer prevention:

Unlike genetic markers, epigenetic markers are modifiable and not only potentially indicate the risk of developing a certain cancer disease but, importantly, can also be used in monitoring the response to preventive measures. A study of 1,092 healthy female volunteers showed that the methylation rate of CpGs, related to colorectal cancer, show a reduced rate of methylation in individuals exposed to cancer-preventive drugs such as acetylsalicylic acid or hormone replacement therapy, and an increased rate of methylation in smokers and in women with high BMI[155]. The observation that time since cessation of smoking is reflected in the epigenome of easily accessible organs not primarily at risk for smoke-induced cancers[85,156,157] indicates that it may be feasible to monitor preventive strategies for inaccessible organs by means of DNAme in easy to access samples. Besides smoking, DNAme changes associated with obesity have also been shown to be similar between adipose and blood cells [67], further supporting this principle. Ongoing work will determine which easy to access surrogate tissue best reflects the epigenetic state in those organs at risk for which epigenetic field defects are likely drivers of carcinogenesis[57,118] – this is a long-term requirement for effectively monitoring cancer-preventive measures.

Tissue specificity of the epigenome:

Although the specific tissue from which the cancer arises would be the ideal target for the retrieval of cells with an epigenetic risk signature, apart from a few exceptions (e.g. cervical

smear for cervical cancer), it is not typically feasible to access the tissue at risk directly as this would require invasive procedures (e.g. bronchial lavage, biopsies of the breast, liver, pancreas, prostate, colon or Fallopian Tube). We therefore propose that surrogate tissue – from blood (i.e. normal blood cells), buccal and cervical cells (and possibly cells from urine) - to be used for this purpose. To date, the vast majority of analyses have been undertaken in blood cells as these samples are readily available in various cohorts (Table 3).

The fact that the tissues used in EWAS represent complex mixtures of many underlying cell-types whereas DNA methylation is cell-type specific [158,159], poses a significant challenge to the analysis and interpretation of EWAS data [160], not encountered in GWAS. For instance, many cancer EWAS conducted in whole blood and peripheral blood have demonstrated that most DNAme changes between cancer cases and controls can be attributed to shifts in the granulocyte/monocyte to lymphocyte proportions, reflecting a specific and major immune-response to the presence of cancer [143,152,161,162]. In women with primary ovarian cancer or residual disease after chemotherapy, such shifts in DNAme provided highly accurate predictions of cancer-status (AUC>0.8) [162]. However, when assessing ovarian cancer patients who had undergone chemotherapy and who did not have evidence of residual disease (ovarian cancer serum marker CA125 < 35 U/mL), DNAme profiles were largely indistinguishable from age-matched controls [162]. While DNAme changes associated with such shifts in cell-type composition could be useful for general diagnostic purposes, they do not represent epigenetic alterations which may potentially drive carcinogenesis. Identifying the latter requires the inference of differentially methylated CpGs (DMCs) that are not driven by underlying changes in cell-type composition. To help address this challenge, efforts such as the IHEC [163] and BLUEPRINT [164] are underway generating reference DNAme profiles for all major human cell-types. These reference DNAme profiles, although derived from specific individuals (and thus potentially confounded by factors such as genotype and age), can be used in the deconvolution of bulk-tissue DNAme profiles [165], providing reasonably accurate estimates of underlying cell-type proportions in independent samples, as confirmed using matched FACS/MACS-based cell count data [166] (Figure 3). These cell-type fraction estimates can subsequently be used to adjust the DNAme data, allowing identification of DMCs that are not driven by changes in cell-type composition [165,166]. Using this approach, a recent meta-analysis of solid cancer EWAS conducted in blood, further confirmed that very few of the DMCs between cancers and controls remain after adjustment for cell-type composition [151]. Although these residual DMCs were found to map to cancer-related pathways [151], their interpretation and relevance for the corresponding cancer-type is still unclear. It is likely that further progress will require the identification of DNAme changes in either the cell of origin of the cancer, or in surrogate tissue/cells that more closely represent the cell of origin in epithelial cancers. Ongoing work will demonstrate whether a combination of the epigenomes

in several surrogate tissues [i.e. blood (capturing the contribution from the stroma/immune-system), cervical and buccal cells (capturing the hormone dependent and independent risk factors, respectively)] might provide the best accuracy.

Cell-free DNA in serum or plasma is currently used to monitor the efficacy of cancer treatment and identify therapy-resistant cancer clones. In this context, somatic genetic or epigenetic alterations which have accumulated in the cancer and are released into the liquid phase are analysed (i.e. "liquid biopsy"). This, by definition, is not useful for cancer-risk prediction as discussed in the context of this review. However, having said this, there is now some preliminary evidence that organ-specific DNAme patterns can be detected in cancer-free individuals [167,168]. Whether this cell-free DNA in plasma/serum can be used to assess future cancer-risk for specific organs needs to be determined once sufficiently large population-based cell-free DNA repositories (which are not massively contaminated with DNA released from blood cells) have become available and their donors followed up for a sufficient amount of time in order to identify those individuals who eventually developed a cancer.

In summary, tissue specificity is a hallmark of the epigenome. The vast majority of EWAS studies have been performed based on peripheral blood cells. To date, not one study has analysed several surrogate tissues (i.e. blood cells and buccal) from the same individuals at the same time in order assess which surrogate tissue is best suited to predict future risk for a specific cancer entity. Thus far, it is also unclear whether epigenetic profiles in blood cells (i.e. the vast majority of EWAS were based on blood epigenomes) are (i) a surrogate of the epigenome in the tissue at risk or (ii) purely an indication of the epigenetic status of immune-cells (and thereby reflective of their anti-neoplastic capacity) or a combination of (i) and (ii).

Translational Challenges:

The development of epigenome-based risk predictors in surrogate tissues face several significant challenges.

<u>Choice of DNAme analysis method:</u>

Box 1 describes the potential tools for discovering DNAme risk predictors and for clinical application of these markers. The choice of tool will depend on the size and costs of the study, the heterogeneity of the samples as well as whether quantitative assessment of single CpG methylation or DNAme patterns in a specific region is required.

<u>Choice of surrogate tissue:</u>

Although recent studies have indicated that cancer risk prediction may be possible using DNAme profiles obtained in blood [83,169], prediction accuracies are low, and require further validation and have an unclear mechanistic basis. In the context of women-specific cancers, cervical smears, representing hormone-responsive tissue, are a more promising alternative. Cervical smears may serve to identify relevant epigenetic cancer-risk biomarkers not only for cervical cancers but also for endometrial and ovarian (due to their common embryological origin) as well as breast (hormonally-triggered) cancers in prospective case/control settings nested within larger prospective clinical trials. Buccal cells (epithelial cells directly exposed to smoke-toxins) may be the best surrogate tissue for predicting lung cancer risk and a urine sample containing epithelial cells from the urethra (the prostate's embryological origin) might be best suited for predicting prostate cancer risk.

Analytical challenges

The identification of DNAme alterations that may indicate cancer-risk is particularly challenging since the relevant comparison is between normal cells at risk and normal cells that are not. Such normal to normal tissue comparison is statistically challenging [116] owing to (i) technical confounders, (ii) biological confounders (e.g. cell-type heterogeneity), and (iii) the likely stochastic nature of DNAme changes preceding carcinogenesis.

Although technical confounders (e.g. batch effects) are frequently observed in -omic datasets [170], there are also many statistical algorithms that can successfully be used to adjust data for these confounders [171] [172,173]. Cervical smears, comprising various types of epithelial and immune cells, exhibit substantial variation in immune-cell fractions between unrelated women, making adjustment vital. Statistical methods, specifically designed for cell-type composition, have also been developed [165,174] and allow for the identification of DMCs not driven by changes in tissue composition (Figure 3).

In the context of cancer risk prediction, an additional statistical challenge arises because differences between normal cells and normal cells at risk of neoplastic transformation are expected to be infrequent and stochastic, which means that standard algorithms based on selecting DMCs may fail [116]. While cancer cells exhibit widespread changes in DNAme which are identifiable using DMC approaches and account for most of the variation in the data [17,118], precursor cancer cells exhibit a much more heterogeneous and stochastic pattern [116,175]. This is possibly due to normal cells not having undergone neoplastic transformation and consequently not being selected for. A recent proof-of-principle study, conducted in the context of cervical cancer, confirmed the aforementioned[116]; it demonstrated that the DNAme patterns of normal cervical smears from women who developed a CIN2+ lesion three years after sample collection could only be distinguished from those of women who remained (pre)cancer-free and only if one adopts a radically different statistical feature selection

paradigm which selects for CpGs with heterogeneous and stochastic patterns, the so called Differentially Variable CpGs (DVCs). Such DVCs manifest as outlier DNAme events that are only seen in a very small fraction of the women who later developed CIN2+. While DVCs appear to be stochastically distributed across independent individuals, the pattern is distinctively non-random across the genome of any individual, highlighting that there are specific regions of the genome that are more susceptible to inter-individual variation in DNAme, as previously observed[116,176-178]. Thus, as shown in the context of cervical carcinogenesis [116], risk prediction may be possible by measuring the accumulation of deviations in DNAme from the normal state across a well-defined set of DVC loci, an approach called EVORA (Epigenetic Variable Outliers for Risk prediction Analysis) [116,179]. While the EVORA framework awaits further validation, independent strong evidence for its validity was obtained recently in the context of breast cancer, by comparing normal breast tissue from women to the normal breast tissue adjacent to breast cancers [175]: EVORA could distinguish normal tissue from breast cancer patients from that of healthy women with an AUC of 0.84.

Sample size:

The search for epigenomic risk markers is often hampered by the analysis of relatively small sample sets, caused by high costs. Consequently, spurious associations between CpGs and cancer risk may be found, and true associations may be exaggerated. The ideal scenario of comprehensive data from a single large-scale, prospective cohort study may not be reached. The evidence-base for associations may be increased by also considering results from other prospective study designs that include only incident cases, matched to well-defined, population-based controls (Table 3). Such studies allow unbiased estimation of relative risks. Applying simulations for EWAS[180] and calculations based on our data[85,116,118,119,135] suggest that 300 cases and 300 controls are sufficient to discover differentially methylated CpGs. Validation studies with independent, population-based data are required to confirm any associations and to validate absolute risks that apply to the general population.

Data storage and sharing:

Adopted by the European Union in 2016 and coming into effect in 2018, the General Data Protection Regulation (GDPR)[181] provides legal guidance for the management of privacy risks based on the data types (e.g. personal data, genetic data, data concerning health, biometric data or sensitive data), levels of identifiability (anonymous, pseudonymised or identifiable data) and data uses (e.g. clinical care, research). While anonymous data fall outside the purview of the GDPR, sharing of pseudonymised (e.g. coded) and identifiable data is strictly regulated.

Therefore, in the context of the epigenetic risk prediction test, the main challenge for the scientific community would appear to be characterising the identifiability of epigenomic information. Does epigenomic information allow for the identification of a natural person, directly or indirectly? Should it be considered as 'personal', 'sensitive', 'genetic' or health-related information? Such questions are key when addressing the specific issues raised by sharing epigenomic information.

Challenges to implement epigenome-based risk predictors as a clinical tool:

Combining genetic variants with environmental and lifestyle risk factors would improve risk stratification. The use of epigenetic changes captures the interaction of observed and unobserved risk factors at each individual's cellular level[182], while the assessment of these risk factors via questionnaires and retrospective self-reporting is of limited reliability and susceptible to, for example, recall bias[183].

The implementation of risk-tailored cancer prevention and early diagnostic programmes is a multi-step process and raises a number of challenges for policymakers and the public they serve (Figure 4). The organisational challenges to be addressed include providing equitable access to risk assessment and risk-tailored interventions, preparing and training the workforce, building an infrastructure for assessing the quality of tests and services, and developing IT platforms and data storage capacity. Using epigenome-based risk assessment poses additional organisational challenges due to the plasticity of the epigenome that requires repeated risk assessment over time and varying intervention recommendations according to risk levels. Based on the available data on smoking and methylation (i.e. DNAme changes as a function of accumulating pack-years and of time after cessation of smoking[85,156]) we speculate that an epigenetic risk predicting test will have to be repeated every 3-5 years in order to re-calculate the risk.

<u>Ethical issues:</u>

The epigenome acts as a surrogate readout for heritable and lifestyle factors, raising several issues: (i) Personal responsibility and healthy lifestyle; how much responsibility can be attributed to the individual and to what extent individuals can be held accountable for their health? (ii) Safeguarding autonomous decision-making; how to guarantee that individuals are making a voluntary and well-considered informed choice for or against a test comprising complex information about risks for different diseases with varying ages of onset. (iii) Risk profiles for one cancer might encapsulate information for other conditions. For example signatures for cervical or breast cancer might reflect the individual response to smoking and obesity and as such also indicate the risks for lung cancer[124-126,156] or type

2 diabetes[67], respectively. This requires new informed consent paradigms (e.g. tiered, staged models)[184], shared decision-making and novel patient decision tools[185].

Legal issues:

The development of genomics and other -omics sciences, including epigenomics, has eroded the once clear boundary that existed between research and clinical care. This new ''translational'' space is conducive to improving healthcare but also raises legal issues due to the reversibility of epigenetic risk factors and the dynamic, sometimes transgenerational, nature of epigenetic data. Relevant legal issues include: (i) Consolidation of a cost-efficient pathway for regulatory approval of new epigenetic tests. (ii) Clarification of the limits of liability for researchers and clinicians (e.g. when returning research results or incidental findings, including epigenetic test results to the medical file, and updating patients on important changes in epigenetic material). (iii) Clarification of privacy and confidentiality rights of the patient vs. those of family members (e.g. siblings, children, etc.); and (iv) Promotion of equality while promoting the data sharing necessary for advancing epigenetic science[186].

Risk communication:

To assess the risk of individuals requires informed consent and the provision and communication of evidence-based information in lay language. Some of the communication challenges associated with epigenome-based risk assessment are identical to already existing tests. Individuals need to be informed upfront (e.g. by fact boxes[187]) concerning their age-adjusted baseline risks, the benefit-harm-ratio of having or not having the test, and the modified benefit-harm ratios of current cancer screening approaches and prevention as a consequence of the test [188]. Epigenetic screening, however, has additional layers of complexity; individuals need to be informed about the complex cancer-specific interplay of genes, environment, and behaviour and additionally that testing for epigenetic factors will reveal some of their past environmental exposures (i.e. smoking, alcohol, etc.). It will therefore be essential that the healthcare workforce is trained in interpretation [189-192] and communication of risk prediction test results.

Decision analysis to evaluate the relationship between benefits and harms:

Scientific evidence needs to demonstrate additional benefit for a new risk-tailored screening or prevention strategy, with an acceptable benefit-harm ratio and cost-effectiveness ratio when compared to current standards of care [193-195]. Decision-analytic modelling is a useful quantitative approach for synthesising the best available scientific evidence such as epidemiologic parameters, test performance, prognosis, treatment effectiveness, quality of life, and economic data. It is also useful to evaluate the trade-off between benefits, harms,

and costs of alternative interventional strategies [196-198]. Decision-analytic models simulate the development of the disease, and the consequences of different screening/prevention strategies including specific medical pathways [197,199] (Figure 5).

Adaptation of the currently established infrastructure:

The leveraging of already existent screening programmes is a key opportunity for rapid real-life evaluation and roll-out of new tests. In most high-resource settings, the infrastructure for cancer screening programmes is already available and could be used for new -omic frontiers in prevention. Such programmes have the inherent potential to test new biomarkers through so-called randomised health services studies (RHS; [200]). Once evaluated by a RHS design, new screening tests – if found to be superior to the old policy – could be immediately implemented since the programme has already been part of the testing phase.

Conclusions and future directions:

Epigenetic based risk models provide state-of-the-art opportunities for personalised medicine and risk-level-tailored interventions to improve human health through the reduction of cancer burden. Although several significant challenges have been identified and further research is required, such risk models are potentially feasible and, when available, would likely meet most criteria needed for effective risk prediction, i.e. the ability to:

(i) encapsulate both genetic and non-genetic risk referring factors using a single -omics platform which is biologically stable and technically reproducible;

(ii) derive a predictive score using easily accessible tissues which are relevant for cancer development or are able to capture risk-referring signals;

(iii) be used to monitor the efficacy of risk reducing measures.

Development and implementation of epigenomic-based cancer prevention and screening/early detection programmes requires international collaboration between multidisciplinary teams with expertise in -omics, bioinformatics, epidemiology, public health, economics, decision analysis, ethics, law, risk communication and engagement of the scientific community with healthcare professionals, policy makers, and the public. In order to develop epigenomic-based cancer prevention, multidisciplinary research through international consortia is needed to overcome the various scientific challenges.

ACKNOWLEDGMENTS

The authors are supported by the European Union's Horizon 2020 Programme (H2020/2014-2020) under grant agreement number 634570 (Project FORECEE: www.forecee.eu/), and

MW also receives support from the European Research Council (ERC Advanced Grant ERC-BRCA) and by The Eve Appeal (www.eveappeal.org.uk/). The authors would also like to acknowledge the support of the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre.

COMPETING INTERESTS STATEMENT

The authors have no competing interests.

TABLES:

Table 1. Percentage of variance in liability for several common cancers

Disease is assumed to arise from a liability threshold model, in which each individual has an unobserved, normally distributed liability that results in disease when it exceeds a threshold. Heritability; variance explained by heritable factors, estimated from twin or family studies. Known genes; variance explained by established risk genes including findings from genome-wide association studies. Environmental, variance explained by environmental exposures.

	Percentage of Variance (95% CI)		
Cancer	**Heritability**	**Known genes**	**Environmental**
Ovary[6,201]	22 (0-41)	1 (1-1)	78 (59-100)
Endometrium[6,202]	24 (14-87)	0 (0-0)	76 (13-86)
Lung[6,201]	26 (0-49)	2 (2-2)	74 (51-100)
Breast[6,201]	27 (4-41)	8 (0-21)	73 (59-96)
Cervical[203,204]	27 (26-29)	2 (0-5)	78 (71-74)
Colorectal[201,205]	35 (10-48)	1[a] (1-1)	65 (52-90)
Pancreas[6,201]	36 (0-53)	2 (2-2)	64 (47-100)
Kidney[6,202]	38 (21-55)	3 (3-3)	62 (45-79)
Prostate[6,201]	42 (29-50)	22 (0-93)	58 (50-71)
Melanoma[6,202]	58 (43-73)	9 (9-9)	42 (57-27)

[a] converted from recurrence risk to liability scale using formula given by Wray et al[206]

Table 2. **Examples of currently recognised and validated risk prediction models**

Cancer type and/or model	Phase (development/ validation/impact assessment)	Endpoint (any cancer /progressive cancer)	Predictors included in the final model	Discriminative ability (AUC)
Breast[207] *IBIS model*	Validation	Any	Age, BMI, age at menarche, age at first birth, age at menopause, no. of breast biopsies, atypical hyperplasia, lobular carcinoma in situ, family history of breast/ovarian cancer	0.76[a]
Ovarian[208]	Validation	Any	Age, Oral contraceptive use, menopausal hormone therapy use, parity, family history of breast/ovarian cancer	0.59[a]
Cervical[11]	Development	CIN1/CIN2+	High DNA-load of high-risk HPV, age, married status, smoking, age at sexual debut	0.76 CIN1[b], 0.90 CIN2+[b]
Prostate[209] *ERSPC risk calculator*	Impact assessment	Any	Ultrasound volume, digital rectal exam, transrectal ultrasound, PSA	0.76[a]
Lung[210] *$PLCO_{M2012}$ model*	Validation	Any	Age, race, education, BMI, COPD, personal history of cancer, family history of lung cancer, smoking status, smoking duration, smoking intensity, years since cessation	0.69-0.79[a]
Esophageal[211]	Validation	Any	age, sex, smoking status, body mass index, highest level of education, frequency of use of acid suppressant medications	0.61[a]
Colorectal[212]	Validation	Any	Sigmoidoscopy results, colonoscopy results, history of polyps, relative with CRC, aspirin/nonsteroidal anti-inflammatory drug use, smoking, vegetables, body mass index, leisure time activity (men only),	0.61[a]

		leisure exercise time (women only), oestrogen status (women only)

[a] Performance at external validation

[b] Performance at internal validation

AUC, Area Under the ROC Curve.

Table 3. Studies predicting risk for incident cancers using DNAme markers

Only studies using population–based samples with incident cancers (i.e. volunteers cancer-free at the time of sample collection) were used irrespective of how many CpGs were analysed.

Cancer	Source of DNA	Technique	Markers	Numbers	Remarks	Study Design
Breast[150]	Blood	Bisulfite pyrosequencing	ATM	640 cases, 741 controls	top quintile OR 1.89	Nested case-control/case-control
Breast[149]	Blood	Illumina 27k array	250 CpGs	298 cases, 612 controls	AUC 0.66	Nested case-cohort
Breast[135]	Blood	Illumina 27k & 450k array	1829 CpGs	210 cases, 271 controls	AUC 0.67 for fatal breast cancer	Case-control
Breast[147]	Blood	Illumina 450k array	mean beta values across all CpGs	420 cases, 420 controls	top quartile OR 0.42	Nested case-control
Breast[148]	Blood	Illumina 450k array	mean beta values across all CpGs	358 cases, 358 controls	top quartile OR 0.34 and 0.99 for 2 studies	Nested case-control
Breast[84]	Blood	Illumina 450k array	353 CpG age signature	451 cases, 451 controls	1 unit increase of epigenetic age acceleration leads to 4% increased breast cancer risk	Nested case-control

Lung[83]	Blood	Illumina 450k array	353 CpG age signature	43 cases, 1986 controls	1 unit increase of epigenetic age acceleration leads to 50% increased lung cancer risk	Case-control
Lung[156]	Blood	Illumina 450k array	AHRR, F2RL3	789 cases, 789 controls	AUC 0.76 adjusted for smoking	Nested case-control/case-control
Lung[213]	Blood	MassARRAY	F2RL3	318 cases, 4669 controls	AUC 0.77	Cohort study
Lung[124]	Blood	Bisulfite pyrosequencing	AHRR, F2RL3	143 cases, 453 controls	bottom quartile OR 15.9 (AHRR) and 10.55 (F2RL3) adjusted for smoking and other factors	Nested case-control
Lung[125]	Blood	Real time PCR	AHRR	352 cases, 8859 controls	bottom quintile HR 4.9	Cohort study
Cervical (pre-invasive)[116]	Cervical	Illumina 27k array	140 CpGs	77 incident CIN2+ cases and 77 controls	EVORA algorithm in normal cervical cells predicts future risk to develop CIN2+	Nested case-control
Liver[154]	Blood	MethyLight, Bisulfite pyrosequencing	Sat2, LINE-1	305 cases, 1254 controls	logSat2 1 unit decrease adjusted OR 1.77	Nested case-control
Liver[153]	Blood	Illumina 450k array	WNK2, TPO, MYT1L	159 cases, 312 controls	OR (above vs below median) 1.91 (WNK2), 0.59 (TPO), 0.50 (MYT1L)	Nested case-control
Various[82]	Blood	Illumina 450k array	71 CpG age signature	132 cases, 310 controls (2 samples from most volunteers)	one year increase of epigenetic compared to chronological age leads to 6% increased cancer and 17% mortality risk	Cohort study

BOX:

Box 1| **Potential methods for the assessment of the DNA methylome for risk predicting purposes**

The majority of technologies used to quantify DNA methylation rely on the principle of sodium bisulfite-induced deamination of unmethylated cytosine to uracil, followed by either microarray or sequencing as a read-out.

For discovery (i.e. feature selection):

- **Whole Genome Bisulfite Sequencing (WGBS):** A labour intensive method involving DNA fragmentation, ligation of adapters, purification of ligation products, bisulfite modification (BM), polymerase chain reaction (PCR), and sequencing. Theoretically, WGBS is able to capture all CpGs in the genome at single nucleotide resolution.
- **Reduced Representation Bisulfite Sequencing (RRBS):** Sequencing method that enriches for CpG rich regions of the genome, by digesting genomic DNA with Msp1. RRBS covers 85% of CpG islands and 60% of promoters. Steps involve DNA digestion, end-repair, A-tailing, adapter ligation, fragment size selection, BM and sequencing.
- **Methylation Arrays:** Arrays targeted to the methylated regions (CpG islands) of the genome. The Methylation EPIC BeadChip (Illumina5), covers 99% of RefSeq genes and 95% of CpG islands and allows interrogation of >850,000 methylation sites. Arrays also rely on BM but is less labour-intensive than sequencing.
- **Affinity Enrichment methods:** Based on the affinity purification of methylated DNA regions using either an antibody directed against 5-methylcytosine (MeDIP6) or against methyl-binding proteins (MethylCap7). Isolated methylated DNA can be assessed by PCR, microarray or sequencing.

For clinical assays:

Clinical assays require a targeted approach, allowing for the screening of large sample sets but only covering the regions of interest. This allows for a reduction in work-load and over-all cost.

- **Custom Arrays:** Specific regions of the genome can be studied with custom designed arrays. Various companies (Illumina, Agilent, Roche) offer custom array services for the creation of targeted assays.

- **Targeted Bisulfite Sequencing:** Use of specifically designed primers and NGS technology for the analysis of targeted genomic regions of interest. Cost per sample is reduced, but single nucleotide resolution is maintained.
- **Pyrosequencing:** DNA sequencing based on the "sequencing by synthesis" principle. It relies on the detection of pyrophosphate release upon nucleotide incorporation. A light signal is generated that allows for quantitative methylation analysis.
- **Quantitative PCR:** Amplification of BM-DNA with fluorescent primers that hybridise to predefined methylated regions, such as, in Methylight or digital PCR.

FIGURES:

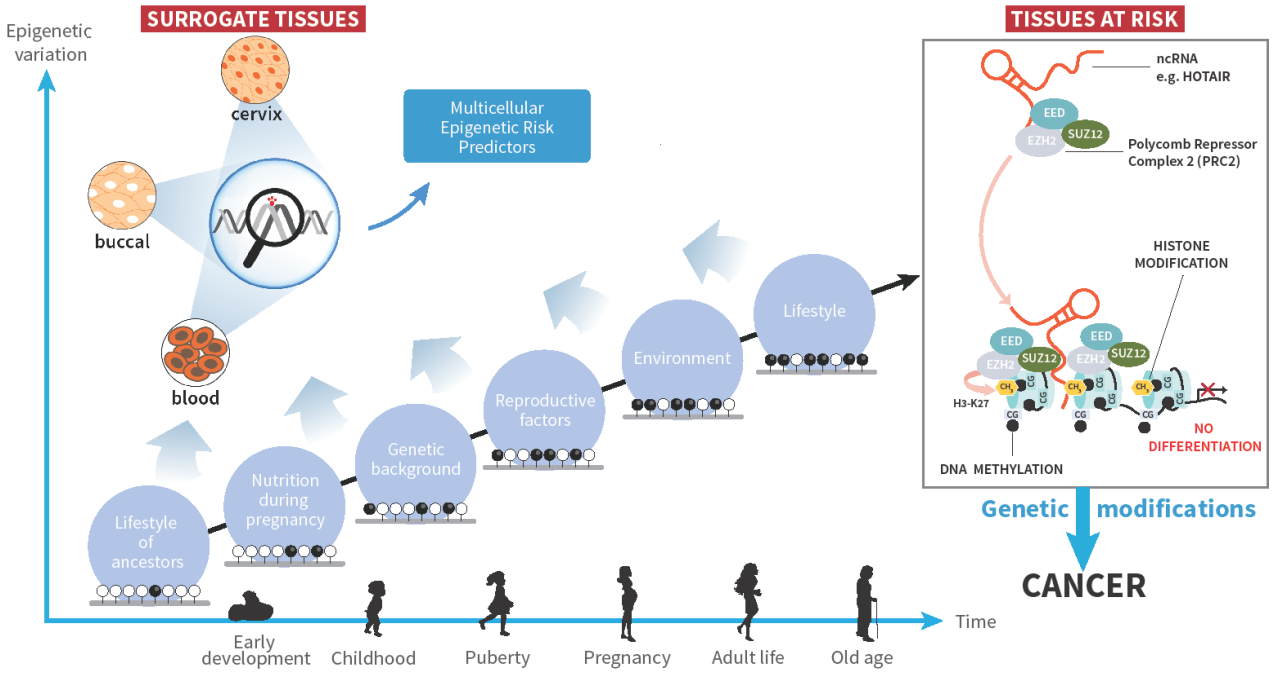
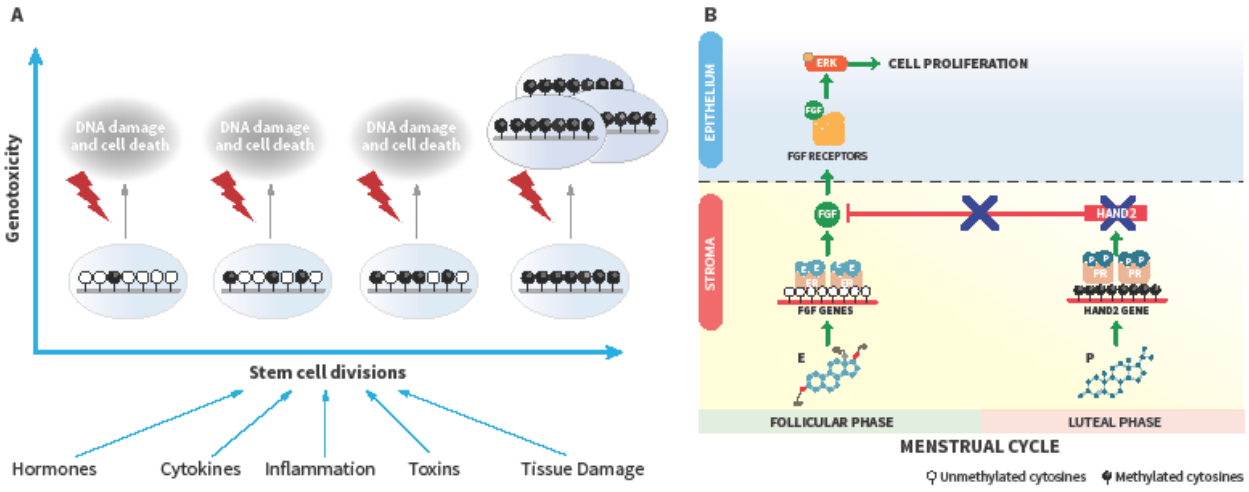
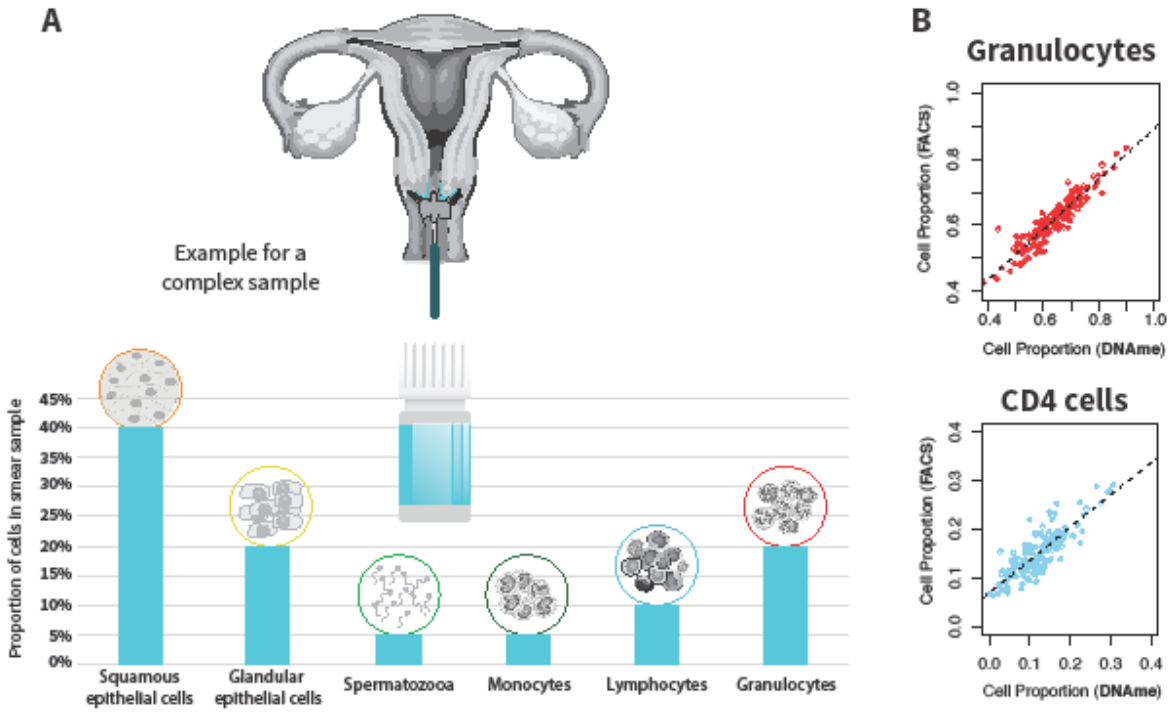


Figure 1. Multicellular epigenetic risk predictor. Factors that trigger epigenetic mis-programming in the inaccessible tissue at risk can be assessed in easily accessible surrogate tissue.



<u>Figure 2.</u> Examples illustrating how epigenetic alterations contribute to cancer development. (A) A general, epigenome-genome unifying concept of cancer formation: Accumulation of epigenetic alterations as a function of stem cell divisions may fix stem-ness, a state which is compatible with genotoxicity-induced DNA damage leading to cancer formation. **(B)** A specific example of epigenome-mediated cancer formation: Functional oestrogen dominance in epithelial cells due to epigenetic silencing of essential progesterone downstream gene *HAND2* in the endometrial stromal cells lead to precancerous complex atypical hyperplasia. E, oestrogen; P, progesterone; FGF, fibroblast growth factor; ERK, extracellular signal regulated kinase



<u>Figure 3.</u> The use of the epigenome in adjusting for sample heterogeneity. (A) Depiction of the potential cellular heterogeneity within a complex cervical smear sample. (B) Cell type specific DNAme signatures (x-axis) are used to predict the actual proportion of cell subtypes in a sample verified by FACS analysis (y-axis); the examples are given for granulocytes and CD4 lymphocytes in blood samples.

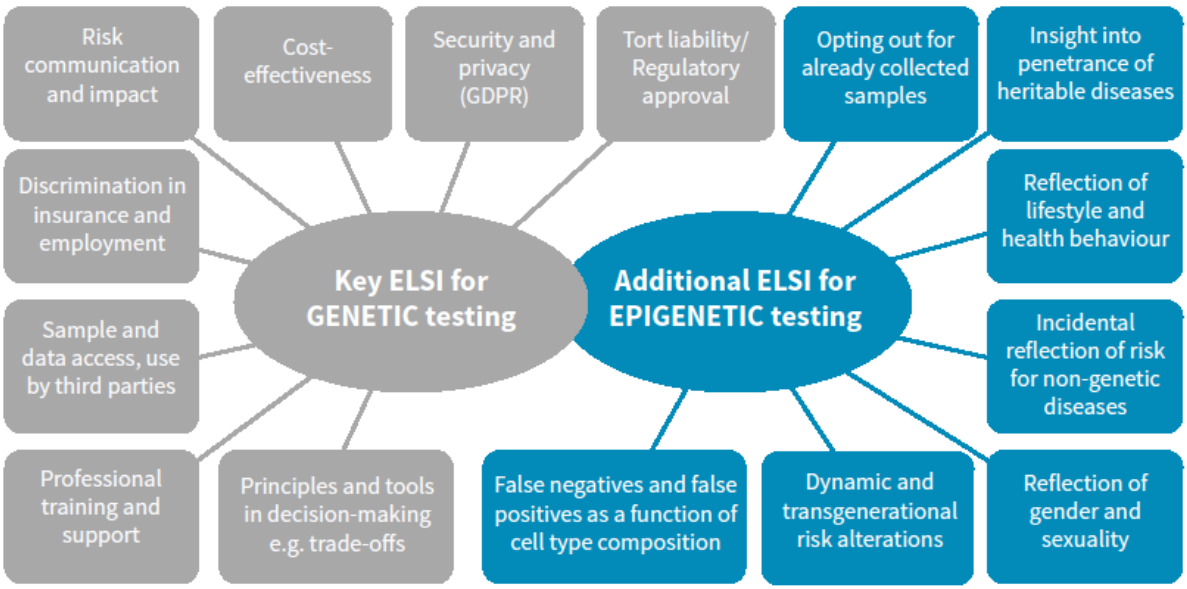
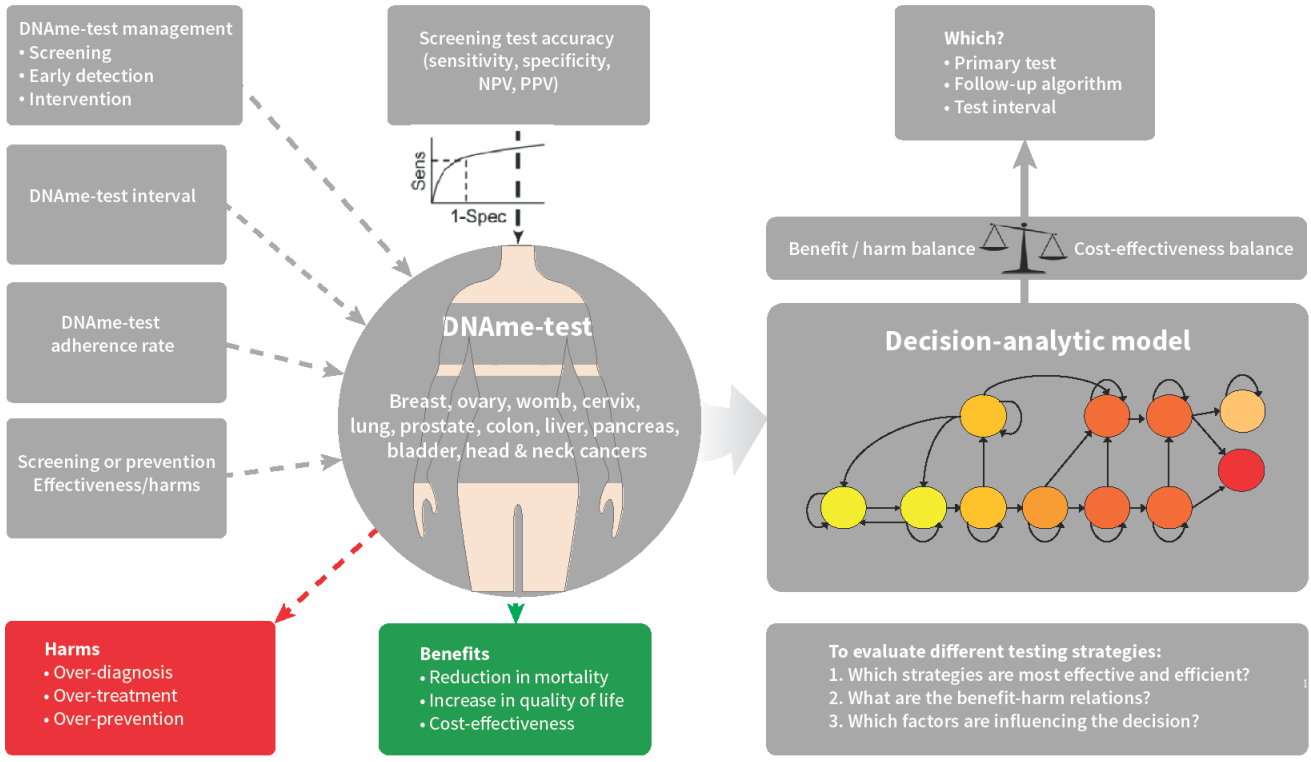


Figure 4. Organisational, ethical, legal and social issues (ELSI) to be considered when implementing epigenome-based risk predictors. 'Key ELSI' for risk-stratification based on the genome have already been identified by the COGS consortium[214] and 'Novel ELSI' are additional issues for the Women's cancer risk IDentification test (WID-test) specific to epigenome-based risk prediction tests.



<u>Figure 5.</u> Decision-analysis to evaluate the consequences of the DNA methylation (DNAme) test-based intervention strategies.

REFERENCES:

1 Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359-386, doi:10.1002/ijc.29210 (2015).

2 Thun, M. J., DeLancey, J. O., Center, M. M., Jemal, A. & Ward, E. M. The global burden of cancer: priorities for prevention. *Carcinogenesis* **31**, 100-110, doi:10.1093/carcin/bgp263 (2010).

3 Wu, S., Powers, S., Zhu, W. & Hannun, Y. A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529**, 43-47, doi:10.1038/nature16166 (2016).

4 Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330-1334, doi:10.1126/science.aaf9011 (2017).

5 Wodarz, D. & Zauber, A. G. Cancer: Risk factors and random chances. *Nature* **517**, 563-564, doi:10.1038/517563a (2015).

6 Lu, Y. *et al.* Most common 'sporadic' cancers have a significant germline genetic component. *Hum Mol Genet* **23**, 6112-6118, doi:10.1093/hmg/ddu312 (2014).

7 Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49**, 131-138, doi:10.1038/ng.3721 (2017).

8 Burton, H. *et al.* Public health implications from COGS and potential for risk stratification and screening. *Nat Genet* **45**, 349-351, doi:10.1038/ng.2582 (2013).

9 Pashayan, N. *et al.* Reducing overdiagnosis by polygenic risk-stratified screening: findings from the Finnish section of the ERSPC. *Br J Cancer* **113**, 1086-1093, doi:10.1038/bjc.2015.289 (2015).

10 Pashayan, N. *et al.* Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genet Med* **17**, 789-795, doi:10.1038/gim.2014.192 (2015).

11 Lee, C. H. *et al.* Risk evaluation for the development of cervical intraepithelial neoplasia: development and validation of risk-scoring schemes. *Int J Cancer* **136**, 340-349, doi:10.1002/ijc.28982 (2015).

12 Sporn, M. B. & Liby, K. T. Cancer chemoprevention: scientific promise, clinical uncertainty. *Nat. Clin. Pract. Oncol* **2**, 518-525, doi:ncponc0319 [pii];10.1038/ncponc0319 [doi] (2005).

13 Damen, J. A. *et al.* Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* **353**, i2416, doi:10.1136/bmj.i2416 (2016).

14 Heyn, H. *et al.* DNA methylation contributes to natural human variation. *Genome Res* **23**, 1363-1372, doi:10.1101/gr.154187.112 (2013).

15 Bergman, Y. & Cedar, H. DNA methylation dynamics in health and disease. *Nat. Struct. Mol. Biol* **20**, 274-281, doi:nsmb.2518 [pii];10.1038/nsmb.2518 [doi] (2013).

16 Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet* **13**, 97-109, doi:nrg3142 [pii];10.1038/nrg3142 [doi] (2011).

17 Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet* **39**, 157-158 (2007).
18 Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet* **39**, 232-236 (2007).
19 Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet* **39**, 237-242 (2007).
20 Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071-1076 (2010).
21 Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693 (2010).
22 Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007).
23 Laugesen, A., Hojfeldt, J. W. & Helin, K. Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. *Cold Spring Harb. Perspect. Med*, doi:cshperspect.a026575 [pii];10.1101/cshperspect.a026575 [doi] (2016).
24 Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14**, 735-751, doi:S1934-5909(14)00193-3 [pii];10.1016/j.stem.2014.05.006 [doi] (2014).
25 Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-874, doi:nature04431 [pii];10.1038/nature04431 [doi] (2006).
26 Wienken, M. *et al.* MDM2 Associates with Polycomb Repressor Complex 2 and Enhances Stemness-Promoting Chromatin Modifications Independent of p53. *Mol. Cell* **61**, 68-83, doi:S1097-2765(15)00942-9 [pii];10.1016/j.molcel.2015.12.008 [doi] (2016).
27 Zhuang, J. *et al.* The Dynamics and Prognostic Potential of DNA Methylation Changes at Stem Cell Gene Loci in Women's Cancer. *PLoS. Genet* **8**, e1002517, doi:10.1371/journal.pgen.1002517 [doi];PGENETICS-D-11-02221 [pii] (2012).
28 Iliou, M. S. *et al.* Bivalent histone modifications in stem cells poise miRNA loci for CpG island hypermethylation in human cancer. *Epigenetics* **6**, 1344-1353, doi:18021 [pii];10.4161/epi.6.11.18021 [doi] (2011).
29 Yang, Z. *et al.* Correlation of an epigenetic mitotic clock with cancer risk. *Genome biology* **17**, 205, doi:10.1186/s13059-016-1064-3 (2016).
30 Klutstein, M., Nejman, D., Greenfield, R. & Cedar, H. DNA Methylation in Cancer and Aging. *Cancer Res* **76**, 3446-3450, doi:10.1158/0008-5472.CAN-15-3278 (2016).
31 Klutstein, M., Moss, J., Kaplan, T. & Cedar, H. Contribution of epigenetic mechanisms to variation in cancer risk among tissues. *Proc Natl Acad Sci U S A* **114**, 2230-2234, doi:10.1073/pnas.1616556114 (2017).
32 Morel, A. P. *et al.* A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. *Nat Med* **23**, 568-578, doi:10.1038/nm.4323 (2017).
33 Li, Q. *et al.* The antiproliferative action of progesterone in uterine epithelium is mediated by Hand2. *Science* **331**, 912-916, doi:331/6019/912 [pii];10.1126/science.1197454 [doi] (2011).
34 Jones, A. *et al.* Role of DNA methylation and epigenetic silencing of HAND2 in endometrial cancer development. *PLoS. Med* **10**, e1001551, doi:10.1371/journal.pmed.1001551 [doi];PMEDICINE-D-13-01308 [pii] (2013).
35 Horn, L. C., Schnurrbusch, U., Bilek, K., Hentschel, B. & Einenkel, J. Risk of progression in complex and atypical endometrial hyperplasia: clinicopathologic analysis in cases with and without progestogen treatment. *Int J Gynecol Cancer* **14**, 348-353, doi:10.1111/j.1048-891x.2004.014220.x (2004).
36 Hanson, J. A. *et al.* Gene promoter methylation in prostate tumor-associated stromal cells. *J Natl Cancer Inst* **98**, 255-261, doi:10.1093/jnci/djj051 (2006).
37 Valcz, G. *et al.* Myofibroblast-derived SFRP1 as potential inhibitor of colorectal carcinoma field effect. *PLoS One* **9**, e106143, doi:10.1371/journal.pone.0106143 (2014).
38 Fiegl, H. *et al.* Breast cancer DNA methylation profiles in cancer cells and tumor stroma: association with HER-2/neu status in primary breast cancer. *Cancer Res* **66**, 29-33, doi:10.1158/0008-5472.CAN-05-2508 (2006).
39 Paterson, R. F. *et al.* Molecular genetic alterations in the laser-capture-microdissected stroma adjacent to bladder carcinoma. *Cancer* **98**, 1830-1836, doi:10.1002/cncr.11747 (2003).
40 Lin, H. J. *et al.* Breast cancer-associated fibroblasts confer AKT1-mediated epigenetic silencing of Cystatin M in epithelial cells. *Cancer Res* **68**, 10257-10266, doi:10.1158/0008-5472.CAN-08-0288 (2008).

41 Widschwendter, M. *et al.* HOXA methylation in normal endometrium from premenopausal women is associated with the presence of ovarian cancer: a proof of principle study. *Int. J. Cancer* **125**, 2214-2218 (2009).
42 Ongen, H. *et al.* Putative cis-regulatory drivers in colorectal cancer. *Nature* **512**, 87-90, doi:10.1038/nature13602 (2014).
43 Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C. & Gehrke, C. W. DNA cytosine methylation and heat-induced deamination. *Biosci Rep* **6**, 387-393 (1986).
44 Poulos, R. C., Olivier, J. & Wong, J. W. H. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. *Nucleic Acids Res* **45**, 7786-7795, doi:10.1093/nar/gkx463 (2017).
45 Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat Genet* **47**, 1402-1407, doi:10.1038/ng.3441 (2015).
46 Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* **17**, 61, doi:10.1186/s13059-016-0926-z (2016).
47 Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414 e1324, doi:10.1016/j.cell.2016.10.026 (2016).
48 Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* **12**, R10, doi:10.1186/gb-2011-12-1-r10 (2011).
49 Zeng, H. & Gifford, D. K. Predicting the impact of non-coding variants on DNA methylation. *Nucleic Acids Res*, doi:10.1093/nar/gkx177 (2017).
50 Heyn, H. *et al.* Linkage of DNA methylation quantitative trait loci to human cancer risk. *Cell Rep* **7**, 331-338, doi:10.1016/j.celrep.2014.03.016 (2014).
51 Liu, Y. *et al.* A Mouse Model That Reproduces the Developmental Pathways and Site Specificity of the Cancers Associated With the Human BRCA1 Mutation Carrier State. *eBioMedicine* **2**, 1318-1330, doi:10.1016/j.ebiom.2015.08.034 [doi];S2352-3964(15)30117-1 [pii] (2015).
52 Widschwendter, M. *et al.* Osteoprotegerin (OPG), The Endogenous Inhibitor of Receptor Activator of NF-kappaB Ligand (RANKL), is Dysregulated in BRCA Mutation Carriers. *eBioMedicine* **2**, 1331-1339, doi:10.1016/j.ebiom.2015.08.037 [doi];S2352-3964(15)30119-5 [pii] (2015).
53 Widschwendter, M. *et al.* The sex hormone system in carriers of BRCA1/2 mutations: a case-control study. *Lancet Oncol* **14**, 1226-1232, doi:S1470-2045(13)70448-0 [pii];10.1016/S1470-2045(13)70448-0 [doi] (2013).
54 Chodankar, R. *et al.* Cell-nonautonomous induction of ovarian and uterine serous cystadenomas in mice lacking a functional Brca1 in ovarian granulosa cells. *Curr. Biol* **15**, 561-565, doi:S0960-9822(05)00161-2 [pii];10.1016/j.cub.2005.01.052 [doi] (2005).
55 Hong, H. *et al.* Changes in the mouse estrus cycle in response to BRCA1 inactivation suggest a potential link between risk factors for familial and sporadic ovarian cancer. *Cancer Res* **70**, 221-228, doi:0008-5472.CAN-09-3232 [pii];10.1158/0008-5472.CAN-09-3232 [doi] (2010).
56 Yen, H. Y. *et al.* Alterations in Brca1 expression in mouse ovarian granulosa cells have short-term and long-term consequences on estrogen-responsive organs. *Lab Invest* **92**, 802-811, doi:labinvest201258 [pii];10.1038/labinvest.2012.58 [doi] (2012).
57 Bartlett, T. E. *et al.* Epigenetic reprogramming of fallopian tube fimbriae in BRCA mutation carriers defines early ovarian cancer evolution. *Nat. Commun* **7**, 11620, doi:ncomms11620 [pii];10.1038/ncomms11620 [doi] (2016).
58 Benowitz, N. L., St Helen, G., Dempsey, D. A., Jacob, P., 3rd & Tyndale, R. F. Disposition kinetics and metabolism of nicotine and cotinine in African American smokers: impact of CYP2A6 genetic variation and enzymatic activity. *Pharmacogenet Genomics* **26**, 340-350, doi:10.1097/FPC.0000000000000222 (2016).
59 Zhang, Y., Florath, I., Saum, K. U. & Brenner, H. Self-reported smoking, serum cotinine, and blood DNA methylation. *Environ Res* **146**, 395-403, doi:10.1016/j.envres.2016.01.026 (2016).
60 Miska, E. A. & Ferguson-Smith, A. C. Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. *Science* **354**, 59-63, doi:10.1126/science.aaf4945 (2016).
61 Pembrey, M., Saffery, R., Bygren, L. O., Network in Epigenetic, E. & Network in Epigenetic, E. Human transgenerational responses to early-life experience: potential impact on development, health and biomedical research. *J Med Genet* **51**, 563-572, doi:10.1136/jmedgenet-2014-102577 (2014).

62 Bygren, L. O. *et al.* Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. *BMC Genet* **15**, 12, doi:10.1186/1471-2156-15-12 (2014).
63 Northstone, K., Golding, J., Davey Smith, G., Miller, L. L. & Pembrey, M. Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses. *Eur J Hum Genet* **22**, 1382-1386, doi:10.1038/ejhg.2014.31 (2014).
64 Kuhnen, P. *et al.* Interindividual Variation in DNA Methylation at a Putative POMC Metastable Epiallele Is Associated with Obesity. *Cell Metab* **24**, 502-509, doi:10.1016/j.cmet.2016.08.001 (2016).
65 Carone, B. R. *et al.* Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084-1096, doi:S0092-8674(10)01426-1 [pii];10.1016/j.cell.2010.12.008 [doi] (2010).
66 Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. & Zwahlen, M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* **371**, 569-578, doi:S0140-6736(08)60269-X [pii];10.1016/S0140-6736(08)60269-X [doi] (2008).
67 Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81-86, doi:10.1038/nature20784 (2017).
68 Hoover, R. N. *et al.* Adverse health outcomes in women exposed in utero to diethylstilbestrol. *N. Engl. J. Med* **365**, 1304-1314, doi:10.1056/NEJMoa1013961 [doi] (2011).
69 Bhan, A. *et al.* Bisphenol-A and diethylstilbestrol exposure induces the expression of breast cancer associated long noncoding RNA HOTAIR in vitro and in vivo. *J Steroid Biochem Mol Biol* **141**, 160-170, doi:10.1016/j.jsbmb.2014.02.002 (2014).
70 Bromer, J. G., Wu, J., Zhou, Y. & Taylor, H. S. Hypermethylation of homeobox A10 by in utero diethylstilbestrol exposure: an epigenetic mechanism for altered developmental programming. *Endocrinology* **150**, 3376-3382, doi:en.2009-0071 [pii];10.1210/en.2009-0071 [doi] (2009).
71 Soto, A. M. & Sonnenschein, C. Environmental causes of cancer: endocrine disruptors as carcinogens. *Nat Rev Endocrinol* **6**, 363-370, doi:10.1038/nrendo.2010.87 (2010).
72 Jorgensen, E. M., Alderman, M. H., 3rd & Taylor, H. S. Preferential epigenetic programming of estrogen response after in utero xenoestrogen (bisphenol-A) exposure. *FASEB J* **30**, 3194-3201, doi:10.1096/fj.201500089R (2016).
73 Kim, J. Y., Tavare, S. & Shibata, D. Counting human somatic cell replications: methylation mirrors endometrial stem cell divisions. *Proc Natl Acad Sci U S A* **102**, 17739-17744, doi:10.1073/pnas.0503976102 (2005).
74 Zhou, D. *et al.* High fat diet and exercise lead to a disrupted and pathogenic DNA methylome in mouse liver. *Epigenetics* **12**, 55-69, doi:10.1080/15592294.2016.1261239 (2017).
75 Rossi, E. L. *et al.* Obesity-Associated Alterations in Inflammation, Epigenetics, and Mammary Tumor Growth Persist in Formerly Obese Mice. *Cancer Prev Res (Phila)* **9**, 339-348, doi:10.1158/1940-6207.CAPR-15-0348 (2016).
76 Li, R. *et al.* Obesity, rather than diet, drives epigenomic alterations in colonic epithelium resembling cancer progression. *Cell Metab* **19**, 702-711, doi:S1550-4131(14)00116-8 [pii];10.1016/j.cmet.2014.03.012 [doi] (2014).
77 Bhaskaran, K. *et al.* Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet* **384**, 755-765, doi:10.1016/S0140-6736(14)60892-8 (2014).
78 Harvey, A. E., Lashinger, L. M. & Hursting, S. D. The growing challenge of obesity and cancer: an inflammatory issue. *Ann. N. Y. Acad. Sci* **1229**, 45-52, doi:10.1111/j.1749-6632.2011.06096.x [doi] (2011).
79 O'Hagan, H. M. *et al.* Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell* **20**, 606-619, doi:S1535-6108(11)00359-X [pii];10.1016/j.ccr.2011.09.012 [doi] (2011).
80 Wang, T. *et al.* Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol* **18**, 57, doi:10.1186/s13059-017-1186-2 (2017).
81 Cole, J. J. *et al.* Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol* **18**, 58, doi:10.1186/s13059-017-1185-3 (2017).

82 Zheng, Y. *et al.* Blood Epigenetic Age may Predict Cancer Incidence and Mortality. *EBioMedicine* **5**, 68-73, doi:10.1016/j.ebiom.2016.02.008 (2016).
83 Levine, M. E. *et al.* DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)* **7**, 690-700, doi:10.18632/aging.100809 (2015).
84 Ambatipudi, S. *et al.* DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur J Cancer* **75**, 299-307, doi:10.1016/j.ejca.2017.01.014 (2017).
85 Teschendorff, A. E. *et al.* Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. *Jama Oncol* **1**, 476-485, doi:2293216 [pii];10.1001/jamaoncol.2015.1053 [doi] (2015).
86 Philibert, R. A., Beach, S. R. & Brody, G. H. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. *Epigenetics* **7**, 1331-1338, doi:10.4161/epi.22520 (2012).
87 Zeilinger, S. *et al.* Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* **8**, e63812, doi:10.1371/journal.pone.0063812 (2013).
88 Joubert, B. R. *et al.* 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* **120**, 1425-1431, doi:10.1289/ehp.1205412 (2012).
89 Wan, E. S. *et al.* Smoking-Associated Site-Specific Differential Methylation in Buccal Mucosa in the COPDGene Study. *Am J Respir Cell Mol Biol* **53**, 246-254, doi:10.1165/rcmb.2014-0103OC (2015).
90 de Martel, C. *et al.* Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol* **13**, 607-615, doi:10.1016/S1470-2045(12)70137-7 (2012).
91 Schwabe, R. F. & Jobin, C. The microbiome and cancer. *Nat Rev Cancer* **13**, 800-812, doi:10.1038/nrc3610 (2013).
92 Paul, B. *et al.* Influences of diet and the gut microbiome on epigenetic modulation in cancer and other diseases. *Clin Epigenetics* **7**, 112, doi:10.1186/s13148-015-0144-7 (2015).
93 Alenghat, T. Epigenomics and the microbiota. *Toxicol Pathol* **43**, 101-106, doi:10.1177/0192623314553805 (2015).
94 Elinav, E. *et al.* Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer* **13**, 759-771, doi:10.1038/nrc3611 (2013).
95 Tang, Y. *et al.* Jak/Stat3 signaling promotes somatic cell reprogramming by epigenetic regulation. *Stem Cells* **30**, 2645-2656, doi:10.1002/stem.1225 (2012).
96 Munoz, D. P. *et al.* Activation-induced cytidine deaminase (AID) is necessary for the epithelial-mesenchymal transition in mammary epithelial cells. *Proc. Natl. Acad. Sci. U. S. A* **110**, E2977-E2986, doi:1301021110 [pii];10.1073/pnas.1301021110 [doi] (2013).
97 Matsumoto, Y. *et al.* Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat. Med* **13**, 470-476, doi:nm1566 [pii];10.1038/nm1566 [doi] (2007).
98 Wijetunga, N. A. *et al.* A pre-neoplastic epigenetic field defect in HCV-infected liver at transcription factor binding sites and polycomb targets. *Oncogene* **36**, 2030-2044, doi:10.1038/onc.2016.340 (2017).
99 Hahn, M. A. *et al.* Methylation of polycomb target genes in intestinal cancer is mediated by inflammation. *Cancer Res* **68**, 10280-10289, doi:10.1158/0008-5472.CAN-08-1957 (2008).
100 Atashgaran, V., Wrin, J., Barry, S. C., Dasari, P. & Ingman, W. V. Dissecting the Biology of Menstrual Cycle-Associated Breast Cancer Risk. *Front Oncol* **6**, 267, doi:10.3389/fonc.2016.00267 (2016).
101 Beral, V., Doll, R., Hermon, C., Peto, R. & Reeves, G. Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. *Lancet* **371**, 303-314, doi:S0140-6736(08)60167-1 [pii];10.1016/S0140-6736(08)60167-1 [doi] (2008).
102 Hennessy, B. T., Coleman, R. L. & Markman, M. Ovarian cancer. *Lancet* **374**, 1371-1382, doi:S0140-6736(09)61338-6 [pii];10.1016/S0140-6736(09)61338-6 [doi] (2009).
103 Amant, F. *et al.* Endometrial cancer. *Lancet* **366**, 491-505, doi:S0140-6736(05)67063-8 [pii];10.1016/S0140-6736(05)67063-8 [doi] (2005).
104 Pauklin, S., Sernandez, I. V., Bachmann, G., Ramiro, A. R. & Petersen-Mahrt, S. K. Estrogen directly activates AID transcription and function. *J Exp Med* **206**, 99-111, doi:10.1084/jem.20080521 (2009).
105 Levine, M. E. *et al.* Menopause accelerates biological aging. *Proc. Natl. Acad. Sci. U. S. A*, doi:1604558113 [pii];10.1073/pnas.1604558113 [doi] (2016).

106 Maldonado-Carceles, A. B. *et al.* Anogenital Distance, a Biomarker of Prenatal Androgen Exposure Is Associated With Prostate Cancer Severity. *Prostate* **77**, 406-411, doi:10.1002/pros.23279 (2017).
107 Rahman, A. A. *et al.* Hand pattern indicates prostate cancer risk. *Br J Cancer* **104**, 175-177, doi:10.1038/sj.bjc.6605986 (2011).
108 Issa, J. P. Aging and epigenetic drift: a vicious cycle. *J Clin Invest* **124**, 24-29, doi:10.1172/JCI69735 (2014).
109 Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78-81, doi:10.1126/science.1260825 (2015).
110 Zhu, L. *et al.* Multi-organ Mapping of Cancer Risk. *Cell* **166**, 1132-1146 e1137, doi:10.1016/j.cell.2016.07.045 (2016).
111 Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* **109**, 10522-10527, doi:10.1073/pnas.1120658109 (2012).
112 Ahuja, N., Li, Q., Mohan, A. L., Baylin, S. B. & Issa, J. P. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer research* **58**, 5489-5494 (1998).
113 Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10604-10609, doi:10.1073/pnas.0500398102 (2005).
114 Teschendorff, A. E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440-446, doi:10.1101/gr.103606.109 (2010).
115 Zheng, S. C., Widschwendter, M. & Teschendorff, A. E. Epigenetic drift, epigenetic clocks and cancer risk. *Epigenomics* **8**, 705-719, doi:10.2217/epi-2015-0017 (2016).
116 Teschendorff, A. E. *et al.* Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med* **4**, 24, doi:gm323 [pii];10.1186/gm323 [doi] (2012).
117 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115, doi:gb-2013-14-10-r115 [pii];10.1186/gb-2013-14-10-r115 [doi] (2013).
118 Teschendorff, A. E. *et al.* DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat. Commun* **7**, 10478, doi:ncomms10478 [pii];10.1038/ncomms10478 [doi] (2016).
119 Teschendorff, A. E., Jones, A. & Widschwendter, M. Stochastic epigenetic outliers can define field defects in cancer. *BMC Bioinformatics* **17**, 178, doi:10.1186/s12859-016-1056-z (2016).
120 Baba, Y. *et al.* Epigenetic field cancerization in gastrointestinal cancers. *Cancer Lett* **375**, 360-366, doi:10.1016/j.canlet.2016.03.009 (2016).
121 Yang, B. *et al.* Methylation profiling defines an extensive field defect in histologically normal prostate tissues associated with prostate cancer. *Neoplasia* **15**, 399-408 (2013).
122 Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-389, doi:10.1126/science.1109557 (2005).
123 Gao, X., Jia, M., Zhang, Y., Breitling, L. P. & Brenner, H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin. Epigenetics* **7**, 113, doi:10.1186/s13148-015-0148-3 [doi];148 [pii] (2015).
124 Zhang, Y. *et al.* Smoking-associated DNA methylation markers predict lung cancer incidence. *Clin Epigenetics* **8**, 127, doi:10.1186/s13148-016-0292-4 (2016).
125 Bojesen, S. E., Timpson, N., Relton, C., Davey Smith, G. & Nordestgaard, B. G. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax*, doi:10.1136/thoraxjnl-2016-208789 (2017).
126 Baglietto, L. *et al.* DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. *Int J Cancer* **140**, 50-61, doi:10.1002/ijc.30431 (2017).
127 Zhang, Y. *et al.* Comparison and combination of blood DNA methylation at smoking-associated genes and at lung cancer related genes in prediction of lung cancer mortality. *Int. J Cancer*, doi:10.1002/ijc.30374 [doi] (2016).
128 Perna, L. *et al.* Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* **8**, 64, doi:10.1186/s13148-016-0228-z (2016).
129 Hitchins, M. P. *et al.* Inheritance of a cancer-associated MLH1 germ-line epimutation. *N. Engl. J. Med* **356**, 697-705 (2007).

130 Suter, C. M., Martin, D. I. & Ward, R. L. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* **36**, 497-501, doi:10.1038/ng1342 (2004).
131 Cui, H. M. *et al.* Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. *Science* **299**, 1753-1755 (2003).
132 Ito, Y. *et al.* Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer. *Hum. Mol. Genet* **17**, 2633-2643, doi:ddn163 [pii];10.1093/hmg/ddn163 [doi] (2008).
133 Widschwendter, M. *et al.* Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. *PLoS. One* **3**, e2656 (2008).
134 Teschendorff, A. E. *et al.* An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS. One* **4**, e8274 (2009).
135 Anjum, S. *et al.* A BRCA1-mutation associated DNA methylation signature in blood cells predicts sporadic breast cancer incidence and survival. *Genome Med* **6**, 47, doi:10.1186/gm567 [doi];gm567 [pii] (2014).
136 Teschendorff, A. E. & Widschwendter, M. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. *Bioinformatics* **28**, 1487-1494, doi:bts170 [pii];10.1093/bioinformatics/bts170 [doi] (2012).
137 Mirabello, L., Savage, S. A., Korde, L., Gadalla, S. M. & Greene, M. H. LINE-1 methylation is inherited in familial testicular cancer kindreds. *BMC Med Genet* **11**, 77, doi:10.1186/1471-2350-11-77 (2010).
138 Koestler, D. C. *et al.* Integrative genomic analysis identifies epigenetic marks that mediate genetic risk for epithelial ovarian cancer. *BMC Med Genomics* **7**, 8, doi:10.1186/1755-8794-7-8 (2014).
139 Winham, S. J. *et al.* Genome-wide investigation of regional blood-based DNA methylation adjusted for complete blood counts implicates BNC2 in ovarian cancer. *Genet Epidemiol* **38**, 457-466, doi:10.1002/gepi.21815 (2014).
140 Luo, X. *et al.* Methylation of a panel of genes in peripheral blood leukocytes is associated with colorectal cancer. *Sci Rep* **6**, 29922, doi:10.1038/srep29922 (2016).
141 Gupta, S. *et al.* Methylation of the BRCA1 promoter in peripheral blood DNA is associated with triple-negative and medullary breast cancer. *Breast Cancer Res Treat* **148**, 615-622, doi:10.1007/s10549-014-3179-0 (2014).
142 Flanagan, J. M. *et al.* Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum. Mol. Genet* **18**, 1332-1342, doi:ddp033 [pii];10.1093/hmg/ddp033 [doi] (2009).
143 Langevin, S. M. *et al.* Peripheral blood DNA methylation profiles are indicative of head and neck squamous cell carcinoma: an epigenome-wide association study. *Epigenetics* **7**, 291-299, doi:10.4161/epi.7.3.19134 (2012).
144 Shen, J. *et al.* Global methylation of blood leukocyte DNA and risk of melanoma. *Int J Cancer* **140**, 1503-1509, doi:10.1002/ijc.30577 (2017).
145 Pergoli, L. *et al.* Blood DNA methylation, nevi number, and the risk of melanoma. *Melanoma Res* **24**, 480-487, doi:10.1097/CMR.0000000000000112 (2014).
146 Liao, L. M. *et al.* LINE-1 methylation levels in leukocyte DNA and risk of renal cell cancer. *PLoS One* **6**, e27361, doi:10.1371/journal.pone.0027361 (2011).
147 Severi, G. *et al.* Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. *Breast Cancer Res Treat* **148**, 665-673, doi:10.1007/s10549-014-3209-y (2014).
148 van Veldhoven, K. *et al.* Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. *Clin Epigenetics* **7**, 67, doi:10.1186/s13148-015-0104-2 (2015).
149 Xu, Z. *et al.* Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *J Natl. Cancer Inst* **105**, 694-700, doi:djt045 [pii];10.1093/jnci/djt045 [doi] (2013).
150 Brennan, K. *et al.* Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res*, doi:0008-5472.CAN-11-3157 [pii];10.1158/0008-5472.CAN-11-3157 [doi] (2012).
151 Langevin, S. M. *et al.* Leukocyte-adjusted epigenome-wide association studies of blood from solid tumor patients. *Epigenetics : official journal of the DNA Methylation Society* **9** (2014).
152 Marsit, C. J. *et al.* DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. *J Clin Oncol* **29**, 1133-1139, doi:10.1200/JCO.2010.31.3577 (2011).

153 Wu, H. C. *et al.* Blood DNA methylation markers in prospectively identified hepatocellular carcinoma cases and controls from Taiwan. *World J Hepatol* **8**, 301-306, doi:10.4254/wjh.v8.i5.301 (2016).
154 Wu, H. C. *et al.* Global DNA methylation levels in white blood cells as a biomarker for hepatocellular carcinoma risk: a nested case-control study. *Carcinogenesis* **33**, 1340-1345, doi:10.1093/carcin/bgs160 (2012).
155 Noreen, F. *et al.* Modulation of age- and cancer-associated DNA methylation change in the healthy colon by aspirin and lifestyle. *J Natl. Cancer Inst* **106**, doi:dju161 [pii];10.1093/jnci/dju161 [doi] (2014).
156 Fasanelli, F. *et al.* Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat Commun* **6**, 10192, doi:10.1038/ncomms10192 (2015).
157 Florath, I., Butterbach, K., Muller, H., Bewerunge-Hudler, M. & Brenner, H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* **23**, 1186-1201, doi:10.1093/hmg/ddt531 (2014).
158 Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-481, doi:Doi 10.1038/Nature12433 (2013).
159 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
160 Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome biology* **15**, R31, doi:10.1186/gb-2014-15-2-r31 (2014).
161 Koestler, D. C. *et al.* Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. *Cancer Epidemiol Biomarkers Prev* **21**, 1293-1302, doi:10.1158/1055-9965.EPI-12-0361 (2012).
162 Teschendorff, A. E. *et al.* An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One* **4**, e8274, doi:10.1371/journal.pone.0008274 (2009).
163 Stunnenberg, H. G., International Human Epigenome, C. & Hirst, M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1145-1149, doi:10.1016/j.cell.2016.11.007 (2016).
164 Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nature biotechnology* **30**, 224-226, doi:10.1038/nbt.2153 (2012).
165 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).
166 Teschendorff, A. E., Breeze, C. E., Zheng, S. C. & Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC bioinformatics* **18**, 105, doi:10.1186/s12859-017-1511-5 (2017).
167 Lehmann-Werman, R. *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A* **113**, E1826-1834, doi:10.1073/pnas.1519286113 (2016).
168 Kang, S. *et al.* CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol* **18**, 53, doi:10.1186/s13059-017-1191-5 (2017).
169 Zhang, Y. *et al.* DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat Commun* **8**, 14617, doi:10.1038/ncomms14617 (2017).
170 Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature reviews. Genetics* **11**, 733-739, doi:10.1038/nrg2825 (2010).
171 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).
172 Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics* **3**, 1724-1735, doi:10.1371/journal.pgen.0030161 (2007).
173 Teschendorff, A. E., Zhuang, J. & Widschwendter, M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics* **27**, 1496-1505, doi:10.1093/bioinformatics/btr171 (2011).
174 Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **30**, 1431-1439, doi:10.1093/bioinformatics/btu029 (2014).
175 Teschendorff, A. E. *et al.* DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nature communications* **7**, 10478, doi:10.1038/ncomms10478 (2016).

176 van Dongen, J. *et al.* Epigenetic variation in monozygotic twins: a genome-wide analysis of DNA methylation in buccal cells. *Genes* **5**, 347-365, doi:10.3390/genes5020347 (2014).
177 Slieker, R. C. *et al.* Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome biology* **17**, 191, doi:10.1186/s13059-016-1053-6 (2016).
178 Hansen, K. D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nature genetics* **43**, 768-U777, doi:Doi 10.1038/Ng.865 (2011).
179 Teschendorff, A. E. & Widschwendter, M. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. *Bioinformatics* **28**, 1487-1494, doi:10.1093/bioinformatics/bts170 (2012).
180 Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet* **12**, 529-541 (2011).
181 Krzysztofek, M. *Post-reform personal data protection in the European Union : general data protection regulation (EU) 2016/679*. (Kluwer Law International B. V., 2017).
182 Pashayan, N., Reisel, D. & Widschwendter, M. Integration of genetic and epigenetic markers for risk stratification: opportunities and challenges. *Per Med* **13**, 93-95, doi:10.2217/pme.15.53 (2016).
183 Garcia-Closas, M., Gunsoy, N. B. & Chatterjee, N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst* **106**, doi:10.1093/jnci/dju305 (2014).
184 Bunnik, E. M., Janssens, A. C. & Schermer, M. H. A tiered-layered-staged model for informed consent in personal genome testing. *Eur. J Hum. Genet* **21**, 596-601, doi:ejhg2012237 [pii];10.1038/ejhg.2012.237 [doi] (2013).
185 Ploug, T., Holm, S. & Brodersen, J. To nudge or not to nudge: cancer screening programmes and the limits of libertarian paternalism. *J Epidemiol Community Health* **66**, 1193-1196, doi:10.1136/jech-2012-201194 (2012).
186 Rothstein, M. A., Cai, Y. & Marchant, G. E. The ghost in our genes: legal and ethical implications of epigenetics. *Health Matrix Clevel* **19**, 1-62 (2009).
187 McDowell, M., Rebitschek, F., Gigerenzer, G. & Wegwarth, O. A simple tool for communicating the benefits and harms of health interventions: a guide for creating a fact box. *Medical Decision Making Policy & Practice* **1:2381468316665365** (2016).
188 Steckelberg, A., Berger, B., Köpke, S., Heesen, C. & Mühlhauser, I. Kriterien für evidenzbasierte Patienteninformationen [Criteria for evidence-based information for patients]. *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen* **99**, 343–351 (2005).
189 Wegwarth, O., Schwartz, L. M., Woloshin, S., Gaissmaier, W. & Gigerenzer, G. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the U.S. *Annals of Internal Medicine*, 340–349 (2012).
190 Wegwarth, O. & Gigerenzer, G. "There is nothing to worry about": Gynecologists' counseling on mammography. *Patient Education and Counseling* **84**, 251-256, doi:10.1016/j.pec.2010.07.025 (2011).
191 Wegwarth, O., Gaissmaier, W. & Gigerenzer, G. Deceiving numbers: survival rates and their impact on doctors' risk communication. *Medical Decision Making* **31**, 386–394, doi:10.1177/0272989X10391469 (2011).
192 Prinz, R., Feufel, M. A., Gigerenzer, G. & Wegwarth, O. What counselors tell low-risk clients about HIV test performance. *Current HIV Research* **13**, 369–380 (2015).
193 Gold, M. R., Siegel, J. E., Russell, L. B. & Weinstein, M. C. *Cost-Effectiveness in Health and Medicine*. (Oxford University Press Inc., 1996).
194 Hunink, M. & Glasziou, P. *Decision making in health and medicine. Integrating evidence and values.*, (Cambridge University Press, 2001).
195 Weinstein, M. C. High-priced technology can be good value for money. *Ann Intern Med* **130**, 857-858 (1999).
196 Siebert, U. When should decision-analytic modeling be used in the economic evaluation of health care? [Editorial]. *European Journal of Health Economics* **4**, 143-150 (2003).
197 Siebert, U. *et al.* State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Value in Health* **15**, 812-820 (2012).
198 Trikalinos, T. A., Siebert, U. & Lau, J. Decision-Analytic Modeling to Evaluate Benefits and Harms of Medical Tests: Uses and Limitations. *Med Decis Making* **29**, E22-29 (2009).
199 Caro, J. J., Briggs, A. H., Siebert, U., Kuntz, K. M. & Force, o. b. o. t. I.-S. M. G. R. P. T. Modeling Good Research Practices - Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force -1. *Medical Decision Making* **32**, 667-677 (2012).

200 Hakama, M., Malila, N. & Dillner, J. Randomised health services studies. *Int J Cancer* **131**, 2898-2902, doi:10.1002/ijc.27561 (2012).
201 Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**, 78-85, doi:10.1056/NEJM200007133430201 (2000).
202 Mucci, L. A. *et al.* Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **315**, 68-76, doi:10.1001/jama.2015.17703 (2016).
203 Magnusson, P. K., Lichtenstein, P. & Gyllensten, U. B. Heritability of cervical tumours. *Int J Cancer* **88**, 698-701 (2000).
204 Chen, D. *et al.* Analysis of the genetic architecture of susceptibility to cervical cancer indicates that common SNPs explain a large proportion of the heritability. *Carcinogenesis* **36**, 992-998, doi:10.1093/carcin/bgv083 (2015).
205 Al-Tassan, N. A. *et al.* A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Scientific reports* **5**, 10442, doi:10.1038/srep10442 (2015).
206 Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet* **6**, e1000864, doi:10.1371/journal.pgen.1000864 (2010).
207 Tyrer, J., Duffy, S. W. & Cuzick, J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* **23**, 1111-1130, doi:10.1002/sim.1668 (2004).
208 Pfeiffer, R. M. *et al.* Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. *PLoS Med* **10**, e1001492, doi:10.1371/journal.pmed.1001492
PMEDICINE-D-13-00139 [pii] (2013).
209 Roobol, M. J. *et al.* A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. *Eur Urol* **57**, 79-85, doi:S0302-2838(09)00893-8 [pii]
10.1016/j.eururo.2009.08.025 (2010).
210 Tammemagi, M. C. *et al.* Selection criteria for lung-cancer screening. *N Engl J Med* **368**, 728-736, doi:10.1056/NEJMoa1211776 (2013).
211 Thrift, A. P. *et al.* A clinical risk prediction model for Barrett esophagus. *Cancer Prev Res (Phila)* **5**, 1115-1123, doi:1940-6207.CAPR-12-0010 [pii]
10.1158/1940-6207.CAPR-12-0010 (2012).
212 Freedman, A. N. *et al.* Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol* **27**, 686-693, doi:JCO.2008.17.4797 [pii]
10.1200/JCO.2008.17.4797 (2009).
213 Zhang, Y. *et al.* F2RL3 methylation, lung cancer incidence and mortality. *Int J Cancer* **137**, 1739-1748, doi:10.1002/ijc.29537 (2015).
214 Chowdhury, S. *et al.* Incorporating genomics into breast and prostate cancer screening: assessing the implications. *Genet. Med* **15**, 423-432, doi:gim2012167 [pii];10.1038/gim.2012.167 [doi] (2013).