WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.453
Volume 13, Issue 9, 528-543. Review Article ISSN 2277–7105

DNA METHYLATION RISK SCORE AS A PREDICTIVE PARAMETER FOR DISEASE VULNERABILITY

***[1]Mayuri N. Jagtap, [2]Jayesh R. Baviskar and [3]Avinash B. Darekar**

[1,2,3]Krantiveer Vasantrao Narayanrao Naik Shikshan Prasarak Sanstha's, Institute of Pharmaceutical Education & Research, Canada Corner, Nashik - (422002) Maharashtra, India.

Article Received on
06 March 2024,

Revised on 27 March 2024,
Accepted on 17 April 2024

DOI: 10.20959/wjpr20249-32106

***Corresponding Author**
Mayuri N. Jagtap
Krantiveer Vasantrao Narayanrao Naik Shikshan Prasarak Sanstha's, Institute of Pharmaceutical Education & Research, Canada Corner, Nashik - (422002) Maharashtra, India.

ABSTRACT

Methylation risk scores (MRS) have emerged as promising tools in the realm of personalized medicine, offering insights into disease susceptibility, prognosis, and treatment purpose. DNA methylation, an essential epigenetic modification, plays a vital role in regulating gene expression, genomic stability, and cellular identity. This review provides a comprehensive overview of the current understanding of DNA methylation dynamics, encompassing its mechanisms, regulation, and functional implications. In this we come to know about the DNA Methylation and Methylation risk score (MRS), highlighting their methodologies, applications and methylation associated with multiple diseases. Genetic risk scores (GRS) are extensively used for risk prediction as well as in association and interaction studies. Recently, interest has been increasing in transferring GRS approaches to DNA methylation data (methylation risk scores, MRS), which can be used 1) as biomarkers for environmental exposures, 2) in association analyses in which single CpG sites do not achieve significance. Finally, we discuss the constructing methylation-based risk scores, and examine their performance in diverse disease contexts, such as gestational diabetes and breast cancer. This review serves as a comprehensive guide for researchers and clinicians alike, illuminating the current landscape of MRS research and charting future directions towards their clinical implementation and impact on patient care.

KEYWORDS: Methylation risk scores (MRS), Epigenetic modification, gestational diabetes, breast cancer, CpG sites.

INTRODUCTION TO METHYLATION

DNA methylation, a significant epigenetic mechanism, can alter the function of a DNA segment without making alterations in its sequence through the addition of a methyl group to cytosine residues.[1] DNA methylation is a reversible alteration to DNA that entails the covalent linking of methyl groups (CH3) to the fifth carbon position of cytosine by enzymes known as DNA methyltransferases.[2] Certain CpG sites exhibit a vigorous correlation between DNA methylation levels and age, while at other sites, alterations in DNA methylation occur in response to environmental factors. Among these influences, the impact of smoking on DNA methylation is extensively reported.[3,4,5,6] DNAm is primarily detected through the transformation of unmethylated cytosines with sodium bisulphite to uracil, allowing methylated and unmethylated cytosines to be distinguished using array-based or sequencing-based technologies. The assessment of by array technology is cheaper and more high-throughout than by sequencing.[7] The methylation-wide association study (MWAS) data sets are many-fold minute than genome-wide association study (GWAS) data sets that rely firstly Single nuclear polymorphism (SNP) array data.[8] Epigenome-wide association studies (EWAS) have proven effective in pinpointing numerous cytosine guanine dinucleotide sites (CpGs) linked to diverse diseases. These findings hold promise for applications such as disease detection and prognosis, the recognition of drug targets, and the estimation of drug reactions.[9,10,11,12,13,14]

While the DNA sequence is constant cell types throughout lifetime (other than somatic mutations), DNAm be distinguishable between cell types (within an individual) and between people (within a cell type) for multiple reasons.[15] Biomarkers of DNA methylation influenced by environmental risk factors could pave the way for novel methods to swiftly assess the effectiveness of lifestyle interventions, benefaction valuable feedback to both clinicians and patients.[16] Methylation risk scores (MRS) are specify as aggregated values obtained from the beta values of selected CpG sites within an individual's methylation profile. They serve multiple purposes, incorporate acting as indicators of environmental exposures like smoking.[17,18] The predominant polygenic method includes enrol weighted genetic risk scores (GRS), also referred to as polygenic risk scores (PRS). These scores are calculated as weighted totals of risk alleles from a predetermined group of single nucleotide polymorphisms (SNPs).[19] In a recent examination, findings revealed that a risk model amalgamating lung cancer polygenic risk scores (PRS), a smoking-related methylation score (MS), and environmental variables like pack years, signify enhanced predictive capability for

lung cancer compared to models comprises separate scores. The joint model yielded a higher accuracy (AUC joint=0.654) than those utilizing individual scores (AUCPRS=0.571, AUCMS=0.628), with the considerable increase in AUC primarily attributable to the methylation score.[20] In In the context of Pancreatic Cancer (PC), researchers have reconnoitred DNA methylation as a procedure linked to tumorigenesis, influencing both genome stability and the regulation of gene transcription.[21] Especially for complex diseases, GRS are statistically strong to test for marginal genetic effects and gene–environment interaction result, as well as to predict individual trait values or risks of disease.[22] Utilizing DNA methylation as a biomarker presents several advantages compared to other epigenetic modifications. These include, but are not limited to, heightened sensitivity and dynamic range, multiple altered sites within each targeted region, and a vast array of targeted regions within a disease context.[23]

METHODS

P+T Method

A commonly employed strategy in the field of genetics to address single nucleotide polymorphisms (SNPs) that reveal high linkage disequilibrium (LD) and to determine optimal p-value thresholds for accurate prediction is the pruning and threshold (P+T) method.[24]

Within the P+T procedure, the correlation square (R2) is computed for SNPs situated in close genetic proximity, and less statistically significant SNPs correlated with an R2 exceeding a specified threshold (LD pruning) are eliminated.[25] Following that, various p-value thresholds are evaluate to optimize the predictive accuracy of the resulting PRS (p-value threshold optimization)[26,27] In theory, while the P + T method could be employed to obtain MRS, there exists no established protocol for conducting pruning in DNAm data, and the efficacy of such MRS across diverse ancestral backgrounds is yet to be determined. Here, we suggest to use the Co-Methylation with genomic CpG Background (CoMeBack) approach, which employs a sliding window technique to gauge DNA co-methylation, in order to address connection among DNAm at adjacent CpG sites.[28] We carried out simulation studies using data from an adult cohort consist of three distinct ancestral groups (Indian, White, and Black, n = 1,199) to assess the predictive capabilities of the P+T CoMeBack. MRS and its variations across diverse ancestral populations. Afterwards, we implemented the P+T CoMeBack methodology on DNAm data from the Drakenstein Child Health Study (n = 260).[29] In the P+T CoMeBack

method, MRS computation involves informed co-methylation pruning (P) utilizing CoMeBack and subsequent P-value thresholding (T). firstly, summary statistics derived from an EWAS (typically encompassing effect size, standard error, and P-value for each CpG site) must be evaluate in an independent dataset (training dataset) to mitigate overfitting. These statistics are then use to generate MRS within a testing dataset (samples employed for MRS performance evaluation). Within our P+T CoMeBack method, co-methylation pruning is carried out by applying CoMeBack to DNAm data within the testing dataset or a reference panel.[30]

Microarray – based DNA methylation profiling

In a technique called bisulfite methylation profiling (BiMP), bisulfite-treated DNA undertake whole genome amplification (WGA) employing random tetranucleotide primers, enzymatic fragmentation, and subsequent microarray hybridization.[31] The microarray design involves using differentially labelled oligonucleotide pairs that are complementary to the unaltered methylated sequence. Consequently, methylation is identified as a signal, and discrepancies arising from the conversion of unmethylated cytosines do not build a signal. However, this technique often yields a low hybridization signal overall and may not be suitable for regions with minimal methylation. The Infinium approach follows a similar sample preparation protocol, which includes bisulfite modification of genomic DNA followed by WGA.[32,33] The DNA is subsequently hybridized onto Bead Chip microarrays, which are equipped with oligonucleotide pairs targeting certain CpG sites. One pair is complementary to the unchanged, methylated sequence, while the other targets the converted unmethylated sequence. Following hybridization, a polymerase chain reaction (PCR) reaction is conducted utilizing fluorescently labelled universal PCR primers, enabling identification of methylation levels by comparing the fluorescence emitted by each dye. While most microarray platforms feature a standard array of probes covering a library of CGIs, certain companies offer custom microarrays to accommodate diverse experimental designs and promote methylation analysis of CGIs and/or organisms not covered by standard arrays. Looking ahead to the era of personalized medicine custom microarrays will prove invaluable for analysing specific individual methylation signatures.

- **Microarray expression profiling**

Genome-wide methylation profiling of samples showing diseased and normal state in search for biomarkers can be costly and time consuming. Therefore, some investigators prefer to

narrow down the search using an expression-array through treatment with demethylating agents such as 5-aza-2′-deoxycytidine[34,35] This method facilitates the identification of genes exhibiting signs of methylation-dependent gene regulation in a diseased state, aiding in the comprehension of disease pathobiology and progression. It also identifies potential biomarkers, namely genes reactivated after treatment. However, this approach is vulnerable to yielding false outcomes and is not deemed a reliable indicator of DNA methylation. This is due to the fact that treatment with demethylating drugs alters the expression of numerous genes that might not be associated with the disease state and can induce the expression of additional secondary targets. Hence, methylation profiles of candidate biomarkers identified through this approach undergo moreover validation using alternative strategies.

APPLICATIONS

1. DNAm data could be used to present patients with objectively quantified outcomes, potentially enhancing the likelihood of successful behaviour quantification through positive reinforcement.
2. Additionally, Methylation profile scores (MPS) could aid in clinically identifying treatment failure, particularly when DNAm data are assessed before and after therapeutic intervention.
3. Although longitudinal datasets spanning several time points are uncommon, recent research has revealed distinct DNAm signatures linked to therapeutic intervention within 4–12 weeks of commencement.[36,37]
4. Phenotype recognition finds applications in forensic science, particularly in cases where a biological sample is available but the associated individual remains unidentified. Unlike DNA profiling, which serves as evidence, MPS can complement suspect profiling in the investigative process (as DNA profiling becomes sufficient once a person is identified). In criminal investigations, demographic traits are important for offender identification. While highly heritable traits like height can be anticipate from genetic data, less heritable traits such as weight or body mass index (BMI) might be more prefectly predicted by MPS or a combination of MPS and PGS.[38,39]
5. Smoking constitute a crucial risk factor in epidemiological analyses. In cases where smoking status is not documented, accurate prediction of smoking can be gained from DNAm data, with an AUC statistic of 0.98 (where AUC represents the probability that a smoker ranks higher than a non-smoker on the MPS).[40,41,42,43,44]

6. Prognosticating, unrecorded phenotypes holds significance in association studies, wherein incorporating confounding variables can aid in mitigating the false-positive rate.[45,46]
7. Body weight loss has also been linked to differences in DNA.[47]

METHYLATION RISK SCORES

Identifying trait prediction using DNAm data necessitates a minimum of two distinct datasets containing measurements of both genome wide DNA m and the trait under investigation. The MWAS "discovery" sample is employed to pinpoint DNAm probes associate with the trait, producing a roster of probes and their respective weights. These findings can be used to formulate a MPS for each individual within the separate "target" sample. The effectiveness of trait prediction is asses by examining the correlation between the MPS and the directly measured trait in the target sample. For both MPS and PGS, there's no essential for the scores to signify functional or causal mechanisms; rather, it's essential that an association is observed in independent datasets. For both PGS and MPS, the precision of prediction may be constrained, yet the information conveyed by the scores could still prove valuable. Enhancing the accuracy of prediction can be obtained by integrating PGS, MPS, and other recognized risk factors. Even with such a merged predictor, the likelihood of high prediction error for a specific individual remains probable, thereby suggesting utility primarily at the level of stratification, where a high-risk group will be enriched with individuals who subsequently develop the disease. We employ the term "trait prediction" for MPS cautiously, as it's crucial to highlight a fundamental difference between MPS and PGS. While PGS can be computed at birth and remains constant throughout one's lifetime (unless the SNPs and their weights used in constructing the PGS are updated), MPS is subject to variability since measured DNAm levels can fluctuate. Consequently, any trait-specific MPS for an individual has the capability to change over their lifetime. Therefore, considerably more deliberation is necessary, in comparison to PGS applications, regarding the timing of biological sample collection (e.g., blood) in the discovery, target, and final application samples to regulate whether the developed MPS can genuinely be regarded as a predictor of future events.[48,49]

POLYGENIC RISK SCORE – PRS

The PRS consist of a collection of independent risk variants associated with a disorder, relying on current evidence gleaned from the most substantial or most informative genome-wide association studies. For every individual, the count of risk alleles carried at each variant (0, 1, or 2) is totalled, with weighting found on its effect size (i.e., log (OR) for binary traits

or beta coefficient for continuous traits). This yields a singular score representing each individual's genetic predisposition for a disease or a continuous trait.[50]

Various techniques subsist for computing polygenic risk scores. Among these are 'clumping/pruning and thresholding' procedure, which involve identifying a condensed set of genetic variants through pruning situate on linkage disequilibrium and taking into account evidence of association with the studied trait (clumping). Polygenic risk scores are subsequently calculated by summing over all SNPs meeting particular p-value thresholds or sets of thresholds, as exemplified in PRSice (11) and PLINK (12).[51,52] There is growing recognition that genetic risk scores not only encapsulate genetic information but also comprises environmental factors. This is evidenced in family-based studies demonstrating that contributions from the PRS computed utilizing non-transmitted alleles of parents also impact offspring phenotypes. This phenomenon of 'genetic nurture', recommend for instance, that educational attainment is influenced not only by the genetics of the offspring but also by the non-transmitted genetics of the parents, which may shape the family environment.[53]

DNA METHYLATION ASSOCIATED WITH VARIOUS DISEASES

Gestational Diabetes

Type 2 Diabetes Mellitus (T2DM) affects approximately 10.5% of the global population aged 20 to 79 years. T2DM poses a significant threat to both individual health and healthcare systems due to its numerous obstacles and the substantial healthcare expenses it incurs. Additionally, it ranks among the top ten leading causes of mortality worldwide.[54]

Throughout pregnancy, insulin resistance escalates to ensure a sufficient glucose supply to the offspring. However, if pancreatic beta cells fail to compensate by secreting sufficient insulin, it can lead to hyperglycaemia and Gestational Diabetes Mellitus (GDM), distinguish by the onset of hyperglycaemia during pregnancy.[55] Certain epigenome-wide association studies (EWAS) direct the involvement of epigenetic mechanisms in the development of T2DM (5–7). Genetic risk scores (GRS) have seen developing utilization in evaluating disease susceptibility, including for T2DM.[56,57]

-Methods

A subset of 480 women was selected from the STORK Groruddalen (STORK G) pregnancy cohort, which comprised 823 healthy women of various ethnic backgrounds (European, South Asian, African, Middle Eastern, and South American) attending three different Child

Health Clinics in the Groruddalen area of Oslo, Norway, between 2008 and 2010 (13). Ethnic origin was recognized based on either the individual's country of birth or their mother's country of birth if the latter was born outside of Europe. The EPIPREG subgroup included 312 European subjects (EPIPREG_EU), of which 73 were diagnosed with GDM, and 168 South Asians (EPIPREG_SA), of which 68 were diagnosed with GDM. European and South Asian ancestry was examine using genetic principal components (14). Fasting blood samples were collected, and a 75g oral glucose tolerance test (OGTT) was offered to all women (universal testing) around week 28 ± 2 of pregnancy. For this study, GDM was classified according to slightly modified International Association of the Diabetes and Pregnancy (IADPSG) criteria (fasting glucose ≥ 5.1 mmol/l and/or 2-hour glucose ≥ 8.5 mmol/l, as 1-hour glucose values were not available).[58,59,60]

-Methylation risk score

To generate the weighted DNA methylation risk score (MRS) for T2DM in our sample, we male use of summary data from CpG sites identified in an EWAS of T2DM across five prospective European cohorts ($n = 5,859$) (7, 18–21). The regression coefficients for the EWAS of T2DM were derived from β-values. Subsequently, we calculated the MRS for each of the identified CpG sites obtainable in our two cohorts. Due to QC filters applied to CpGs in each cohort, only 42 out of 72 CpG sites could be included in the MRS. The MRS was constructed by multiplying the regression coefficient from the EWAS of T2DM with individual β-values in EPIPREG and EPIDG (EPIPREG-EU, EPIPREG-SA, and EPIDG). Subsequently, we clumped the individual scores for each of the 42 CpG sites to obtain the MRS.[61,62,63,64]

Breast Cancer

DNA Methylation score forecast the breast cancer risk. Breast cancer (BC) appear in the epithelial tissues of the breast gland and stands as the most prevalent form of cancer in women. It obvious with symptoms including the presence of a breast lump, deformity of the breast, skin dimpling, and nipple discharge.[65] Primary risk factors for BC comprise a family history of the disease, vulnerability to ionizing radiation, obesity, sedentary lifestyle, excessive alcohol and tobacco use, early onset of menstruation, and delayed childbirth or infertility.[66] DNA methylation can enable the development and progression of cancer by triggering aberrant gene expression and genomic instability. perception of methylation is crucial for identifying individuals susceptible to tumor development.[67,68] Incorporating a

polygenic risk score comprises numerous single nucleotide polymorphisms can enhance the predictive capability of the Breast Cancer Surveillance Consortium risk model. This aids in choosing the most acceptable treatment modalities for patients with breast cancer.[69,70,71]

- **Data source**

From the TCGA database (https://cancergenome.nih.gov/) accessed on September 10, 2019, methylation data for breast cancer (BC) was extracted, used the Illumina Infinium Human Methylation 450 BeadChip platform. Following alignment with accessible clinical information, a total of 872 samples were included in this study as the training set, comprising 776 BC samples with survival prognosis details and 96 normal samples. Additionally, methylation data from the GSE37754 dataset, also utilizing the Illumina Infinium Human Methylation 450 BeadChip platform, was acquire from the Gene Expression Omnibus (GEO) database. Within GSE37754, there were 62 BC samples with survival prognosis information and 10 normal samples, creating the validation set. Utilizing the clinical prognosis information of the 776 BC samples in the training set, differentially methylated regions (DMRs) significantly correlated with prognosis were recognized through univariate Cox regression analysis using the R package survival.[72] Employing the Cox-Proportional Hazards (Cox-PH) model in the R package penalized19 (version 0.9.50), we selected the optimal fusion of independent prognosis-associated DMRs based on their methylation levels. The parameter "lambda" in the model was fine-tuned through 1000 cycles of calculation utilizing cross-validation likelihood (cvl).[73,74] Based on the prognostic coefficients and methylation levels of the DMRs in the optimal combination, we established the backing risk score system: The risk score (RS) for each sample in the training set was enumerate, followed by categorization of the samples into high-risk and low-risk groups based on the median RS value. Subsequently, the Kaplan-Meier (KM) method in the survival package was utilized.[75]

We extracted the methylation levels of the DMRs significantly correspond with prognosis from the training set. Subsequently, we calculated the Pearson correlation coefficients (PCCs) of the methylation levels between the lncRNAs and mRNAs among, these DMRs operate the cor.test function.[76]

REFERENCES

1. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science, 1985 Apr 12; 228(4696): 187-90.
2. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews genetics, 2012 Jul; 13(7): 484-92.
3. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: In the right place at the right time. Science, 2018 Sep 28; 361(6409): 1336-40.
4. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, Morris SW, Bermingham ML, Campbell A, Murray AD, Whalley HC. Epigenetic prediction of complex traits and death. Genome biology, 2018 Dec; 19: 1-1.
5. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PloS one, 2013 May 17; 8(5): e63812.
6. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, Davey Smith G, Hughes AD, Chaturvedi N, Relton CL. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clinical epigenetics, 2014 Dec; 6: 10.
7. Seiler Vellame D, Castanho I, Dahir A, Mill J, Hannon E. Characterizing the properties of bisulfite sequencing data: maximizing power and sensitivity to identify between-group differences in DNA methylation. BMC genomics, 2021 Dec; 22: 1-6.
8. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, Carnero-Montoro E, Lawson DJ, Burrows K, Suderman M, Bretherick AD. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nature genetics, 2021 Sep; 53(9): 1311-21.
9. Wei S, Tao J, Xu J, Chen X, Wang Z, Zhang N, Zuo L, Jia Z, Chen H, Sun H, Yan Y. Ten years of EWAS. Advanced Science, 2021 Oct; 8(20): 2100727.
10. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y, Tan S. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature, 2017 Jan 5; 541(7635): 81-6.
11. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, Morris SW, Bermingham ML, Campbell A, Murray AD, Whalley HC. Epigenetic prediction of complex traits and death. Genome biology, 2018 Dec; 19: 1-1.

12. Heiss JA, Brenner H. Epigenome-wide discovery and evaluation of leukocyte DNA methylation markers for the detection of colorectal cancer in a screening setting. Clinical epigenetics, 2017 Dec; 9: 1-9.
13. Gelato KA, Adler D, Ocker M, Haendler B. Targeting epigenetic regulators for cancer therapy: modulation of bromodomain proteins, methyltransferases, demethylases, and microRNAs. Expert opinion on therapeutic targets, 2016 Jul 2; 20(7): 783-99.
14. Krushkal J, Silvers T, Reinhold WC, Sonkin D, Vural S, Connelly J, Varma S, Meltzer PS, Kunkel M, Rapisarda A, Evans D. Epigenome-wide DNA methylation analysis of small cell lung cancer cell lines suggests potential chemotherapy targets. Clinical epigenetics, 2020 Dec; 12: 1-28.
15. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics, 2012 Dec; 13: 1-6.
16. Sarris J, Thomson R, Hargraves F, Eaton M, de Manincor M, Veronese N, Solmi M, Stubbs B, Yung AR, Firth J. Multiple lifestyle factors and depressed mood: a cross-sectional and longitudinal analysis of the UK Biobank (N= 84,860). BMC medicine, 2020 Dec; 18: 1-0.
17. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, Davey Smith G, Hughes AD, Chaturvedi N, Relton CL. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clinical epigenetics, 2014 Dec; 6: 1-0.
18. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, St Clair D, Mustard C, Breen G, Therman S, Kaprio J. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. Genome biology, 2016 Dec; 17: 1-6.
19. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. Journal of child psychology and psychiatry, 2014 Oct; 55(10): 1068-87.
20. Yu H, Raut JR, Schöttker B, Holleczek B, Zhang Y, Brenner H. Individual and joint contributions of genetic and methylation risk scores for enhancing lung cancer risk stratification: data from a population-based cohort in Germany. Clinical epigenetics, 2020 Dec; 12: 1-1.
21. Brancaccio M, Natale F, Falco G, Angrisano T. Cell-free DNA methylation: the new frontiers of pancreatic cancer biomarkers' discovery. Genes, 2019 Dec 23; 11(1): 14.

22. Hüls A, Ickstadt K, Schikowski T, Krämer U. Detection of gene-environment interactions in the presence of linkage disequilibrium and noise by using genetic risk scores with internal weights from elastic net regression. BMC genetics, 2017 Dec; 18: 1-4.
23. Brancaccio M, Natale F, Falco G, Angrisano T. Cell-free DNA methylation: the new frontiers of pancreatic cancer biomarkers' discovery. Genes, 2019 Dec 23; 11(1): 14.
24. Wu J, Pfeiffer RM, Gail MH. Strategies for developing prediction models from genome-wide association studies. Genetic epidemiology, 2013 Dec; 37(8): 768-77.
25. Privé F, Vilhjálmsson BJ, Aschard H, Blum MG. Making the most of clumping and thresholding for polygenic scores. The American journal of human genetics, 2019 Dec 5; 105(6): 1213-21.
26. Privé F, Vilhjálmsson BJ, Aschard H, Blum MG. Making the most of clumping and thresholding for polygenic scores. The American journal of human genetics, 2019 Dec 5; 105(6): 1213-21.
27. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience, 2019 Jul; 8(7): giz082.
28. Gatev E, Gladish N, Mostafavi S, Kobor MS. CoMeBack: DNA methylation array data analysis for co-methylated regions. Bioinformatics, 2020 May 1; 36(9): 2675-83.
29. Vanker A, Barnett W, Workman L, Nduru PM, Sly PD, Gie RP, Zar HJ. Early-life exposure to indoor air pollution or tobacco smoke and lower respiratory tract illness and wheezing in African infants: a longitudinal birth cohort study. The Lancet Planetary Health, 2017 Nov 1; 1(8): e328-36.
30. Gatev E, Gladish N, Mostafavi S, Kobor MS. CoMeBack: DNA methylation array data analysis for co-methylated regions. Bioinformatics, 2020 May 1; 36(9): 2675-83.
31. Reinders J, Vivier CD, Theiler G, Chollet D, Descombes P, Paszkowski J. Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. Genome research, 2008 Mar 1; 18(3): 469-76.
32. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E, Goldmann T. High-throughput DNA methylation profiling using universal bead arrays. Genome research, 2006 Mar 1; 16(3): 383-93.
33. Weisenberger DJ, Van Den Berg D, Pan F, Berman BP, Laird PW. Comprehensive DNA methylation analysis on the Illumina Infinium assay platform. Illumina, San Diego. 2008.
34. Suzuki H, Gabrielson E, Chen W, Anbazhagan R, Van Engeland M, Weijenberg MP, Herman JG, Baylin SB. A genomic screen for genes upregulated by demethylation and

histone deacetylase inhibition in human colorectal cancer. Nature genetics, 2002 Jun; 31(2): 141-9.

35. Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. Cell, 1980 May 1; 20(1): 85-93.
36. Nair N, Plant D, Verstappen SM, Isaacs JD, Morgan AW, Hyrich KL, Barton A, Wilson AG, MATURA investigators. Differential DNA methylation correlates with response to methotrexate in rheumatoid arthritis. Rheumatology, 2020 Jun 1; 59(6): 1364-71.
37. Julia A, Gomez A, Lopez-Lasanta M, Blanco F, Erra A, Fernandez-Nebro A, Mas AJ, Perez-García C, Vivar ML, Sanchez-Fernandez S, Alperi-Lopez M. Longitudinal analysis of blood DNA methylation identifies mechanisms of response to tumor necrosis factor inhibitor therapy in rheumatoid arthritis. EBio Medicine, 2022 Jun 1; 80.
38. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, Harris SE, Liewald D, Henders AK, Mendelson MM, Liu C. Improving phenotypic prediction by combining genetic and epigenetic associations. The American Journal of Human Genetics, 2015 Jul 2; 97(1): 75-85.
39. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, He J, Fan D, Fowdar J, Kennedy M, Pitcher T. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome medicine, 2019 Dec; 11: 1-1.
40. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, Morris SW, Bermingham ML, Campbell A, Murray AD, Whalley HC. Epigenetic prediction of complex traits and death. Genome biology, 2018 Dec; 19: 1-1.
41. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, Davey Smith G, Hughes AD, Chaturvedi N, Relton CL. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clinical epigenetics, 2014 Dec; 6: 1-0.
42. Trejo Banos D, McCartney DL, Patxot M, Anchieri L, Battram T, Christiansen C, Costeira R, Walker RM, Morris SW, Campbell A, Zhang Q. Bayesian reassessment of the epigenetic architecture of complex traits. Nature communications, 2020 Jun 8; 11(1): 2865.
43. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, Flanagan JM, Vineis P. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology, 2013 Sep 1; 24(5): 712-6.
44. Zhang Y, Florath I, Saum KU, Brenner H. Self-reported smoking, serum cotinine, and blood DNA methylation. Environmental research, 2016 Apr 1; 146: 395-403.

45. Zhang F, Chen W, Zhu Z, Zhang Q, Nabais MF, Qi T, Deary IJ, Wray NR, Visscher PM, McRae AF, Yang J. OSCA: a tool for omic-data-based complex trait analysis. Genome biology, 2019 Dec; 20: 1-3.
46. Zhang F, Chen W, Zhu Z, Zhang Q, Nabais MF, Qi T, Deary IJ, Wray NR, Visscher PM, McRae AF, Yang J. OSCA: a tool for omic-data-based complex trait analysis. Genome biology, 2019 Dec; 20: 1-3.
47. Keller M, Yaskolka Meir A, Bernhart SH, Gepner Y, Shelef I, Schwarzfuchs D, Tsaban G, Zelicha H, Hopp L, Müller L, Rohde K. DNA methylation signature in blood mirrors successful weight-loss during lifestyle interventions: the CENTRAL trial. Genome medicine, 2020 Dec; 12: 1-8.
48. Bernabeu E, McCartney DL, Gadd DA, Hillary RF, Lu AT, Murphy L, Wrobel N, Campbell A, Harris SE, Liewald D, Hayward C. Refining epigenetic prediction of chronological and biological age. Genome Medicine, 2023 Feb 28; 15(1): 12.
49. Gadd DA, Hillary RF, McCartney DL, Zaghlool SB, Stevenson AJ, Cheng Y, Fawns-Ritchie C, Nangle C, Campbell A, Flaig R, Harris SE. Epigenetic scores for the circulating proteome as tools for disease prediction. Elife, 2022 Jan 13; 11: e71802.
50. Aschard H. A perspective on interaction effects in genetic association studies. Genetic epidemiology, 2016 Dec; 40(8): 678-88.
51. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience, 2019 Jul; 8(7): giz082.
52. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 2015 Dec 1; 4(1): s13742-015.
53. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, Benonisdottir S, Oddsson A, Halldorsson BV, Masson G, Gudbjartsson DF. The nature of nurture: Effects of parental genotypes. Science, 2018 Jan 26; 359(6374): 424-8.
54. Abdul Basith Khan M, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. Journal of epidemiology and global health, 2020 Mar; 10(1): 107-11.
55. Odgers-Jewell K, Ball LE, Kelly JT, Michaleff ZA, Clark J, Jones MA, Reidlinger DP. Group-based self-management education for people with type 2 diabetes mellitus. The Cochrane Database of Systematic Reviews, 2023; 2023(9).
56. Iwata M, Maeda S, Kamura Y, Takano A, Kato H, Murakami S, Higuchi K, Takahashi A, Fujita H, Hara K, Kadowaki T. Genetic risk score constructed using 14 susceptibility

alleles for type 2 diabetes is associated with the early onset of diabetes and may predict the future requirement of insulin injections among Japanese individuals. Diabetes care, 2012 Aug 1; 35(8): 1763-70.

57. Stanislawski MA, Litkowski E, Raghavan S, Harrall KK, Shaw J, Glueck DH, Lange EM, Dabelea D, Lange LA. Genetic risk score for type 2 diabetes and traits related to glucose-insulin homeostasis in youth: The exploring perinatal outcomes among children (EPOCH) study. Diabetes Care, 2021 Sep 1; 44(9): 2018-24.
58. Jenum AK, Mørkrid K, Sletner L, Vange S, Torper JL, Nakstad B, Voldner N, Rognerud-Jensen OH, Berntsen S, Mosdøl A, Skrivarhaug T. Impact of ethnicity on gestational diabetes identified with the WHO and the modified International Association of Diabetes and Pregnancy Study Groups criteria: a population-based cohort study. European journal of endocrinology, 2012 Feb; 166(2): 317-24.
59. Fragoso-Bargas N. DNA methylation and cardiometabolic health: associations with insulin resistance, folate, and physical activity in pregnancy.
60. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics, 2012 Dec; 13: 1-6.
61. Fraszczyk E, Spijkerman AM, Zhang Y, Brandmeier S, Day FR, Zhou L, Wackers P, Dollé ME, Bloks VW, Gào X, Gieger C. Epigenome-wide association study of incident type 2 diabetes: a meta-analysis of five prospective European cohorts. Diabetologia, 2022 May; 65(5): 763-76.
62. Verschuren WM, Blokstra A, Picavet HS, Smit HA. Cohort profile: the Doetinchem cohort study. International journal of epidemiology, 2008 Dec 1; 37(6): 1236-41.
63. Raum E, Rothenbacher D, Löw M, Stegmaier C, Ziegler H, Brenner H. Changes of cardiovascular risk factors and their implications in subsequent birth cohorts of older adults in Germany: a life course approach. European Journal of Preventive Cardiology, 2007 Dec 1; 14(6): 809-14.
64. Holle R, Happich M, Löwel H, Wichmann HE, null for the MONICA/KORA Study Group. KORA-a research platform for population-based health research. Das Gesundheitswesen, 2005 Aug; 67(S 01): 19-25.
65. Sun Y, Wang R. A Risk Score System Based on the Methylation Levels of 15 RNAs in Breast Cancer. Cancer Biotherapy & Radiopharmaceuticals, 2022 Oct 1; 37(8): 697-707.

66. Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. Risk factors and preventions of breast cancer. International journal of biological sciences, 2017; 13(11): 1387.
67. Saif I, Kasmi Y, Allali K, Ennaji MM. Prediction of DNA methylation in the promoter of gene suppressor tumor. Gene, 2018 Apr 20; 651: 166-73.
68. Xie W, Kagiampakis I, Pan L, Zhang YW, Murphy L, Tao Y, Kong X, Kang B, Xia L, Carvalho FL, Sen S. DNA methylation patterns separate senescence from transformation potential and indicate cancer risk. Cancer cell, 2018 Feb 12; 33(2): 309-21.
69. Shieh Y, Hu D, Ma L, Huntsman S, Gard CC, Leung JW, Tice JA, Vachon CM, Cummings SR, Kerlikowske K, Ziv E. Breast cancer risk prediction using a clinical risk model and polygenic risk score. Breast cancer research and treatment, 2016 Oct; 159: 513-25.
70. Li H, Feng B, Miron A, Chen X, Beesley J, Bimeh E, Barrowdale D, John EM, Daly MB, Andrulis IL, Buys SS. Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. Genetics in Medicine, 2017 Jan; 19(1): 30-5.
71. Rudolph A, Song M, Brook MN, Milne RL, Mavaddat N, Michailidou K, Bolla MK, Wang Q, Dennis J, Wilcox AN, Hopper JL. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. International journal of epidemiology, 2018 Apr 1; 47(2): 526-36.
72. Wang P, Wang Y, Hang B, Zou X, Mao JH. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget, 2016 Aug 8; 7(34): 55343.
73. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. Biometrical journal, 2010 Feb; 52(1): 70-84.
74. Horne JS, Garton EO. Likelihood cross-validation versus least squares cross-validation for choosing the smoothing parameter in kernel home-range analysis. The Journal of Wildlife Management, 2006 Jun; 70(3): 641-8.
75. Wang P, Wang Y, Hang B, Zou X, Mao JH. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget, 2016 Aug 8; 7(34): 55343.
76. Sun Y, Wang R. A Risk Score System Based on the Methylation Levels of 15 RNAs in Breast Cancer. Cancer Biotherapy & Radiopharmaceuticals, 2022 Oct 1; 37(8): 697-707.