Genome-wide measures of peripheral blood DNA methylation and prostate cancer risk in a prospective nested case-control study

Liesel M. FitzGerald[1,2], Haroon Naeem[3], Enes Makalic[3], Daniel F. Schmidt[3], James G. Dowty[3], Jihoon E. Joo[4], Chol-Hee Jung[5], Julie K. Bassett[1], Pierre-Antoine Dugue[1], Jessica Chung[5], Andrew Lonie[5], Roger L. Milne[1,3], Ee Ming Wong[4], John L. Hopper[3], Dallas R. English[1,3], Gianluca Severi[1,6,7,8], Laura Baglietto[1,6], John Pedersen[9], Graham G. Giles[1,3], Melissa C. Southey[4].

[1] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC, Australia

[2] Cancer Genetics and Immunology, Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia

[3] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, VIC, Australia

[4] Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Parkville, VIC, Australia

[5] VLSCI Life Sciences Computation Centre, University of Melbourne, Carlton VIC, Australia

[6] Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, France

[7] Gustave Roussy, F-94805, Villejuif, France

[8] HuGeF, Human Genetics Foundation, 10126 Torino, Italy

[9] TissuPath, Mount Waverley, Melbourne, VIC, Australia

Running Title: Genome-wide methylation and prostate cancer risk

Correspondence to: Graham Giles, Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Rd, VIC 3004, Melbourne, Australia, Graham.Giles@cancervic.org.au

Disclosure Statement: The authors declare that they have no affiliations that they consider to be relevant to this study, nor any competing financial interests.

ABSTRACT

Background: Global measures of peripheral blood DNA methylation have been associated with risk of some malignancies, including breast, bladder and gastric cancer. Here, we examined genome-wide measures of peripheral blood DNA methylation in prostate cancer and its non-aggressive and aggressive disease forms.

Methods: We used a matched, case-control study of 687 incident prostate cancer samples, nested within a larger prospective cohort study. DNA methylation was measured in pre-diagnostic, peripheral blood samples using the Illumina Infinium HM450K BeadChip. Genome-wide measures of DNA methylation were computed as the median M-value of all CpG sites and according to CpG site location and regulatory function. We used conditional logistic regression to test for associations between genome-wide measures of DNA methylation and risk of prostate cancer and its subtypes, and by time between blood draw and diagnosis.

Results: We observed no associations between the genome-wide measure of DNA methylation based on all CpG sites and risk of prostate cancer or aggressive disease. Risk of non-aggressive disease was associated with higher methylation of CpG islands (OR=0.80; 95% CI=0.68-0.94), promoter regions (OR=0.79; 95% CI=0.66-0.93) and high density CpG regions (OR=0.80; 95% CI=0.68-0.94). Additionally, higher methylation of all CpGs (OR=0.66; 95% CI=0.48-0.89), CpG shores (OR=0.62; 95% CI=0.45-0.84) and regulatory regions (OR=0.68; 95% CI=0.51-0.91) was associated with a reduced risk of overall prostate cancer within 5 years of blood draw but not thereafter.

Conclusions: A reduced risk of overall prostate cancer within 5 years of blood draw and non-aggressive prostate cancer was associated with higher genome-wide methylation of peripheral blood DNA. While these data have no immediate clinical utility, with further work they may provide insight into the early events of prostate carcinogenesis.

Keywords: Prostate cancer, DNA methylation, peripheral blood, biomarker, HM450K array

INTRODUCTION

The concept and measurement of global methylation has recently been enhanced via the capacity to move from the measurement of surrogate markers, such as repetitive genomic elements, to platforms that achieve genome-wide assessment. Genome-wide changes in DNA methylation that occur prior to cancer diagnosis have now been described by several studies and could provide new information to improve personalised cancer risk prediction models. A study of pre-diagnostic blood samples from the Shanghai Women's Health Study cohort estimated global methylation via bisulphite pyrosequencing of repetitive Alu and long interspersed nucleotide elements (LINE)-1 and found Alu methylation (but not (LINE)-1) was inversely associated with an increased risk of gastric cancer [1]. Using similar technology, Barry *et al.* (2015) estimated DNA methylation at Alu and (LINE)-1 repetitive elements in pre-diagnostic blood samples from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and although there was no evidence for an association with prostate cancer risk overall, hypermethylation of Alu repeats (only) was associated with an increased risk in DNA prepared from blood samples taken at least four years prior to diagnosis [2]. Andreotti *et al.* (2014) also examined pre-diagnostic blood samples from the PLCO and reported increased (LINE)-1 methylation to be associated with increased bladder cancer risk for males, especially in male smokers [3]. In a recent review, Brennan and Flanagan (2012) stressed the limitations of surrogate genome-wide methylation measures and highlighted the importance of

examining methylation at specific genomic locations and regulatory regions for prospective studies testing for associations between methylation and disease risk [4]. Since this review, a small number of studies have utilised newer technologies to measure genome-wide methylation, which also enable sub-analyses of methylation measures at specific genomic locations and variously defined regulatory regions. Two recent prospective studies of breast cancer (within the Melbourne Collaborative Cohort Study (MCCS) [5] and the European Prospective Investigation into Cancer and Nutrition (EPIC) study [6]) have utilised the Illumina Infinium HumanMethylation450 (HM450K) BeadChip array and report genome-wide (mean beta values across the CpGs) hypomethylation to be associated with increased breast cancer risk [5,6]. Analysis of functional subcategories of CpG sites in the MCCS found that genome-wide DNA methylation of genomic regions within functional promoters was associated with increased breast cancer risk and both studies found that hypomethylation of CpG sites outside of promoter regions (e.g. CpG shelves, non-regulatory regions, gene bodies) was associated with decreased breast cancer risk [5,6]. These results were also supported by whole-genome bisulphite sequencing data from the prospective Breakthrough Generations Study (BGS), however, using the HM450K BeadChip, the Norwegian Women and Cancer (NOWAC) study provided no evidence for an association between breast cancer risk and genome-wide measures of methylation [6]. A nested MCCS study of mature B-cell neoplasms (MBCN) reported genome-wide hypomethylation calculated from HM450K BeadChip measures to be associated with an increased risk of this disease [7]. Interestingly, analysis of functional subcategories of CpG sites demonstrated an association between hypermethylation of promoter regions and increased risk of MBCN, which was also observed in the MCCS breast cancer study [5].

Here, we report a nested case-control study of 1,374 male participants from the MCCS. Using the Illumina Infinium HM450K BeadChip, we derived genome-wide measures of methylation from blood DNA to test the hypothesis that differential global methylation of DNA from peripheral blood is associated with overall prostate cancer risk or its subtypes defined by aggressive features.

MATERIALS AND METHODS

Participants

Between 1990 and 1994, 41,514 healthy adult volunteers (17,045 men) aged between 27 and 76 years (99% aged 40-69) were recruited into the prospective MCCS [8]. Men with a history of any cancer prior to enrolment, apart from non-melanocytic skin cancer, and men without a dried blood spot baseline sample were excluded. Incident cases of invasive (including metastatic) adenocarcinoma of the prostate were identified up to 31 December 2010 by linkage with the centralised Victorian Cancer Registry, which receives mandatory notification of all new cancer cases in Victoria, and the Australian Cancer Database maintained by the Australian Institute of Health and Welfare in Canberra. Aggressive cases were defined as: Gleason score 8-10; tumour stage T4, N+ or M+; and/or, suffered fatal prostate cancer. Male controls from the MCCS were individually matched to cases on year of birth, country of birth and DNA source using density sampling with age as the time scale.

Study participants provided written, informed consent in accordance with the Declaration of Helsinki. The study was approved by Cancer Council Victoria's Human Research Ethics Committee and performed in accordance with the institution's ethical guidelines.

DNA source and extraction

Peripheral blood specimens were collected at the time of study enrolment (baseline) in the form of peripheral blood mononuclear cells (PBMC), buffy coats or dried blood spots collected onto filter paper (Whatman, United Kingdom). DNA was prepared from PBMC and buffy coat specimens using QIAamp mini spin columns (Qiagen, Germany). DNA was prepared from twenty-one 3.2mm diameter punches from the dried blood spots after lysis in phosphate buffered saline using a TissueLyser (Qiagen, Germany). The supernatant was processed using Qiagen mini spin columns according to the manufacturer's protocol as described previously [9].

DNA bisulfite conversion and hybridisation

Bisulfite conversion was performed on a minimum of 0.3µg DNA (assessed using the Quant-iT™ Picogreen® dsDNA assay measured on the Qubit® Fluorometer (Life Technologies, USA)), using the Zymo Gold single tube kit according to manufacturer's instructions (Zymo Research, USA). Post-conversion quality control was performed using SYBR Green-based quantitative PCR, designed to determine the success of bisulfite conversion by comparing amplification of the test sample with unconverted and converted high-quality DNA controls. Samples with late amplification (Cp>5 compared to high-quality converted DNA) were re-extracted and re-converted, if feasible.

Samples were processed in batches of 96 (8 HM450K BeadChips per batch). In order to minimise potential batch effects, matched cases and controls were processed together and run on the same BeadChip and cancer subtypes (non-aggressive and aggressive) were evenly distributed across the chips and plates. Paired cases and controls were randomly positioned on each BeadChip to reduce any possible position effects within chips. Two pairs of technical replicates and a reference duplicate (DNA prepared from the multiple myeloma cell line U266) were included on each plate. A total of 200ng of bisulfite converted DNA was whole-genome amplified and hybridised onto the BeadChips according to the manufacturer's instructions. The TECAN automated liquid handler (Tecan Group Ltd, Switzerland) was used for the single-base extension and BeadChip staining steps.

HM450K BeadChip pre-processing and quality control

Raw intensity signals were imported into R programming software (www.r-project.org) using the *minfi* Bioconductor package [10]. Data were pre-processed and normalised to control probes using the "preprocessIllumina" *minfi* function. Subset-quantile within array normalization (SWAN) was performed to correct for type I and II probe bias [11]. Samples were excluded if >5% of CpG probes (excluding chromosome X and Y probes) had a detection p-value >0.01, while CpG probes were excluded from further analysis if they had missing values in >20% of samples. In addition, a

total of 240,050 CpG sites were excluded as they were considered to be affected by genomic factors (i.e., single nucleotide polymorphisms, small insertions and deletions, repetitive DNA regions and regions with reduced genomic complexity) [12]. Finally to minimize chip and batch effects, ComBat normalisation was applied to the remaining 245,462 CpG sites [13].

Statistical Analyses

Statistical analyses were performed using median M-values calculated using *minfi* [10], where the M-value was defined as log_2 *(Meth/Unmeth)*, and *Meth* and *Unmeth* were the intensities of the methylated and unmethylated probes, respectively.

DNA methylation was defined as overall (i.e. the median M-value across all 245,462 CpGs) and as categories of CpG sites based on their genomic location and/or regulatory function (e.g. the median M-value of CpGs located in CpG islands). These CpG categories were defined using the Illumina annotation file v1.2 and included: CpG islands, shores or shelves; and promoter, regulatory, non-regulatory and gene body regions. Promoter regions were further stratified by their CpG content and ratio, as differential CpG content within promoter regions is known to influence the methylation profile and gene expression [14,15]. High density CpG promoters, intermediate density CpG promoters, and low density CpG promoters were classified using a published annotation file [16].

Principal components analysis was performed on all samples and CpGs that passed quality control criteria to determine if specimen source (PBMC, buffy coat or Guthrie card) affected methylation profiles (**Supplementary Figure I**). Due to distinct clustering by specimen source and the fact that the majority of specimens were from Guthrie cards, only DNA methylation measurements from Guthrie cards were included in subsequent analyses.

Associations between DNA methylation and overall, non-aggressive and aggressive prostate cancer risk were assessed using conditional logistic regression to compute odds ratios (OR) per standard deviation of genome-wide methylation measure and 95% confidence intervals (CI). The effect of

time interval since blood draw on the association between methylation and prostate cancer risk was also analysed using conditional logistic regression where case-control pairs were categorised considering <5 vs. ≥5-year time intervals between blood draw and the case's diagnosis. For all analyses, DNA methylation measures were treated as continuous and statistical significance was determined using the likelihood ratio test (LRT). Tests for heterogeneity were implemented as tests for interactions. As a sensitivity analysis to determine the effect of excluding 240,050 CpGs, all of the above analyses were repeated with all 485,512 CpGs.

RESULTS

Between baseline and 31 December 2010, 1,464 male MCCS participants were diagnosed with adenocarcinoma of the prostate, of whom 442 were diagnosed with aggressive disease. Three hundred and twenty-nine aggressive cases had an available Guthrie card DNA sample and were assayed using the HM450K BeadChip along with matched controls and a random sample of 359 non-aggressive case-control sets. One non-aggressive case-control set, which failed quality control, and 240,050 CpGs described in Naeem *et al*. (2014) [12] were excluded from the final analyses (**Supplementary Figure II**). The clinical characteristics of cases are summarised in **Table I**.

No associations were observed between risk of overall prostate cancer and genome-wide measures of methylation, either for all CpGs or categories of CpG sites based on location or regulatory function (**Table II**). When cases were stratified by aggressiveness, several significant associations were observed with non-aggressive prostate cancer. Hypermethylation of CpG sites within CpG islands (OR=0.80; 95% CI=0.68-0.94), promoter (OR=0.79; 95% CI=0.66-0.93), high density CpG promoter regions (OR=0.80; 95% CI=0.68-0.94), and other regulatory regions (OR=0.84; 95% CI=0.72-0.98) was associated with a reduced risk of non-aggressive disease (**Table II**). With the exception of methylation in other regulatory regions, these associations were significantly different (p_{het}<0.02) from those observed for aggressive cases, where no associations were observed. Supplemental case-case analyses also highlighted CpG islands, promoter and high density CpG

promoter regions as having significantly different methylation patterns between aggressive and non-aggressive cases (**Supplementary Table I**).

When participants were stratified by time since blood draw, hypermethylation of all CpGs (OR=0.66; 95% CI=0.48-0.89) and hypermethylation of CpG sites within CpG island shores (OR=0.62; 95% CI=0.45-0.84) and other regulatory regions (OR=0.68; 95% CI=0.51-0.91) were associated with a reduced risk of overall prostate cancer within five years of blood draw (**Table III**). These results were significantly different to those for overall prostate cancer risk five or more years after blood draw (all $p<0.03$), where no associations were observed.

To investigate whether the exclusion of probes that are predicted to be affected by genomic factors influenced the results, analyses were repeated with all 485,512 probes. Similar results were observed between the analyses of the filtered probe set and the full probe set across all categories of global methylation and disease subtype stratifications, except in the time since blood draw analyses where no association was observed between global DNA hypermethylation and risk of prostate cancer within five years of blood draw (**Supplementary Tables I - III**).

DISCUSSION

In some contexts, methylation profiling of DNA from peripheral blood indicates promise for cancer risk prediction, early detection and prognosis [1-5,17,18]. Several previous studies have shown an association between measures of methylation and cancer risk using both surrogate global measures and, more recently, high-density genome-wide measures [1-3,5-7]. While it is not clear yet why genome-wide methylation patterns are altered in pre-diagnostic, peripheral blood DNA, it is likely to be a combination of influences, including altered epigenetic regulation, immunological changes from inherited/acquired genetic variants or accumulated exposures, or result from the very early stages of carcinogenesis.

Here, we used the Illumina Infinium HM450K BeadChip to assess whether genome-wide DNA methylation patterns from peripheral blood are associated with risk of prostate cancer, with specific interest in risk of non-aggressive and aggressive disease. Application of the HM450K BeadChip provided us with the capacity to perform more sophisticated analyses (e.g. the ability to conduct sub-analyses of predefined groups of measured CpGs) but it also distinguishes our work from the prior literature that has utilised surrogate markers of methylation, such as Alu and (LINE)-1 repetitive elements. In a prior study by Barry and colleagues (2015), higher Alu methylation was found to be associated with an increase in prostate cancer risk when diagnosis was four or more years after blood draw [2]. This is in contrast to our findings, where we observed genome-wide hypermethylation (and hypermethylation of CpG shores and other regulatory regions) to be associated with a reduced risk of prostate cancer within five years of blood draw and not later. A limited number of other studies, applying a range of methodological approaches, have also noted differences in DNA methylation patterns depending on the time between blood draw and diagnosis [1,2,5]. Examining methylation of repetitive elements, Gao and colleagues (2012) observed an association between hypomethylation and decreased bladder cancer risk when diagnosis was within 1 year of blood draw, and an association between hypomethylation and increased cancer risk when diagnosis was a year or more after blood draw [1]. In a study of breast cancer risk, using genome-wide methylation measurements derived from the HM450K BeadChip, we found a difference in genome-wide DNA methylation between cases and controls that decreased with increasing time since blood draw, where the difference was negligible after 5 years [5]. Some of the apparent contradictions in the above-mentioned studies could be due to methodological differences and more work is required to understand the relationship between these different global methylation measures.

It has been well documented that a number of probes included on the HM450K array do not perform well technically. During our analyses, we excluded CpG probes that Naeem *et al.* (2014) have suggested to be affected by genomic factors [12]. Their study compared HM450K

methylation data to whole-genome bisulfite sequencing data and observed that the methylation status of particular probes can be affected by being able to hybridise to multiple locations or repetitive regions in the genome, by binding to regions containing insertions or deletions (especially relevant in population-wide studies such as ours) and by the presence of a genetic polymorphism either within the CpG site itself or within the probe-binding region [12]. Naeem *et al.* demonstrated that removing the measurements of these probes from their analyses resulted in a set of high-quality probes that reduced false-positive signals, decreased within tissue standard deviation, yet still provided adequate genome-wide coverage, with only the HLA region on chromosome 6 being devoid of probes. Taking this approach, we removed 240,050 CpG probes from the analyses reported here but also repeated the analyses with the complete probe dataset. These analyses yielded very similar results except that the association between genome-wide DNA methylation measures and risk of prostate cancer within five years of blood draw did not persist. It is possible that when considering global methylation, the removal of a select number of technically deficient probes may not have a significant effect on the outcome of analyses provided there is still adequate genome coverage.

We consider the design of our prospective study to be a strength. Detailed information collected at blood draw allowed us to carefully match cases and controls on age, ethnicity and DNA source, and also allowed us to stratify cases according to clinical disease aggressiveness. Additionally, potential batch effects were mitigated by placing matched cases and controls adjacent to each other, but in a random position, on the same chip, resulting in minimal technical bias [19].

One shortcoming of our study is its lack of replication. To our knowledge, there are currently no other prostate cancer studies with HM450K methylation data available for pre-diagnostic blood DNA samples. It is important that other cohorts examine associations between genome-wide DNA methylation and prostate cancer risk, and for data pooling to occur in order to sufficiently investigate the individual CpG components of the genome-wide methylation measure.

CONCLUSION

Our study of genome-wide measures of methylation in peripheral blood DNA provides no evidence of association with overall prostate cancer risk. The data provide evidence that methylation of peripheral blood DNA is associated with non-aggressive disease; the hypomethylation of CpG islands, promoter and other regulatory regions being associated with an increased risk of indolent disease. This is interesting, as few of the common genetic risk factors for prostate cancer, identified through genome-wide association studies, distinguish aggressive from non-aggressive disease [20,21], suggesting that the underlying driver(s) of these phenotypes are still yet to be identified. We did detect an association with risk of overall prostate cancer when measured within five years of diagnosis but we had inadequate statistical power to stratify our analysis more substantially. While these data have no immediate clinical utility, they may provide some insight regarding early events in prostate carcinogenesis. It is important for this work to be extended, both to confirm these findings and also to accumulate data sufficient to conduct association studies at the resolution of individual CpGs. The findings of this study do not exclude the possibility that some individual CpGs (or groups of CpGs) within this dataset are associated with overall prostate cancer risk or its indolent and aggressive subtypes.

ACKNOWLEDGMENTS

This study was made possible by the contribution of many people, including the original investigators and the diligent team who recruited the participants and who continue working on follow-up. We would also like to express our gratitude to the many thousands of Melbourne residents who took part in the study and provided blood samples.

FUNDING

This work was supported by the National Health and Medical Research Council project grant number 1026892.

References

1. Gao Y, Baccarelli A, Shu XO, Ji BT, Yu K, Tarantini L, Yang G, Li HL, Hou L, Rothman N, Zheng W, Gao YT, Chow WH. Blood leukocyte Alu and LINE-1 methylation and gastric cancer risk in the Shanghai Women's Health Study. British Journal of Cancer 2012;106(3):585-591.
2. Barry KH, Moore LE, Liao LM, Huang WY, Andreotti G, Poulin M, Berndt SI. Prospective study of DNA methylation at LINE-1 and Alu in peripheral blood and the risk of prostate cancer. The Prostate 2015;75(15):1718-1725.
3. Andreotti G, Karami S, Pfeiffer RM, Hurwitz L, Liao LM, Weinstein SJ, Albanes D, Virtamo J, Silverman DT, Rothman N, Moore LE. LINE1 methylation levels associated with increased bladder cancer risk in pre-diagnostic blood DNA among US (PLCO) and European (ATBC) cohort study participants. Epigenetics 2014;9(3):404-415.
4. Brennan K, Flanagan JM. Is there a link between genome-wide hypomethylation in blood and cancer risk? Cancer Prevention Research 2012;5(12):1345-1357.
5. Severi G, Southey MC, English DR, Jung CH, Lonie A, McLean C, Tsimiklis H, Hopper JL, Giles GG, Baglietto L. Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. Breast Cancer Research and Treatment 2014;148(3):665-673.
6. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, Mattiello A, Palli D, Masala G, Krogh V, Agnoli C, Tumino R, Frasca G, Flower K, Curry E, Orr N, Tomczyk K, Jones ME, Ashworth A, Swerdlow A, Chadeau-Hyam M, Lund E, Garcia-Closas M, Sandanger TM, Flanagan JM, Vineis P. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clinical Epigenetics 2015;7(1):67.

7. Wong Doo N, Makalic E, Joo JE, Vajdic CM, Schmidt DF, Wong EM, Jung CH, Severi G, Park DJ, Chung J, Baglietto L, Prince HM, Seymour JF, Tam C, Hopper JL, English DR, Milne RL, Harrison SJ, Southey MC, Giles GG. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. Epigenomics 2016;8(1):55-66.

8. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC Scientific Publications 2002;156(69-70.

9. Joo JE, Wong EM, Baglietto L, Jung CH, Tsimiklis H, Park DJ, Wong NC, English DR, Hopper JL, Severi G, Giles GG, Southey MC. The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. BMC Biotechnology 2013;13(23.

10. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30(10):1363-1369.

11. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biology 2012;13(6):R44.

12. Naeem H, Wong NC, Chatterton Z, Hong MK, Pedersen JS, Corcoran NM, Hovens CM, Macintyre G. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. BMC Genomics 2014;15(51.

13. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118-127.

14. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics 2007;39(4):457-466.

15. Martin-Subero JI, Kreuz M, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E, Rosolowski M, Richter J, Lopez-Serra L, Ballestar E, Berger H, Agirre X, Bernd HW,

Calvanese V, Cogliatti SB, Drexler HG, Fan JB, Fraga MF, Hansmann ML, Hummel M, Klapper W, Korn B, Kuppers R, Macleod RA, Moller P, Ott G, Pott C, Prosper F, Rosenwald A, Schwaenen C, Schubeler D, Seifert M, Sturzenhofecker B, Weber M, Wessendorf S, Loeffler M, Trumper L, Stein H, Spang R, Esteller M, Barker D, Hasenclever D, Siebert R. New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. Blood 2009;113(11):2488-2497.

16. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, Robinson WP, Kobor MS. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & Chromatin 2013;6(1):4.

17. Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. Journal of Clinical Oncology 2011;29(9):1133-1139.

18. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M. An epigenetic signature in peripheral blood predicts active ovarian cancer. PloS One 2009;4(12):e8274.

19. Harper KN, Peters BA, Gamble MV. Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. Cancer Epidemiology, Biomarkers & Prevention 2013;22(6):1052-1060.

20. Berndt SI, Wang Z, Yeager M, Alavanja MC, Albanes D, Amundadottir L, Andriole G, Beane Freeman L, Campa D, Cancel-Tassin G, Canzian F, Cornu JN, Cussenot O, Diver WR, Gapstur SM, Gronberg H, Haiman CA, Henderson B, Hutchinson A, Hunter DJ, Key TJ, Kolb S, Koutros S, Kraft P, Le Marchand L, Lindstrom S, Machiela MJ, Ostrander EA, Riboli E, Schumacher F, Siddiq A, Stanford JL, Stevens VL, Travis RC, Tsilidis KK, Virtamo J, Weinstein S, Wilkund F, Xu J, Lilly Zheng S, Yu K, Wheeler W, Zhang H,

Sampson J, Black A, Jacobs K, Hoover RN, Tucker M, Chanock SJ. Two susceptibility loci identified for prostate cancer aggressiveness. Nature Communications 2015;6(6889.

21. FitzGerald LM, Kwon EM, Conomos MP, Kolb S, Holt SK, Levine D, Feng Z, Ostrander EA, Stanford JL. Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer. Cancer Epidemiology, Biomarkers & Prevention 2011;20(6):1196-1203.

Supplementary Figure I: Principal components analysis plot demonstrating the variance in methylation data due to specimen source. BC: DNA extracted from stored buffy coat, dark green circles; DNA BC: Specimen from stored buffy coat DNA, orange circles; DNA PBMC; Specimen from stored peripheral blood mononuclear cell (PBMC) DNA, purple circles; GC: DNA extracted from stored Guthrie card, pink circles; and PBMC: DNA extracted from stored PBMC samples, light green circles.

Supplementary Figure II: Flow diagram depicting the selection of MCCS prostate cancer cases and controls for whole-genome methylation analyses.

Genome-wide measures of peripheral blood DNA methylation and prostate cancer risk in a prospective nested case-control study

Liesel M. FitzGerald[1,2], Haroon Naeem[3], Enes Makalic[3], Daniel F. Schmidt[3], James G. Dowty[3], Jihoon E. Joo[4], Chol-Hee Jung[5], Julie K. Bassett[1], Pierre-Antoine Dugue[1], Jessica Chung[5], Andrew Lonie[5], Roger L. Milne[1,3], Ee Ming Wong[4], John L. Hopper[3], Dallas R. English[1,3], Gianluca Severi[1,6,7,8], Laura Baglietto[1,6], John Pedersen[9], Graham G. Giles[1,3], Melissa C. Southey[4].

[1] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC, Australia

[2] Cancer Genetics and Immunology, Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia

[3] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, VIC, Australia

[4] Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Parkville, VIC, Australia

[5] VLSCI Life Sciences Computation Centre, University of Melbourne, Carlton VIC, Australia

[6] Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, France

[7] Gustave Roussy, F-94805, Villejuif, France

[8] HuGeF, Human Genetics Foundation, 10126 Torino, Italy

[9] TissuPath, Mount Waverley, Melbourne, VIC, Australia

Correspondence to: Graham Giles, Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Rd, VIC 3004, Melbourne, Australia, Graham.Giles@cancervic.org.au

Disclosure Statement: The authors declare that they have no affiliations that they consider to be relevant to this study, nor any competing financial interests.

ABSTRACT

Background: Global measures of peripheral blood DNA methylation have been associated with risk of some malignancies, including breast, bladder and gastric cancer. Here, we examined genome-wide measures of peripheral blood DNA methylation in prostate cancer and its non-aggressive and aggressive disease forms.

Methods: We used a matched, case-control study of 687 incident prostate cancer samples, nested within a larger prospective cohort study. DNA methylation was measured in pre-diagnostic, peripheral blood samples using the Illumina Infinium HM450K BeadChip. Genome-wide measures of DNA methylation were computed as the median M-value of all CpG sites and according to CpG

site location and regulatory function. We used conditional logistic regression to test for associations between genome-wide measures of DNA methylation and risk of prostate cancer and its subtypes, and by time between blood draw and diagnosis.

Results: We observed no associations between the genome-wide measure of DNA methylation based on all CpG sites and risk of prostate cancer or aggressive disease. Risk of non-aggressive disease was associated with higher methylation of CpG islands (OR=0.80; 95% CI=0.68-0.94), promoter regions (OR=0.79; 95% CI=0.66-0.93) and high density CpG regions (OR=0.80; 95% CI=0.68-0.94). Additionally, higher methylation of all CpGs (OR=0.66; 95% CI=0.48-0.89), CpG shores (OR=0.62; 95% CI=0.45-0.84) and regulatory regions (OR=0.68; 95% CI=0.51-0.91) was associated with a reduced risk of overall prostate cancer within 5 years of blood draw but not thereafter.

Conclusions: A reduced risk of overall prostate cancer within 5 years of blood draw and non-aggressive prostate cancer was associated with higher genome-wide methylation of peripheral blood DNA. While these data have no immediate clinical utility, with further work they may provide insight into the early events of prostate carcinogenesis.

Keywords: Prostate cancer, DNA methylation, peripheral blood, biomarker, HM450K array

INTRODUCTION

The concept and measurement of global methylation has recently been enhanced via the capacity to move from the measurement of surrogate markers, such as repetitive genomic elements, to platforms that achieve genome-wide assessment. Genome-wide changes in DNA methylation that occur prior to cancer diagnosis have now been described by several studies and could provide new information to improve personalised cancer risk prediction models. A study of pre-diagnostic blood samples from the Shanghai Women's Health Study cohort estimated global methylation via bisulphite pyrosequencing of repetitive Alu and long interspersed nucleotide elements (LINE)-1 and

found Alu methylation (but not (LINE)-1) was inversely associated with an increased risk of gastric cancer [1]. Using similar technology, Barry *et al.* (2015) estimated DNA methylation at Alu and (LINE)-1 repetitive elements in pre-diagnostic blood samples from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and although there was no evidence for an association with prostate cancer risk overall, hypermethylation of Alu repeats (only) was associated with an increased risk in DNA prepared from blood samples taken at least four years prior to diagnosis [2]. Andreotti *et al.* (2014) also examined pre-diagnostic blood samples from the PLCO and reported increased (LINE)-1 methylation to be associated with increased bladder cancer risk for males, especially in male smokers [3]. In a recent review, Brennan and Flanagan (2012) stressed the limitations of surrogate genome-wide methylation measures and highlighted the importance of examining methylation at specific genomic locations and regulatory regions for prospective studies testing for associations between methylation and disease risk [4]. Since this review, a small number of studies have utilised newer technologies to measure genome-wide methylation, which also enable sub-analyses of methylation measures at specific genomic locations and variously defined regulatory regions. Two recent prospective studies of breast cancer (within the Melbourne Collaborative Cohort Study (MCCS) [5] and the European Prospective Investigation into Cancer and Nutrition (EPIC) study [6]) have utilised the Illumina Infinium HumanMethylation450 (HM450K) BeadChip array and report genome-wide (mean beta values across the CpGs) hypomethylation to be associated with increased breast cancer risk [5,6]. Analysis of functional subcategories of CpG sites in the MCCS found that genome-wide DNA methylation of genomic regions within functional promoters was associated with increased breast cancer risk and both studies found that hypomethylation of CpG sites outside of promoter regions (e.g. CpG shelves, non-regulatory regions, gene bodies) was associated with decreased breast cancer risk [5,6]. These results were also supported by whole-genome bisulphite sequencing data from the prospective Breakthrough Generations Study (BGS), however, using the HM450K BeadChip, the Norwegian Women and Cancer (NOWAC) study provided no evidence for an association between breast

cancer risk and genome-wide measures of methylation [6]. A nested MCCS study of mature B-cell neoplasms (MBCN) reported genome-wide hypomethylation calculated from HM450K BeadChip measures to be associated with an increased risk of this disease [7]. Interestingly, analysis of functional subcategories of CpG sites demonstrated an association between hypermethylation of promoter regions and increased risk of MBCN, which was also observed in the MCCS breast cancer study [5].

Here, we report a nested case-control study of 1,374 male participants from the MCCS. Using the Illumina Infinium HM450K BeadChip, we derived genome-wide measures of methylation from blood DNA to test the hypothesis that differential global methylation of DNA from peripheral blood is associated with overall prostate cancer risk or its subtypes defined by aggressive features.

MATERIALS AND METHODS

Participants

Between 1990 and 1994, 41,514 healthy adult volunteers (17,045 men) aged between 27 and 76 years (99% aged 40-69) were recruited into the prospective MCCS [8]. Men with a history of any cancer prior to enrolment, apart from non-melanocytic skin cancer, and men without a dried blood spot baseline sample were excluded. Incident cases of invasive (including metastatic) adenocarcinoma of the prostate were identified up to 31 December 2010 by linkage with the centralised Victorian Cancer Registry, which receives mandatory notification of all new cancer cases in Victoria, and the Australian Cancer Database maintained by the Australian Institute of Health and Welfare in Canberra. Aggressive cases were defined as: Gleason score 8-10; tumour stage T4, N+ or M+; and/or, suffered fatal prostate cancer. Male controls from the MCCS were individually matched to cases on year of birth, country of birth and DNA source using density sampling with age as the time scale.

Study participants provided written, informed consent in accordance with the Declaration of Helsinki. The study was approved by Cancer Council Victoria's Human Research Ethics Committee and performed in accordance with the institution's ethical guidelines.

DNA source and extraction

Peripheral blood specimens were collected at the time of study enrolment (baseline) in the form of peripheral blood mononuclear cells (PBMC), buffy coats or dried blood spots collected onto filter paper (Whatman, United Kingdom). DNA was prepared from PBMC and buffy coat specimens using QIAamp mini spin columns (Qiagen, Germany). DNA was prepared from twenty-one 3.2mm diameter punches from the dried blood spots after lysis in phosphate buffered saline using a TissueLyser (Qiagen, Germany). The supernatant was processed using Qiagen mini spin columns according to the manufacturer's protocol as described previously [9].

DNA bisulfite conversion and hybridisation

Bisulfite conversion was performed on a minimum of 0.3µg DNA (assessed using the Quant-iT™ Picogreen® dsDNA assay measured on the Qubit® Fluorometer (Life Technologies, USA)), using the Zymo Gold single tube kit according to manufacturer's instructions (Zymo Research, USA). Post-conversion quality control was performed using SYBR Green-based quantitative PCR, designed to determine the success of bisulfite conversion by comparing amplification of the test sample with unconverted and converted high-quality DNA controls. Samples with late amplification (Cp>5 compared to high-quality converted DNA) were re-extracted and re-converted, if feasible.

Samples were processed in batches of 96 (8 HM450K BeadChips per batch). In order to minimise potential batch effects, matched cases and controls were processed together and run on the same BeadChip and cancer subtypes (non-aggressive and aggressive) were evenly distributed across the chips and plates. Paired cases and controls were randomly positioned on each BeadChip to reduce any possible position effects within chips. Two pairs of technical replicates and a reference duplicate (DNA prepared from the multiple myeloma cell line U266) were included on each plate.

A total of 200ng of bisulfite converted DNA was whole-genome amplified and hybridised onto the BeadChips according to the manufacturer's instructions. The TECAN automated liquid handler (Tecan Group Ltd, Switzerland) was used for the single-base extension and BeadChip staining steps.

HM450K BeadChip pre-processing and quality control

Raw intensity signals were imported into R programming software (www.r-project.org) using the *minfi* Bioconductor package [10]. Data were pre-processed and normalised to control probes using the "preprocessIllumina" *minfi* function. Subset-quantile within array normalization (SWAN) was performed to correct for type I and II probe bias [11]. Samples were excluded if >5% of CpG probes (excluding chromosome X and Y probes) had a detection p-value >0.01, while CpG probes were excluded from further analysis if they had missing values in >20% of samples. In addition, a total of 240,050 CpG sites were excluded as they were considered to be affected by genomic factors (i.e., single nucleotide polymorphisms, small insertions and deletions, repetitive DNA regions and regions with reduced genomic complexity) [12]. Finally to minimize chip and batch effects, ComBat normalisation was applied to the remaining 245,462 CpG sites [13].

Statistical Analyses

Statistical analyses were performed using median M-values calculated using *minfi* [10], where the M-value was defined as log_2 *(Meth/Unmeth)*, and *Meth* and *Unmeth* were the intensities of the methylated and unmethylated probes, respectively.

DNA methylation was defined as overall (i.e. the median M-value across all 245,462 CpGs) and as categories of CpG sites based on their genomic location and/or regulatory function (e.g. the median M-value of CpGs located in CpG islands). These CpG categories were defined using the Illumina annotation file v1.2 and included: CpG islands, shores or shelves; and promoter, regulatory, non-regulatory and gene body regions. Promoter regions were further stratified by their CpG content and ratio, as differential CpG content within promoter regions is known to influence the methylation

profile and gene expression [14,15]. High density CpG promoters, intermediate density CpG promoters, and low density CpG promoters were classified using a published annotation file [16].

Principal components analysis was performed on all samples and CpGs that passed quality control criteria to determine if specimen source (PBMC, buffy coat or Guthrie card) affected methylation profiles (**Supplementary Figure I**). Due to distinct clustering by specimen source and the fact that the majority of specimens were from Guthrie cards, only DNA methylation measurements from Guthrie cards were included in subsequent analyses.

Associations between DNA methylation and overall, non-aggressive and aggressive prostate cancer risk were assessed using conditional logistic regression to compute odds ratios (OR) per standard deviation of genome-wide methylation measure and 95% confidence intervals (CI). The effect of time interval since blood draw on the association between methylation and prostate cancer risk was also analysed using conditional logistic regression where case-control pairs were categorised considering <5 vs. ≥5-year time intervals between blood draw and the case's diagnosis. For all analyses, DNA methylation measures were treated as continuous and statistical significance was determined using the likelihood ratio test (LRT). Tests for heterogeneity were implemented as tests for interactions. As a sensitivity analysis to determine the effect of excluding 240,050 CpGs, all of the above analyses were repeated with all 485,512 CpGs.

RESULTS

Between baseline and 31 December 2010, 1,464 male MCCS participants were diagnosed with adenocarcinoma of the prostate, of whom 442 were diagnosed with aggressive disease. Three hundred and twenty-nine aggressive cases had an available Guthrie card DNA sample and were assayed using the HM450K BeadChip along with matched controls and a random sample of 359 non-aggressive case-control sets. One non-aggressive case-control set, which failed quality control,

and 240,050 CpGs described in Naeem *et al*. (2014) [12] were excluded from the final analyses (**Supplementary Figure II**). The clinical characteristics of cases are summarised in **Table I**.

No associations were observed between risk of overall prostate cancer and genome-wide measures of methylation, either for all CpGs or categories of CpG sites based on location or regulatory function (**Table II**). When cases were stratified by aggressiveness, several significant associations were observed with non-aggressive prostate cancer. Hypermethylation of CpG sites within CpG islands (OR=0.80; 95% CI=0.68-0.94), promoter (OR=0.79; 95% CI=0.66-0.93), high density CpG promoter regions (OR=0.80; 95% CI=0.68-0.94), and other regulatory regions (OR=0.84; 95% CI=0.72-0.98) was associated with a reduced risk of non-aggressive disease (**Table II**). With the exception of methylation in other regulatory regions, these associations were significantly different (p_{het}<0.02) from those observed for aggressive cases, where no associations were observed. Supplemental case-case analyses also highlighted CpG islands, promoter and high density CpG promoter regions as having significantly different methylation patterns between aggressive and non-aggressive cases (**Supplementary Table I**).

When participants were stratified by time since blood draw, hypermethylation of all CpGs (OR=0.66; 95% CI=0.48-0.89) and hypermethylation of CpG sites within CpG island shores (OR=0.62; 95% CI=0.45-0.84) and other regulatory regions (OR=0.68; 95% CI=0.51-0.91) were associated with a reduced risk of overall prostate cancer within five years of blood draw (**Table III**). These results were significantly different to those for overall prostate cancer risk five or more years after blood draw (all p<0.03), where no associations were observed.

To investigate whether the exclusion of probes that are predicted to be affected by genomic factors influenced the results, analyses were repeated with all 485,512 probes. Similar results were observed between the analyses of the filtered probe set and the full probe set across all categories of global methylation and disease subtype stratifications, except in the time since blood draw analyses where no association was observed between global DNA hypermethylation and risk of prostate cancer within five years of blood draw (**Supplementary Tables I - III**).

DISCUSSION

In some contexts, methylation profiling of DNA from peripheral blood indicates promise for cancer risk prediction, early detection and prognosis [1-5,17,18]. Several previous studies have shown an association between measures of methylation and cancer risk using both surrogate global measures and, more recently, high-density genome-wide measures [1-3,5-7]. While it is not clear yet why genome-wide methylation patterns are altered in pre-diagnostic, peripheral blood DNA, it is likely to be a combination of influences, including altered epigenetic regulation, immunological changes from inherited/acquired genetic variants or accumulated exposures, or result from the very early stages of carcinogenesis.

Here, we used the Illumina Infinium HM450K BeadChip to assess whether genome-wide DNA methylation patterns from peripheral blood are associated with risk of prostate cancer, with specific interest in risk of non-aggressive and aggressive disease. Application of the HM450K BeadChip provided us with the capacity to perform more sophisticated analyses (e.g. the ability to conduct sub-analyses of predefined groups of measured CpGs) but it also distinguishes our work from the prior literature that has utilised surrogate markers of methylation, such as Alu and (LINE)-1 repetitive elements. In a prior study by Barry and colleagues (2015), higher Alu methylation was found to be associated with an increase in prostate cancer risk when diagnosis was four or more years after blood draw [2]. This is in contrast to our findings, where we observed genome-wide hypermethylation (and hypermethylation of CpG shores and other regulatory regions) to be associated with a reduced risk of prostate cancer within five years of blood draw and not later. A limited number of other studies, applying a range of methodological approaches, have also noted differences in DNA methylation patterns depending on the time between blood draw and diagnosis [1,2,5]. Examining methylation of repetitive elements, Gao and colleagues (2012) observed an association between hypomethylation and decreased bladder cancer risk when diagnosis was within 1 year of blood draw, and an association between hypomethylation and increased cancer risk when

diagnosis was a year or more after blood draw [1]. In a study of breast cancer risk, using genome-wide methylation measurements derived from the HM450K BeadChip, we found a difference in genome-wide DNA methylation between cases and controls that decreased with increasing time since blood draw, where the difference was negligible after 5 years [5]. Some of the apparent contradictions in the above-mentioned studies could be due to methodological differences and more work is required to understand the relationship between these different global methylation measures.

It has been well documented that a number of probes included on the HM450K array do not perform well technically. During our analyses, we excluded CpG probes that Naeem *et al.* (2014) have suggested to be affected by genomic factors [12]. Their study compared HM450K methylation data to whole-genome bisulfite sequencing data and observed that the methylation status of particular probes can be affected by being able to hybridise to multiple locations or repetitive regions in the genome, by binding to regions containing insertions or deletions (especially relevant in population-wide studies such as ours) and by the presence of a genetic polymorphism either within the CpG site itself or within the probe-binding region [12]. Naeem *et al.* demonstrated that removing the measurements of these probes from their analyses resulted in a set of high-quality probes that reduced false-positive signals, decreased within tissue standard deviation, yet still provided adequate genome-wide coverage, with only the HLA region on chromosome 6 being devoid of probes. Taking this approach, we removed 240,050 CpG probes from the analyses reported here but also repeated the analyses with the complete probe dataset. These analyses yielded very similar results except that the association between genome-wide DNA methylation measures and risk of prostate cancer within five years of blood draw did not persist. It is possible that when considering global methylation, the removal of a select number of technically deficient probes may not have a significant effect on the outcome of analyses provided there is still adequate genome coverage.

We consider the design of our prospective study to be a strength. Detailed information collected at blood draw allowed us to carefully match cases and controls on age, ethnicity and DNA source, and also allowed us to stratify cases according to clinical disease aggressiveness. Additionally, potential batch effects were mitigated by placing matched cases and controls adjacent to each other, but in a random position, on the same chip, resulting in minimal technical bias [19].

One shortcoming of our study is its lack of replication. To our knowledge, there are currently no other prostate cancer studies with HM450K methylation data available for pre-diagnostic blood DNA samples. It is important that other cohorts examine associations between genome-wide DNA methylation and prostate cancer risk, and for data pooling to occur in order to sufficiently investigate the individual CpG components of the genome-wide methylation measure.

CONCLUSION

Our study of genome-wide measures of methylation in peripheral blood DNA provides no evidence of association with overall prostate cancer risk. The data provide evidence that methylation of peripheral blood DNA is associated with non-aggressive disease; the hypomethylation of CpG islands, promoter and other regulatory regions being associated with an increased risk of indolent disease. This is interesting, as few of the common genetic risk factors for prostate cancer, identified through genome-wide association studies, distinguish aggressive from non-aggressive disease [20,21], suggesting that the underlying driver(s) of these phenotypes are still yet to be identified. We did detect an association with risk of overall prostate cancer when measured within five years of diagnosis but we had inadequate statistical power to stratify our analysis more substantially. While these data have no immediate clinical utility, they may provide some insight regarding early events in prostate carcinogenesis. It is important for this work to be extended, both to confirm these findings and also to accumulate data sufficient to conduct association studies at the resolution of individual CpGs. The findings of this study do not exclude the possibility that some individual

CpGs (or groups of CpGs) within this dataset are associated with overall prostate cancer risk or its indolent and aggressive subtypes.

ACKNOWLEDGMENTS

This study was made possible by the contribution of many people, including the original investigators and the diligent team who recruited the participants and who continue working on follow-up. We would also like to express our gratitude to the many thousands of Melbourne residents who took part in the study and provided blood samples.

FUNDING

This work was supported by the National Health and Medical Research Council project grant number 1026892.

References

1. Gao Y, Baccarelli A, Shu XO, Ji BT, Yu K, Tarantini L, Yang G, Li HL, Hou L, Rothman N, Zheng W, Gao YT, Chow WH. Blood leukocyte Alu and LINE-1 methylation and gastric cancer risk in the Shanghai Women's Health Study. British Journal of Cancer 2012;106(3):585-591.
2. Barry KH, Moore LE, Liao LM, Huang WY, Andreotti G, Poulin M, Berndt SI. Prospective study of DNA methylation at LINE-1 and Alu in peripheral blood and the risk of prostate cancer. The Prostate 2015;75(15):1718-1725.
3. Andreotti G, Karami S, Pfeiffer RM, Hurwitz L, Liao LM, Weinstein SJ, Albanes D, Virtamo J, Silverman DT, Rothman N, Moore LE. LINE1 methylation levels associated with increased bladder cancer risk in pre-diagnostic blood DNA among US (PLCO) and European (ATBC) cohort study participants. Epigenetics 2014;9(3):404-415.
4. Brennan K, Flanagan JM. Is there a link between genome-wide hypomethylation in blood and cancer risk? Cancer Prevention Research 2012;5(12):1345-1357.

5. Severi G, Southey MC, English DR, Jung CH, Lonie A, McLean C, Tsimiklis H, Hopper JL, Giles GG, Baglietto L. Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. Breast Cancer Research and Treatment 2014;148(3):665-673.

6. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, Mattiello A, Palli D, Masala G, Krogh V, Agnoli C, Tumino R, Frasca G, Flower K, Curry E, Orr N, Tomczyk K, Jones ME, Ashworth A, Swerdlow A, Chadeau-Hyam M, Lund E, Garcia-Closas M, Sandanger TM, Flanagan JM, Vineis P. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clinical Epigenetics 2015;7(1):67.

7. Wong Doo N, Makalic E, Joo JE, Vajdic CM, Schmidt DF, Wong EM, Jung CH, Severi G, Park DJ, Chung J, Baglietto L, Prince HM, Seymour JF, Tam C, Hopper JL, English DR, Milne RL, Harrison SJ, Southey MC, Giles GG. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. Epigenomics 2016;8(1):55-66.

8. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC Scientific Publications 2002;156(69-70.

9. Joo JE, Wong EM, Baglietto L, Jung CH, Tsimiklis H, Park DJ, Wong NC, English DR, Hopper JL, Severi G, Giles GG, Southey MC. The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. BMC Biotechnology 2013;13(23.

10. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30(10):1363-1369.

11. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biology 2012;13(6):R44.

12. Naeem H, Wong NC, Chatterton Z, Hong MK, Pedersen JS, Corcoran NM, Hovens CM, Macintyre G. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. BMC Genomics 2014;15(51.

13. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118-127.

14. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics 2007;39(4):457-466.

15. Martin-Subero JI, Kreuz M, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E, Rosolowski M, Richter J, Lopez-Serra L, Ballestar E, Berger H, Agirre X, Bernd HW, Calvanese V, Cogliatti SB, Drexler HG, Fan JB, Fraga MF, Hansmann ML, Hummel M, Klapper W, Korn B, Kuppers R, Macleod RA, Moller P, Ott G, Pott C, Prosper F, Rosenwald A, Schwaenen C, Schubeler D, Seifert M, Sturzenhofecker B, Weber M, Wessendorf S, Loeffler M, Trumper L, Stein H, Spang R, Esteller M, Barker D, Hasenclever D, Siebert R. New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. Blood 2009;113(11):2488-2497.

16. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, Robinson WP, Kobor MS. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & Chromatin 2013;6(1):4.

17. Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. Journal of Clinical Oncology 2011;29(9):1133-1139.

18. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M. An epigenetic signature in peripheral blood predicts active ovarian cancer. PloS One 2009;4(12):e8274.

19. Harper KN, Peters BA, Gamble MV. Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. Cancer Epidemiology, Biomarkers & Prevention 2013;22(6):1052-1060.

20. Berndt SI, Wang Z, Yeager M, Alavanja MC, Albanes D, Amundadottir L, Andriole G, Beane Freeman L, Campa D, Cancel-Tassin G, Canzian F, Cornu JN, Cussenot O, Diver WR, Gapstur SM, Gronberg H, Haiman CA, Henderson B, Hutchinson A, Hunter DJ, Key TJ, Kolb S, Koutros S, Kraft P, Le Marchand L, Lindstrom S, Machiela MJ, Ostrander EA, Riboli E, Schumacher F, Siddiq A, Stanford JL, Stevens VL, Travis RC, Tsilidis KK, Virtamo J, Weinstein S, Wilkund F, Xu J, Lilly Zheng S, Yu K, Wheeler W, Zhang H, Sampson J, Black A, Jacobs K, Hoover RN, Tucker M, Chanock SJ. Two susceptibility loci identified for prostate cancer aggressiveness. Nature Communications 2015;6(6889.

21. FitzGerald LM, Kwon EM, Conomos MP, Kolb S, Holt SK, Levine D, Feng Z, Ostrander EA, Stanford JL. Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer. Cancer Epidemiology, Biomarkers & Prevention 2011;20(6):1196-1203.

Supplementary Figure I: Principal components analysis plot demonstrating the variance in methylation data due to specimen source. BC: DNA extracted from stored buffy coat, dark green circles; DNA BC: Specimen from stored buffy coat DNA, orange circles; DNA PBMC; Specimen from stored peripheral blood mononuclear cell (PBMC) DNA, purple circles; GC: DNA extracted from stored Guthrie card, pink circles; and PBMC: DNA extracted from stored PBMC samples, light green circles.

Supplementary Figure II: Flow diagram depicting the selection of MCCS prostate cancer cases and controls for whole-genome methylation analyses.

Liesel M. FitzGerald[1,2], Haroon Naeem[3], Enes Makalic[3], Daniel F. Schmidt[3], James G. Dowty[3], Jihoon E. Joo[4], Chol-Hee Jung[5], Julie K. Bassett[1], Pierre-Antoine Dugue[1], Jessica Chung[5], Andrew Lonie[5], Roger L. Milne[1,3], Ee Ming Wong[4], John L. Hopper[3], Dallas R. English[1,3], Gianluca Severi[1,6,7,8], Laura Baglietto[1,6], John Pedersen[9], Graham G. Giles[1,3], Melissa C. Southey[4].

[1] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC, Australia

[2] Cancer Genetics and Immunology, Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia

[3] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, VIC, Australia

[4] Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Parkville, VIC, Australia

[5] VLSCI Life Sciences Computation Centre, University of Melbourne, Carlton VIC, Australia

[6] Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, France

[7] Gustave Roussy, F-94805, Villejuif, France

[8] HuGeF, Human Genetics Foundation, 10126 Torino, Italy

[9] TissuPath, Mount Waverley, Melbourne, VIC, Australia

Running Title: Genome-wide methylation and prostate cancer risk

Correspondence to: Graham Giles, Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Rd, VIC 3004, Melbourne, Australia, Graham.Giles@cancervic.org.au

Disclosure Statement: The authors declare that they have no affiliations that they consider to be relevant to this study, nor any competing financial interests.

ABSTRACT

Background: Global measures of peripheral blood DNA methylation have been associated with risk of some malignancies, including breast, bladder and gastric cancer. Here, we examined genome-wide measures of peripheral blood DNA methylation in prostate cancer and its non-aggressive and aggressive disease forms.

Methods: We used a matched, case-control study of 687 incident prostate cancer samples, nested within a larger prospective cohort study. DNA methylation was measured in pre-diagnostic, peripheral blood samples using the Illumina Infinium HM450K BeadChip. Genome-wide measures of DNA methylation were computed as the median M-value of all CpG sites and according to CpG site location and regulatory function. We used conditional logistic regression to test for associations between genome-wide measures of DNA methylation and risk of prostate cancer and its subtypes, and by time between blood draw and diagnosis.

Results: We observed no associations between the genome-wide measure of DNA methylation based on all CpG sites and risk of prostate cancer or aggressive disease. Risk of non-aggressive disease was associated with higher methylation of CpG islands (OR=0.80; 95% CI=0.68-0.94), promoter regions (OR=0.79; 95% CI=0.66-0.93) and high density CpG regions (OR=0.80; 95% CI=0.68-0.94). Additionally, higher methylation of all CpGs (OR=0.66; 95% CI=0.48-0.89), CpG shores (OR=0.62; 95% CI=0.45-0.84) and regulatory regions (OR=0.68; 95% CI=0.51-0.91) was associated with a reduced risk of overall prostate cancer within 5 years of blood draw but not thereafter.

Conclusions: A reduced risk of overall prostate cancer within 5 years of blood draw and non-aggressive prostate cancer was associated with higher genome-wide methylation of peripheral blood DNA. While these data have no immediate clinical utility, with further work they may provide insight into the early events of prostate carcinogenesis.

Keywords: Prostate cancer, DNA methylation, peripheral blood, biomarker, HM450K array

INTRODUCTION

The concept and measurement of global methylation has recently been enhanced via the capacity to move from the measurement of surrogate markers, such as repetitive genomic elements, to platforms that achieve genome-wide assessment. Genome-wide changes in DNA methylation that occur prior to cancer diagnosis have now been described by several studies and could provide new information to improve personalised cancer risk prediction models. A study of pre-diagnostic blood samples from the Shanghai Women's Health Study cohort estimated global methylation via bisulphite pyrosequencing of repetitive Alu and long interspersed nucleotide elements (LINE)-1 and found Alu methylation (but not (LINE)-1) was inversely associated with an increased risk of gastric cancer [1]. Using similar technology, Barry *et al.* (2015) estimated DNA methylation at Alu and (LINE)-1 repetitive elements in pre-diagnostic blood samples from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and although there was no evidence for an association with prostate cancer risk overall, hypermethylation of Alu repeats (only) was associated with an increased risk in DNA prepared from blood samples taken at least four years prior to diagnosis [2]. Andreotti *et al.* (2014) also examined pre-diagnostic blood samples from the PLCO and reported increased (LINE)-1 methylation to be associated with increased bladder cancer risk for males, especially in male smokers [3]. In a recent review, Brennan and Flanagan (2012) stressed the limitations of surrogate genome-wide methylation measures and highlighted the importance of examining methylation at specific genomic locations and regulatory regions for prospective studies testing for associations between methylation and disease risk [4]. Since this review, a small number of studies have utilised newer technologies to measure genome-wide methylation, which also enable sub-analyses of methylation measures at specific genomic locations and variously defined regulatory regions. Two recent prospective studies of breast cancer (within the Melbourne Collaborative Cohort Study (MCCS) [5] and the European Prospective Investigation into Cancer and Nutrition (EPIC) study [6]) have utilised the Illumina Infinium HumanMethylation450 (HM450K) BeadChip array and report genome-wide (mean beta values across the CpGs) hypomethylation to be associated with increased breast cancer risk [5,6]. Analysis of functional

subcategories of CpG sites in the MCCS found that genome-wide DNA methylation of genomic regions within functional promoters was associated with increased breast cancer risk and both studies found that hypomethylation of CpG sites outside of promoter regions (e.g. CpG shelves, non-regulatory regions, gene bodies) was associated with decreased breast cancer risk [5,6]. These results were also supported by whole-genome bisulphite sequencing data from the prospective Breakthrough Generations Study (BGS), however, using the HM450K BeadChip, the Norwegian Women and Cancer (NOWAC) study provided no evidence for an association between breast cancer risk and genome-wide measures of methylation [6]. A nested MCCS study of mature B-cell neoplasms (MBCN) reported genome-wide hypomethylation calculated from HM450K BeadChip measures to be associated with an increased risk of this disease [7]. Interestingly, analysis of functional subcategories of CpG sites demonstrated an association between hypermethylation of promoter regions and increased risk of MBCN, which was also observed in the MCCS breast cancer study [5].

Here, we report a nested case-control study of 1,374 male participants from the MCCS. Using the Illumina Infinium HM450K BeadChip, we derived genome-wide measures of methylation from blood DNA to test the hypothesis that differential global methylation of DNA from peripheral blood is associated with overall prostate cancer risk or its subtypes defined by aggressive features.

MATERIALS AND METHODS

Participants

Between 1990 and 1994, 41,514 healthy adult volunteers (17,045 men) aged between 27 and 76 years (99% aged 40-69) were recruited into the prospective MCCS [8]. Men with a history of any cancer prior to enrolment, apart from non-melanocytic skin cancer, and men without a dried blood spot baseline sample were excluded. Incident cases of invasive (including metastatic) adenocarcinoma of the prostate were identified up to 31 December 2010 by linkage with the

centralised Victorian Cancer Registry, which receives mandatory notification of all new cancer cases in Victoria, and the Australian Cancer Database maintained by the Australian Institute of Health and Welfare in Canberra. Aggressive cases were defined as: Gleason score 8-10; tumour stage T4, N+ or M+; and/or, suffered fatal prostate cancer. Male controls from the MCCS were individually matched to cases on year of birth, country of birth and DNA source using density sampling with age as the time scale.

Study participants provided written, informed consent in accordance with the Declaration of Helsinki. The study was approved by Cancer Council Victoria's Human Research Ethics Committee and performed in accordance with the institution's ethical guidelines.

DNA source and extraction

Peripheral blood specimens were collected at the time of study enrolment (baseline) in the form of peripheral blood mononuclear cells (PBMC), buffy coats or dried blood spots collected onto filter paper (Whatman, United Kingdom). DNA was prepared from PBMC and buffy coat specimens using QIAamp mini spin columns (Qiagen, Germany). DNA was prepared from twenty-one 3.2mm diameter punches from the dried blood spots after lysis in phosphate buffered saline using a TissueLyser (Qiagen, Germany). The supernatant was processed using Qiagen mini spin columns according to the manufacturer's protocol as described previously [9].

DNA bisulfite conversion and hybridisation

Bisulfite conversion was performed on a minimum of 0.3µg DNA (assessed using the Quant-iT™ Picogreen® dsDNA assay measured on the Qubit® Fluorometer (Life Technologies, USA)), using the Zymo Gold single tube kit according to manufacturer's instructions (Zymo Research, USA). Post-conversion quality control was performed using SYBR Green-based quantitative PCR, designed to determine the success of bisulfite conversion by comparing amplification of the test sample with unconverted and converted high-quality DNA controls. Samples with late amplification (Cp>5 compared to high-quality converted DNA) were re-extracted and re-converted, if feasible.

Samples were processed in batches of 96 (8 HM450K BeadChips per batch). In order to minimise potential batch effects, matched cases and controls were processed together and run on the same BeadChip and cancer subtypes (non-aggressive and aggressive) were evenly distributed across the chips and plates. Paired cases and controls were randomly positioned on each BeadChip to reduce any possible position effects within chips. Two pairs of technical replicates and a reference duplicate (DNA prepared from the multiple myeloma cell line U266) were included on each plate. A total of 200ng of bisulfite converted DNA was whole-genome amplified and hybridised onto the BeadChips according to the manufacturer's instructions. The TECAN automated liquid handler (Tecan Group Ltd, Switzerland) was used for the single-base extension and BeadChip staining steps.

HM450K BeadChip pre-processing and quality control

Raw intensity signals were imported into R programming software (www.r-project.org) using the *minfi* Bioconductor package [10]. Data were pre-processed and normalised to control probes using the "preprocessIllumina" *minfi* function. Subset-quantile within array normalization (SWAN) was performed to correct for type I and II probe bias [11]. Samples were excluded if >5% of CpG probes (excluding chromosome X and Y probes) had a detection p-value >0.01, while CpG probes were excluded from further analysis if they had missing values in >20% of samples. In addition, a total of 240,050 CpG sites were excluded as they were considered to be affected by genomic factors (i.e., single nucleotide polymorphisms, small insertions and deletions, repetitive DNA regions and regions with reduced genomic complexity) [12]. Finally to minimize chip and batch effects, ComBat normalisation was applied to the remaining 245,462 CpG sites [13].

Statistical Analyses

Statistical analyses were performed using median M-values calculated using *minfi* [10], where the M-value was defined as log_2 *(Meth/Unmeth),* and *Meth* and *Unmeth* were the intensities of the methylated and unmethylated probes, respectively.

DNA methylation was defined as overall (i.e. the median M-value across all 245,462 CpGs) and as categories of CpG sites based on their genomic location and/or regulatory function (e.g. the median M-value of CpGs located in CpG islands). These CpG categories were defined using the Illumina annotation file v1.2 and included: CpG islands, shores or shelves; and promoter, regulatory, non-regulatory and gene body regions. Promoter regions were further stratified by their CpG content and ratio, as differential CpG content within promoter regions is known to influence the methylation profile and gene expression [14,15]. High density CpG promoters, intermediate density CpG promoters, and low density CpG promoters were classified using a published annotation file [16].

Principal components analysis was performed on all samples and CpGs that passed quality control criteria to determine if specimen source (PBMC, buffy coat or Guthrie card) affected methylation profiles (**Supplementary Figure I**). Due to distinct clustering by specimen source and the fact that the majority of specimens were from Guthrie cards, only DNA methylation measurements from Guthrie cards were included in subsequent analyses.

Associations between DNA methylation and overall, non-aggressive and aggressive prostate cancer risk were assessed using conditional logistic regression to compute odds ratios (OR) per standard deviation of genome-wide methylation measure and 95% confidence intervals (CI). The effect of time interval since blood draw on the association between methylation and prostate cancer risk was also analysed using conditional logistic regression where case-control pairs were categorised considering <5 vs. ≥5-year time intervals between blood draw and the case's diagnosis. For all analyses, DNA methylation measures were treated as continuous and statistical significance was determined using the likelihood ratio test (LRT). Tests for heterogeneity were implemented as tests for interactions. As a sensitivity analysis to determine the effect of excluding 240,050 CpGs, all of the above analyses were repeated with all 485,512 CpGs.

RESULTS

Between baseline and 31 December 2010, 1,464 male MCCS participants were diagnosed with adenocarcinoma of the prostate, of whom 442 were diagnosed with aggressive disease. Three hundred and twenty-nine aggressive cases had an available Guthrie card DNA sample and were assayed using the HM450K BeadChip along with matched controls and a random sample of 359 non-aggressive case-control sets. One non-aggressive case-control set, which failed quality control, and 240,050 CpGs described in Naeem *et al*. (2014) [12] were excluded from the final analyses (**Supplementary Figure II**). The clinical characteristics of cases are summarised in **Table I**.

No associations were observed between risk of overall prostate cancer and genome-wide measures of methylation, either for all CpGs or categories of CpG sites based on location or regulatory function (**Table II**). When cases were stratified by aggressiveness, several significant associations were observed with non-aggressive prostate cancer. Hypermethylation of CpG sites within CpG islands (OR=0.80; 95% CI=0.68-0.94), promoter (OR=0.79; 95% CI=0.66-0.93), high density CpG promoter regions (OR=0.80; 95% CI=0.68-0.94), and other regulatory regions (OR=0.84; 95% CI=0.72-0.98) was associated with a reduced risk of non-aggressive disease (**Table II**). With the exception of methylation in other regulatory regions, these associations were significantly different (p_{het}<0.02) from those observed for aggressive cases, where no associations were observed. Supplemental case-case analyses also highlighted CpG islands, promoter and high density CpG promoter regions as having significantly different methylation patterns between aggressive and non-aggressive cases (**Supplementary Table I**).

When participants were stratified by time since blood draw, hypermethylation of all CpGs (OR=0.66; 95% CI=0.48-0.89) and hypermethylation of CpG sites within CpG island shores (OR=0.62; 95% CI=0.45-0.84) and other regulatory regions (OR=0.68; 95% CI=0.51-0.91) were associated with a reduced risk of overall prostate cancer within five years of blood draw (**Table III**). These results were significantly different to those for overall prostate cancer risk five or more years after blood draw (all p<0.03), where no associations were observed.

To investigate whether the exclusion of probes that are predicted to be affected by genomic factors influenced the results, analyses were repeated with all 485,512 probes. Similar results were observed between the analyses of the filtered probe set and the full probe set across all categories of global methylation and disease subtype stratifications, except in the time since blood draw analyses where no association was observed between global DNA hypermethylation and risk of prostate cancer within five years of blood draw (**Supplementary Tables I - III**).

DISCUSSION

In some contexts, methylation profiling of DNA from peripheral blood indicates promise for cancer risk prediction, early detection and prognosis [1-5,17,18]. Several previous studies have shown an association between measures of methylation and cancer risk using both surrogate global measures and, more recently, high-density genome-wide measures [1-3,5-7]. While it is not clear yet why genome-wide methylation patterns are altered in pre-diagnostic, peripheral blood DNA, it is likely to be a combination of influences, including altered epigenetic regulation, immunological changes from inherited/acquired genetic variants or accumulated exposures, or result from the very early stages of carcinogenesis.

Here, we used the Illumina Infinium HM450K BeadChip to assess whether genome-wide DNA methylation patterns from peripheral blood are associated with risk of prostate cancer, with specific interest in risk of non-aggressive and aggressive disease. Application of the HM450K BeadChip provided us with the capacity to perform more sophisticated analyses (e.g. the ability to conduct sub-analyses of predefined groups of measured CpGs) but it also distinguishes our work from the prior literature that has utilised surrogate markers of methylation, such as Alu and (LINE)-1 repetitive elements. In a prior study by Barry and colleagues (2015), higher Alu methylation was found to be associated with an increase in prostate cancer risk when diagnosis was four or more years after blood draw [2]. This is in contrast to our findings, where we observed genome-wide hypermethylation (and hypermethylation of CpG shores and other regulatory regions) to be

associated with a reduced risk of prostate cancer within five years of blood draw and not later. A limited number of other studies, applying a range of methodological approaches, have also noted differences in DNA methylation patterns depending on the time between blood draw and diagnosis [1,2,5]. Examining methylation of repetitive elements, Gao and colleagues (2012) observed an association between hypomethylation and decreased bladder cancer risk when diagnosis was within 1 year of blood draw, and an association between hypomethylation and increased cancer risk when diagnosis was a year or more after blood draw [1]. In a study of breast cancer risk, using genome-wide methylation measurements derived from the HM450K BeadChip, we found a difference in genome-wide DNA methylation between cases and controls that decreased with increasing time since blood draw, where the difference was negligible after 5 years [5]. Some of the apparent contradictions in the above-mentioned studies could be due to methodological differences and more work is required to understand the relationship between these different global methylation measures.

It has been well documented that a number of probes included on the HM450K array do not perform well technically. During our analyses, we excluded CpG probes that Naeem *et al.* (2014) have suggested to be affected by genomic factors [12]. Their study compared HM450K methylation data to whole-genome bisulfite sequencing data and observed that the methylation status of particular probes can be affected by being able to hybridise to multiple locations or repetitive regions in the genome, by binding to regions containing insertions or deletions (especially relevant in population-wide studies such as ours) and by the presence of a genetic polymorphism either within the CpG site itself or within the probe-binding region [12]. Naeem *et al.* demonstrated that removing the measurements of these probes from their analyses resulted in a set of high-quality probes that reduced false-positive signals, decreased within tissue standard deviation, yet still provided adequate genome-wide coverage, with only the HLA region on chromosome 6 being devoid of probes. Taking this approach, we removed 240,050 CpG probes from the analyses reported here but also repeated the analyses with the complete probe dataset. These analyses yielded

very similar results except that the association between genome-wide DNA methylation measures and risk of prostate cancer within five years of blood draw did not persist. It is possible that when considering global methylation, the removal of a select number of technically deficient probes may not have a significant effect on the outcome of analyses provided there is still adequate genome coverage.

We consider the design of our prospective study to be a strength. Detailed information collected at blood draw allowed us to carefully match cases and controls on age, ethnicity and DNA source, and also allowed us to stratify cases according to clinical disease aggressiveness. Additionally, potential batch effects were mitigated by placing matched cases and controls adjacent to each other, but in a random position, on the same chip, resulting in minimal technical bias [19].

One shortcoming of our study is its lack of replication. To our knowledge, there are currently no other prostate cancer studies with HM450K methylation data available for pre-diagnostic blood DNA samples. It is important that other cohorts examine associations between genome-wide DNA methylation and prostate cancer risk, and for data pooling to occur in order to sufficiently investigate the individual CpG components of the genome-wide methylation measure.

CONCLUSION

Our study of genome-wide measures of methylation in peripheral blood DNA provides no evidence of association with overall prostate cancer risk. The data provide evidence that methylation of peripheral blood DNA is associated with non-aggressive disease; the hypomethylation of CpG islands, promoter and other regulatory regions being associated with an increased risk of indolent disease. This is interesting, as few of the common genetic risk factors for prostate cancer, identified through genome-wide association studies, distinguish aggressive from non-aggressive disease [20,21], suggesting that the underlying driver(s) of these phenotypes are still yet to be identified. We did detect an association with risk of overall prostate cancer when measured within five years of

diagnosis but we had inadequate statistical power to stratify our analysis more substantially. While these data have no immediate clinical utility, they may provide some insight regarding early events in prostate carcinogenesis. It is important for this work to be extended, both to confirm these findings and also to accumulate data sufficient to conduct association studies at the resolution of individual CpGs. The findings of this study do not exclude the possibility that some individual CpGs (or groups of CpGs) within this dataset are associated with overall prostate cancer risk or its indolent and aggressive subtypes.

ACKNOWLEDGMENTS

This study was made possible by the contribution of many people, including the original investigators and the diligent team who recruited the participants and who continue working on follow-up. We would also like to express our gratitude to the many thousands of Melbourne residents who took part in the study and provided blood samples.

FUNDING

This work was supported by the National Health and Medical Research Council project grant number 1026892.

References

1. Gao Y, Baccarelli A, Shu XO, Ji BT, Yu K, Tarantini L, Yang G, Li HL, Hou L, Rothman N, Zheng W, Gao YT, Chow WH. Blood leukocyte Alu and LINE-1 methylation and gastric cancer risk in the Shanghai Women's Health Study. British Journal of Cancer 2012;106(3):585-591.
2. Barry KH, Moore LE, Liao LM, Huang WY, Andreotti G, Poulin M, Berndt SI. Prospective study of DNA methylation at LINE-1 and Alu in peripheral blood and the risk of prostate cancer. The Prostate 2015;75(15):1718-1725.

3. Andreotti G, Karami S, Pfeiffer RM, Hurwitz L, Liao LM, Weinstein SJ, Albanes D, Virtamo J, Silverman DT, Rothman N, Moore LE. LINE1 methylation levels associated with increased bladder cancer risk in pre-diagnostic blood DNA among US (PLCO) and European (ATBC) cohort study participants. Epigenetics 2014;9(3):404-415.
4. Brennan K, Flanagan JM. Is there a link between genome-wide hypomethylation in blood and cancer risk? Cancer Prevention Research 2012;5(12):1345-1357.
5. Severi G, Southey MC, English DR, Jung CH, Lonie A, McLean C, Tsimiklis H, Hopper JL, Giles GG, Baglietto L. Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. Breast Cancer Research and Treatment 2014;148(3):665-673.
6. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, Mattiello A, Palli D, Masala G, Krogh V, Agnoli C, Tumino R, Frasca G, Flower K, Curry E, Orr N, Tomczyk K, Jones ME, Ashworth A, Swerdlow A, Chadeau-Hyam M, Lund E, Garcia-Closas M, Sandanger TM, Flanagan JM, Vineis P. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clinical Epigenetics 2015;7(1):67.
7. Wong Doo N, Makalic E, Joo JE, Vajdic CM, Schmidt DF, Wong EM, Jung CH, Severi G, Park DJ, Chung J, Baglietto L, Prince HM, Seymour JF, Tam C, Hopper JL, English DR, Milne RL, Harrison SJ, Southey MC, Giles GG. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. Epigenomics 2016;8(1):55-66.
8. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC Scientific Publications 2002;156(69-70.
9. Joo JE, Wong EM, Baglietto L, Jung CH, Tsimiklis H, Park DJ, Wong NC, English DR, Hopper JL, Severi G, Giles GG, Southey MC. The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. BMC Biotechnology 2013;13(23.

10. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30(10):1363-1369.

11. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biology 2012;13(6):R44.

12. Naeem H, Wong NC, Chatterton Z, Hong MK, Pedersen JS, Corcoran NM, Hovens CM, Macintyre G. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. BMC Genomics 2014;15(51.

13. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118-127.

14. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics 2007;39(4):457-466.

15. Martin-Subero JI, Kreuz M, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E, Rosolowski M, Richter J, Lopez-Serra L, Ballestar E, Berger H, Agirre X, Bernd HW, Calvanese V, Cogliatti SB, Drexler HG, Fan JB, Fraga MF, Hansmann ML, Hummel M, Klapper W, Korn B, Kuppers R, Macleod RA, Moller P, Ott G, Pott C, Prosper F, Rosenwald A, Schwaenen C, Schubeler D, Seifert M, Sturzenhofecker B, Weber M, Wessendorf S, Loeffler M, Trumper L, Stein H, Spang R, Esteller M, Barker D, Hasenclever D, Siebert R. New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. Blood 2009;113(11):2488-2497.

16. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, Robinson WP, Kobor MS. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & Chromatin 2013;6(1):4.

17. Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. Journal of Clinical Oncology 2011;29(9):1133-1139.

18. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M. An epigenetic signature in peripheral blood predicts active ovarian cancer. PloS One 2009;4(12):e8274.

19. Harper KN, Peters BA, Gamble MV. Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. Cancer Epidemiology, Biomarkers & Prevention 2013;22(6):1052-1060.

20. Berndt SI, Wang Z, Yeager M, Alavanja MC, Albanes D, Amundadottir L, Andriole G, Beane Freeman L, Campa D, Cancel-Tassin G, Canzian F, Cornu JN, Cussenot O, Diver WR, Gapstur SM, Gronberg H, Haiman CA, Henderson B, Hutchinson A, Hunter DJ, Key TJ, Kolb S, Koutros S, Kraft P, Le Marchand L, Lindstrom S, Machiela MJ, Ostrander EA, Riboli E, Schumacher F, Siddiq A, Stanford JL, Stevens VL, Travis RC, Tsilidis KK, Virtamo J, Weinstein S, Wilkund F, Xu J, Lilly Zheng S, Yu K, Wheeler W, Zhang H, Sampson J, Black A, Jacobs K, Hoover RN, Tucker M, Chanock SJ. Two susceptibility loci identified for prostate cancer aggressiveness. Nature Communications 2015;6(6889.

21. FitzGerald LM, Kwon EM, Conomos MP, Kolb S, Holt SK, Levine D, Feng Z, Ostrander EA, Stanford JL. Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer. Cancer Epidemiology, Biomarkers & Prevention 2011;20(6):1196-1203.

Supplementary Figure I: Principal components analysis plot demonstrating the variance in methylation data due to specimen source. BC: DNA extracted from stored buffy coat, dark green circles; DNA BC: Specimen from stored buffy coat DNA, orange circles; DNA PBMC; Specimen from stored peripheral blood mononuclear cell (PBMC) DNA, purple circles; GC: DNA extracted

from stored Guthrie card, pink circles; and PBMC: DNA extracted from stored PBMC samples, light green circles.

Supplementary Figure II: Flow diagram depicting the selection of MCCS prostate cancer cases and controls for whole-genome methylation analyses.

Table I: Summary of clinical characteristics for non-aggressive and aggressive prostate cancer cases.

	Non-Aggressive Cases	Aggressive[1] Cases
	N = 358	N = 329
Age at Diagnosis (mean, SD)	67.3 (7.1)	69.0 (7.0)
Ethnicity		
United Kingdom (Australian + New Zealand)	313 (87.4%)	284 (86.3%)
Greece	19 (5.3%)	17 (5.2%)
Italy	26 (7.3%)	28 (8.5%)
Gleason Score		
2-7	340 (95%)	138 (42%)
8-10	0 (0%)	158 (48%)
Missing	18 (5%)	33 (10%)
Grade		
Well differentiated	25 (7%)	4 (1.2%)
Moderately differentiated	99 (27.7%)	19 (5.8%)
Poorly differentiated	0 (0%)	168 (51.1%)
Undifferentiated	0 (0%)	2 (0.6%)
Missing	234 (65.3%)	136 (41.3%)
Primary tumour stage (TNMT)		
TX	1 (0.3%)	22 (6.8%)
T1 (1A-1C)	248 (69.3%)	210 (63.8%)
T2 (2A-2C)	86 (24%)	55 (16.7%)
T3 (3A-3B)	23 (6.4%)	37 (11.2%)
T4	0 (0%)	5 (1.5%)
Regional lymph nodes tumour stage (TNMN)		
NX	299 (83.5%)	267 (81.2%)
N0	59 (16.5%)	53 (16.1%)
N1	0 (0%)	9 (2.7%)
Distant metastases tumour stage (TNMM)		
M0	358 (100%)	312 (94.8%)
M1 (1A-1C)	0 (0%)	17 (5.2%)
Vital status		
Alive	309 (86.3%)	169 (51.4%)
Other cause of death	49 (13.7%)	38 (11.6%)
Prostate cancer-specific death	0 (0%)	122 (37.0%)

[1]Aggressive disease is defined as: Gleason score 8-10; tumour stage T4, N+ or M+; and/or died of prostate cancer-specific causes.

Table II: Associations between genome-wide measures of DNA methylation and prostate cancer risk

	All cases vs. controls (n=1,374)		Non-aggressive cases vs. controls (n=716)		Aggressive cases vs. controls (n=658)		
CpG Category	**OR[2] (95% CI)[3]**	**LRT[4] P-value**	**OR (95% CI)**	**LRT P-value**	**OR (95% CI)**	**LRT P-value**	**P_{het}[5]**
Genome-wide							
All CpG sites[1]	0.94 (0.84-1.05)	0.26	0.93 (0.80-1.09)	0.39	0.94 (0.80-1.10)	0.4646	0.92
CpG Region							

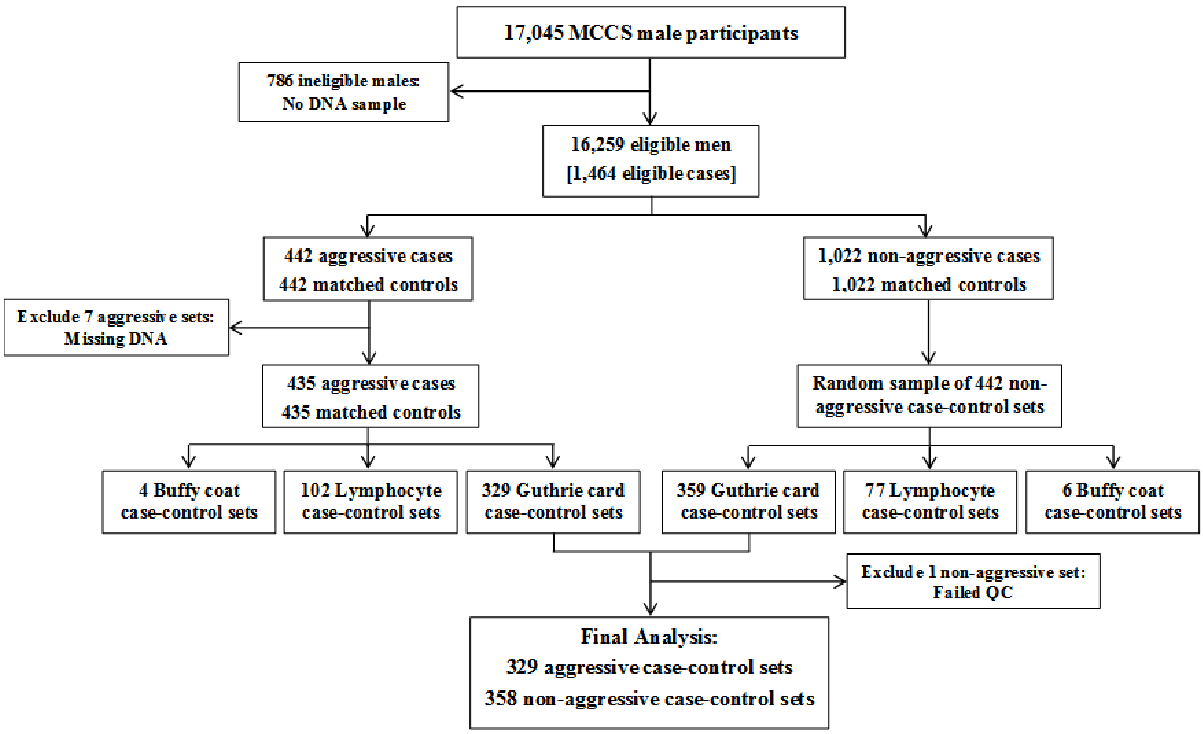
CpG island	0.92 (0.82-1.03)	0.13	**0.80 (0.68-0.94)**	**0.006**	1.07 (0.91-1.26)	0.4325	**0.02**
CpG shore	0.93 (0.83-1.04)	0.18	0.89 (0.76-1.04)	0.13	0.97 (0.83-1.13)	0.6982	0.42
CpG shelf	1.03 (0.91-1.15)	0.67	1.12 (0.96-1.31)	0.15	0.92 (0.77-1.09)	0.3244	0.09
Regulatory Region							
Promoter	0.92 (0.81-1.03)	0.13	**0.79 (0.66-0.93)**	**0.004**	1.08 (0.91-1.27)	0.3943	**0.009**
Other regulatory	0.91 (0.82-1.02)	0.09	**0.84 (0.72-0.98)**	**0.02**	1.00 (0.86-1.16)	0.9664	0.12
Non-regulatory	1.03 (0.92-1.15)	0.61	1.12 (0.96-1.31)	0.13	0.92 (0.78-1.09)	0.3510	0.09
Gene body	1.03 (0.92-1.15)	0.62	1.12 (0.96-1.31)	0.13	0.92 (0.78-1.09)	0.3507	0.09
Promoter Regions							
High density CpG	0.92 (0.82-1.03)	0.14	**0.80 (0.68-0.94)**	**0.005**	1.08 (0.91-1.27)	0.3883	**0.01**
Intermediate density CpG	1.02 (0.91-1.14)	0.75	1.13 (0.97-1.32)	0.13	0.89 (0.75-1.07)	0.2126	0.05
Low density CpG	1.03 (0.92-1.15)	0.58	1.11 (0.95-1.29)	0.18	0.94 (0.80-1.12)	0.4906	0.16

[1]Includes 245,462 probes after quality control filters and filtering based on Naeem *et al.* (2014); [2]OR: Odds ratio; [3]95% CI: 95% Confidence Intervals; [4]LRT: likelihood ratio test; [5]Phet: heterogeneity p-value

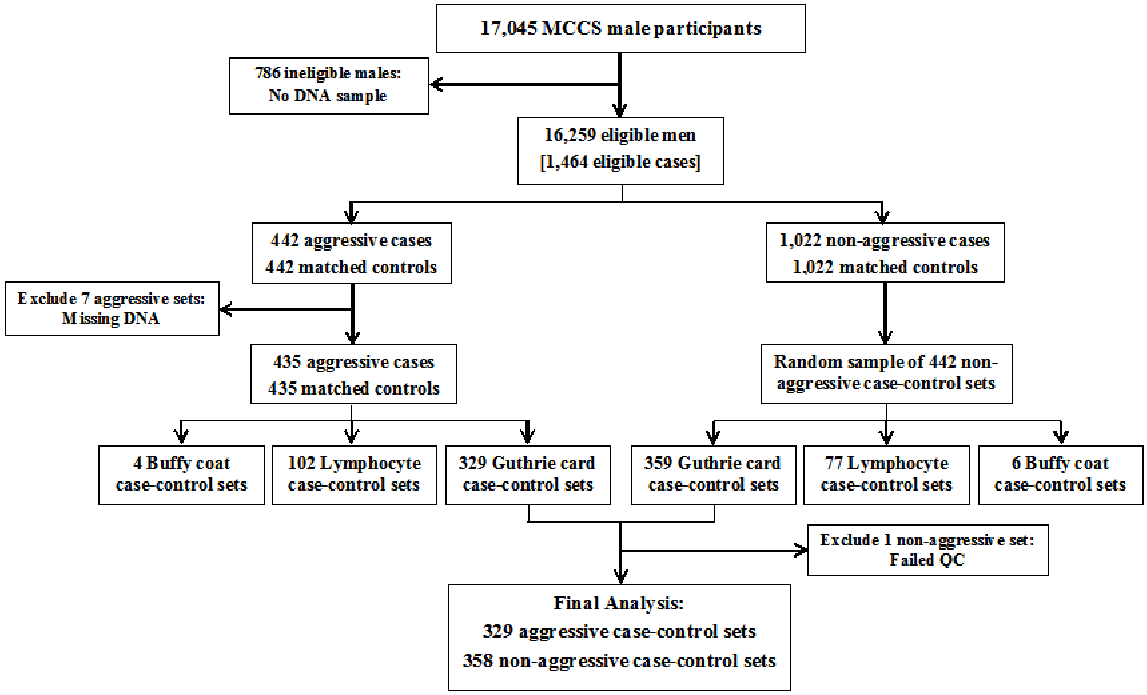
Table III: Associations between genome-wide measures of DNA methylation and prostate cancer risk by time between blood collection and diagnosis

	Time Interval (< 5 years)		Time Interval (>= 5 years)		
	Cases vs. controls (n = 240)		Cases vs. controls (n = 1132)		
CpG Category	**OR[2] (95% CI)[3]**	**LRT[4] P-value**	**OR (95% CI)**	**LRT P-value**	**P_{het}[5]**
Genome-wide					
All CpG sites[1]	**0.66 (0.48-0.89)**	**0.005**	1.00 (0.88-1.13)	0.97	**0.01**
CpG Region					
CpG island	0.78 (0.58-1.04)	0.08	0.94 (0.83-1.07)	0.35	0.21
CpG shore	**0.62 (0.45-0.84)**	**0.001**	0.99 (0.88-1.11)	0.87	**0.003**
CpG shelf	1.07 (0.81-1.42)	0.63	1.02 (0.90-1.15)	0.80	0.74
Regulatory Region					
Promoter	0.81 (0.61-1.08)	0.14	0.94 (0.82-1.07)	0.33	0.35
Other regulatory	**0.68 (0.51-0.91)**	**0.007**	0.96 (0.85-1.08)	0.46	**0.02**
Non-regulatory	1.06 (0.80-1.41)	0.69	1.02 (0.91-1.16)	0.70	0.83
Gene body	1.04 (0.78-1.38)	0.80	1.03 (0.91-1.16)	0.66	0.95
Promoter Regions					
High density CpG	0.79 (0.59-1.06)	0.10	0.94 (0.83-1.07)	0.36	0.25
Intermediate density CpG	0.96 (0.72-1.29)	0.80	1.03 (0.91-1.17)	0.64	0.68
Low density CpG	1.10 (0.83-1.46)	0.49	1.02 (0.90-1.15)	0.77	0.61

[1]Includes 245,462 probes after quality control filters and filtering based on Naeem *et al.* (2014); [2]OR: Odds ratio; [3]95% CI: 95% Confidence Intervals; [4]LRT: likelihood ratio test; [5]Phet: heterogeneity p-value



FitzGerald et al Supplementary Figure 2 .



Supplementary Figure 2: Flow diagram depicting the selection of MCCS prostate cancer cases and controls for whole-genome methylation analyses.

FitzGerald et al Supplementary Figure 2 title .