

Keywords: DNA methylation biomarker; myeloid cell; lymphoid cell; inflammatory disease

Similar blood-borne DNA methylation alterations in cancer and inflammatory diseases determined by subpopulation shifts in peripheral leukocytes

H Li¹, T Zheng¹, B Chen², G Hong¹, W Zhang¹, T Shi¹, S Li¹, L Ao^{2,3}, C Wang² and Z Guo^{*1,2,3}

¹Bioinformatics Centre, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China; ²College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China and ³Department of Bioinformatics, School of Basic Medical College, Fujian Medical University, Fuzhou 350004, China

Background: Although many DNA methylation (DNAm) alterations observed in peripheral whole blood/leukocytes and serum have been considered as potential diagnostic markers for cancer, their origin and their specificity for cancer (e.g., vs inflammatory diseases) remain unclear.

Methods: From publicly available datasets, we identified changes in the methylation of blood-borne DNA for multiple cancers and inflammatory diseases. We compared the identified changes with DNAm difference between myeloid and lymphoid cells extracted from two datasets.

Results: At least 94.7% of the differentially methylated DNA loci (DM loci) observed in peripheral whole blood/leukocytes and serum of cancer patients overlapped with DM loci that distinguish between myeloid and lymphoid cells and >99.9% of the overlapped DM loci had consistent alteration states (hyper- or hypomethylation) in cancer samples compared to normal controls with those in myeloid cells compared to lymphoid cells (binomial test, P -value $< 2.2 \times 10^{-16}$). Similar results were observed for DM loci in peripheral whole blood/leukocytes in patients with rheumatoid arthritis or inflammatory bowel diseases. The direct comparison between DM loci observed in the peripheral whole blood/leukocytes of patients with inflammatory diseases and DM loci observed in the peripheral whole blood of patients with cancer showed that DM loci detected from cancer and inflammatory diseases also had significantly consistent alteration states (binomial test, P -value $< 2.2 \times 10^{-16}$).

Conclusions: DNAm changes observed in the peripheral whole blood/leukocytes and serum of cancer patients and in the peripheral whole blood/leukocytes of inflammatory disease patients are predominantly determined by the increase of myeloid cells and the decrease of lymphoid cells under the disease conditions, in the sense that their alteration states in disease samples compared to normal controls mainly reflect the DNAm difference between myeloid and lymphoid cells. These analyses highlight the importance of comparing cancer and inflammatory disease directly for the identification of cancer-specific diagnostic biomarkers.

DNA methylation (DNAm) is an important epigenetic component of carcinogenesis (Robertson, 2005). Historically, DNAm has been measured in various cancer tissues to understand cancer pathogenesis (Christensen *et al*, 2010), evaluate patient responses

to therapeutic interventions (Maier *et al*, 2005), and find cancer diagnostic biomarkers (Brock *et al*, 2008). Because blood sampling is less invasive and easy to handle (Lonneborg *et al*, 2009; Li *et al*, 2012), investigators have attempted to identify diagnostic DNAm

*Correspondence: Professor Z Guo; E-mail: guoz@ems.hrbmu.edu.cn

Received 13 February 2014; revised 12 April 2014; accepted 15 May 2014; published online 24 June 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14

alterations in peripheral whole blood/leukocytes (Teschendorff *et al*, 2009; Marsit *et al*, 2011; Langevin *et al*, 2012) and serum (Fujiwara *et al*, 2005; Wang *et al*, 2010; Pedersen *et al*, 2011; Garcia-Closas *et al*, 2013) from patients with cancer such as ovarian (Teschendorff *et al*, 2009), breast (Garcia-Closas *et al*, 2013) and head and neck squamous cell carcinoma (HNSCC; Langevin *et al*, 2012). Although some DNAm alterations can distinguish cancers from normal controls (Teschendorff *et al*, 2009; Marsit *et al*, 2011; Langevin *et al*, 2012), it is currently still unclear if these DNAm modifications can distinguish cancers from other diseases, especially inflammatory diseases.

A change commonly observed in peripheral leukocytes of cancer patients is an increase in the number of myeloid cells and a decrease in the number of lymphoid cells (Kuss *et al*, 2004; Cho *et al*, 2009; Accomando *et al*, 2012; Houseman *et al*, 2012). As different leukocyte subtypes have different DNAm patterns, shifts in leukocyte subpopulations can lead to DNAm alterations in the peripheral whole blood of cancer patients (Koestler *et al*, 2012; Reinius *et al*, 2012). However, how and to what extent the population shifts of leukocyte cells could contribute to the DNAm alterations in the peripheral whole blood/leukocytes of cancer patients are still unclear. More importantly, similar changes in subpopulations of peripheral leukocytes also have been observed in various inflammatory diseases such as inflammatory bowel disease (IBD; Hanai *et al*, 2004), chronic hepatitis (Zhang *et al*, 2011), chronic obstructive pulmonary disease (Domagala-Kulawik *et al*, 2007) and rheumatoid arthritis (RA; Liu *et al*, 2013). This suggests that the blood-borne DNAm alterations observed in inflammatory diseases, similarly as in cancer, could also be affected by population shifts in leukocyte subpopulations. This similarity may confound the use of blood-borne DNAm alterations in cancer patients as cancer-specific diagnostic markers because such markers may not be able to distinguish inflammatory diseases from cancer.

In serum, researchers often attempt to identify cancer-specific diagnostic biomarkers by testing candidate aberrant DNAm modifications previously found to be hypermethylated in cancer tissues (Ichikawa *et al*, 2004; Fujiwara *et al*, 2005; Van De Voorde *et al*, 2012), hypothesising that serum exhibits signals of circulating cell-free DNA released from necrotic tumour cells (Lonneborg *et al*, 2009). However, especially in early cancer stages, limited DNAm signals originating from cancer tissues may be masked by DNAm signals released from peripheral leukocytes that tend to undergo apoptosis more in cancer as well as in inflammatory diseases (Courtney *et al*, 1999; Saito *et al*, 2000; Yoshikawa *et al*, 2008; Saito *et al*, 2013). Thus, DNAm alterations in the serum of patients with cancer or inflammatory diseases could also be

affected by population shifts in leukocyte subpopulations, confounding the use of DNAm alterations in serum as cancer-specific diagnostic markers.

Considering that DNAm patterns of leukocytes originating from a common (myeloid or lymphoid) progenitor tend to be similar whereas DNAm patterns of leukocytes originating from different progenitors tend to be different (Reinius *et al*, 2012) and that the proportion of myeloid cells tends to increase while the proportion of lymphoid cells tends to decrease in both cancer and inflammatory diseases (Houseman *et al*, 2012; Liu *et al*, 2013), we hypothesised that DNAm changes observed in the peripheral whole blood/leukocytes and serum of patients with cancer or inflammatory diseases would reflect the differential DNAm between myeloid and lymphoid cells. In this paper, using publicly available DNAm profiles of peripheral whole blood/leukocytes from patients with various cancers (HNSCC, ovarian, small-cell lung cancer (SCLC), and pancreatic cancer) and various inflammatory diseases (RA and IBD), we revealed that almost all alteration states (hyper- or hypomethylation) of differentially methylated DNA loci (DM loci) observed for cancer and inflammatory diseases compared to normal controls were consistent with those in myeloid cells compared to lymphoid cells. Furthermore, we also showed that almost all of the alteration states of DM loci observed in the serum of cancer patients were consistent with those in myeloid cells compared to lymphoid cells. In this sense, DNAm changes in the peripheral whole blood/leukocytes of patients with cancer or inflammatory diseases, as well as in cancer serum, are predominantly determined by population shifts in the myeloid and lymphoid cells. Finally, we report that DNAm alterations observed in the peripheral whole blood/leukocytes of patients with inflammatory diseases were consistent with those observed in the peripheral whole blood of cancer patients. Our analyses suggest that current studies for finding blood-borne DNAm biomarkers between cancer patients and normal controls might provide little information on the identification of cancer-specific diagnosis biomarkers, highlighting the importance of comparing cancer and inflammatory disease directly.

MATERIALS AND METHODS

Data sources. All genome-wide DNAm profile datasets (Table 1) were generated using either the Illumina 450k Human DNA methylation platform (Illumina, San Diego, CA, USA) or the Human Methylation 27 Bead Array platform (Illumina). Datasets were downloaded from the Gene Expression Omnibus (GEO)

Table 1. Datasets analysed in this study

Dataset ^a	Sample size		Platform ^b	GEO accession ID	Reference
N_MTL1	46		27k	GSE39981	Accomando <i>et al</i> , 2012
N_MTL2	48		450k	GSE35069	Reinius <i>et al</i> , 2012
HNSCC_PB	Case 92	Control 92	27k	GSE30229	Langevin <i>et al</i> , 2012
HNSCC_SR	12	12	450k	GSE40005	
OVC_PB	131	274	27k	GSE19711	Teschendorff <i>et al</i> , 2009
RA_PL	354	335	450k	GSE42861	Liu <i>et al</i> , 2013

Abbreviations: GEO = Gene Expression Omnibus database; HNSCC = head and neck squamous cell carcinoma; MTL = myeloid and lymphoid cells; N = normal; OVC = ovarian cancer; PB = peripheral whole blood; RA = rheumatoid arthritis; PL = peripheral leukocytes; SR = serum.

^aEach dataset is denoted by phenotype followed by sample type.

^b27k represents the Human Methylation 27 Bead Array platform and 450k represents the Illumina 450k Human DNA methylation platform.

database (Barrett *et al*, 2009). In all, 25 978 CpG loci located within the proximal promoter regions of the transcription start sites of 14 113 genes were measured in both platforms, which were used as background CpG loci in this study.

Specifically, in the N_MTL1 dataset, DNAm profiles of normal human leukocyte subtypes were divided into two groups: one group was composed of DNAm profiles of myeloid cells including monocytes ($n=5$), granulocytes ($n=4$) and neutrophils ($n=4$) and the second group was composed of DNAm profiles of lymphoid cells including B cells ($n=5$), NK cells (Pan NKR cells, CD16 + NK cells, CD16 – NK cells, CD8 + NK cells and CD8 – NK cells, $n=12$) and T cells (CD4 + T cells, CD8 + T cells, NKT cells, Pan T cells and Tregs, $n=16$; Accomando *et al*, 2012). Similarly, DNAm profiles in the N_MTL2 dataset were divided into two groups according to origin (myeloid or lymphoid; Reinius *et al*, 2012).

The HNSCC_PB dataset included DNAm profiles of peripheral whole blood samples for 92 HNSCC patients and 92 normal controls (Langevin *et al*, 2012). The HNSCC_SR dataset included DNAm profiles of serum samples for 12 HNSCC patients and 12 normal controls. The OVC_PB dataset included DNAm profiles of peripheral whole blood samples for 274 pre-treatment ovarian cancer patients and 131 normal controls after excluding samples of post-treatment ovarian cancer patients (Teschendorff *et al*, 2009). In addition, two reported lists of aberrant DNAm detected in peripheral leukocytes of pancreatic cancer (Pedersen *et al*, 2011) and SCLC (Wang *et al*, 2010) patients were collected.

The RA_PL dataset included DNAm profiles of peripheral leukocytes samples for 354 RA patients and 335 normal controls (Liu *et al*, 2013). In addition, a reported list of aberrant DNAm detected from peripheral whole blood of IBD patients was collected to validate our results (Nimmo *et al*, 2012).

Data preprocessing. For each CpG locus, the detected P -value reported by the Bead Studio (Illumina) was used to evaluate the methylation signal reliability. If the detected P -value of a CpG locus was >0.05 , then the DNAm signal was considered unreliable and its methylation level was considered to be missing. A profile with $>10\%$ missing values was deleted. After examining all DNAm profiles, two and four profiles were excluded from HNSCC_PB and OVC_PB datasets, respectively. The missing values in the remaining profiles were estimated by the k -nearest neighbor algorithm ($k=1$; Troyanskaya *et al*, 2001). For a CpG locus, the methylation level (β) was calculated by

$$\beta = \frac{\max(M, 0)}{|U|+|M|+100} \quad (1)$$

where M and U represent the methylated and unmethylated signal intensity of this locus, respectively (Bibikova *et al*, 2006). The methylation level (β) takes values ranging from 0 to 1. A value of 0 indicates no methylation and 1 indicates complete methylation.

Detection of differentially methylated CpG loci. We employed the student's t -test to detect differentially methylated CpG loci (denoted as DM loci) between two groups of samples with a false discovery rate (FDR; Benjamini and Hochberg, 1995) <0.01 . We also required that the absolute difference of the average methylation levels of a DM locus between two groups ($\Delta\beta$) was >0.03 (Wang *et al*, 2010).

Consistency analysis of two lists of DM loci. If the average methylation level of a DM locus detected in a dataset was higher or lower for the case compared to the control, we defined the alteration state of this locus as hypermethylated or hypomethylated. Suppose two DM loci lists selected separately from two datasets share k DM loci among which s loci have the same alteration states in these two datasets, then the probability of observing at least s loci by chance can be calculated according to

the following cumulative binomial distribution model (Bahn, 1969):

$$P = 1 - \sum_{i=0}^{s-1} \binom{k}{i} p^i (1-p)^{k-i} \quad (2)$$

where $p=0.5$ represents the probability of a locus having the same alteration state in two datasets by random chance. The alteration states of DM loci in the two lists are considered significantly consistent if the binomial P -value is smaller than 0.05.

DNAm alterations determined by shifts in myeloid and lymphoid cell populations. First, we hypothesised about DNAm alterations in the peripheral whole blood/leukocytes of a disease sample. The assumption was that the increase of myeloid cells and the decrease of lymphoid cells are chief contributors to DNAm alterations observed in the peripheral whole blood/leukocytes for a disease. For a DNA locus, suppose its average DNAm level in myeloid and lymphoid cells are β_m and β_l , respectively, then the DNAm level in the normal whole blood/leukocytes can be represented as

$$\beta_{\text{normal}} = \beta_m \times P_m + \beta_l \times P_l \quad (3)$$

where P_m and P_l ($P_l = 1 - P_m$) represent the proportion of myeloid and lymphoid cells in the peripheral whole blood of a normal sample, respectively.

In a disease sample, the DNAm level of the locus can be represented as

$$\beta_{\text{disease}} = (\beta_m + \Delta\beta_m) \times (P_m + \Delta k) + (\beta_l + \Delta\beta_l) \times (P_l - \Delta k) \quad (4)$$

where Δk represents the changed proportion of myeloid cells or lymphoid cells; $\Delta\beta_m$ and $\Delta\beta_l$ represent the difference of DNAm level of myeloid cells and the difference of DNAm level of lymphoid cells between disease sample and normal control, respectively.

The difference in DNAm level of this locus between disease sample and normal control should be

$$\Delta\beta = \beta_{\text{disease}} - \beta_{\text{normal}} = \Delta k \times (\beta_m - \beta_l) + (P_m + \Delta k) \times \Delta\beta_m + (P_l - \Delta k) \times \Delta\beta_l \quad (5)$$

Thus, if there is no DNAm level change of this locus in myeloid and lymphoid cells under the disease condition (i.e., $\Delta\beta_m = 0$ and $\Delta\beta_l = 0$), the DNAm level difference between disease sample and normal control is determined by the proportion changes of myeloid and lymphoid cells and the alteration state of this locus in the disease will be consistent with that in myeloid cells compared to lymphoid cells. When there are DNAm alterations in myeloid and/or lymphoid cells but their influence ($(P_m + \Delta k) \times \Delta\beta_m + (P_l - \Delta k) \times \Delta\beta_l$) is smaller than the influence of proportion changes ($\Delta k \times (\beta_m - \beta_l)$), the alteration state of this locus in the disease will also be consistent with that in myeloid cells compared to lymphoid cells, although the alterations in myeloid and/or lymphoid cells could affect the extent of the DNAm level difference between disease sample and normal control.

In serum, the cell-free DNA may come from multiple sources including the normal tissue cells, the cells with pathological changes in disease tissue and apoptotic or necrotic leukocytes under the disease condition. Suppose that the cell-free DNA from normal tissue cells are appropriately equal in disease and normal serum, then the DNAm level difference of a CpG locus between disease sample and normal control can be represented as:

$$\Delta\beta = \Delta k' \times (\beta_m - \beta_l) + (P_m + \Delta k') \times \Delta\beta_m + (P_l - \Delta k') \times \Delta\beta_l + P_l \times \left(\beta_l - \left((\beta_m + \Delta\beta_m) \times (P_m + \Delta k') + (\beta_l + \Delta\beta_l) \times (P_l - \Delta k') \right) \right) \quad (6)$$

where $\Delta k'$ represents the changed proportion of cell-free DNA from myeloid and lymphoid cells under the disease condition, which is determined by the proportion changes and the apoptosis

rates of these two cell types. P_t represents the proportion of cell-free DNA from the disease tissue cells, and β_t represents the corresponding DNAm level. According to formula (6), apart from cell-free DNA from myeloid and lymphoid cells, the DNAm level difference between disease sample and normal control could also be influenced by cell-free DNA from the disease tissue cells. Similarly, as in peripheral whole blood/leukocytes, if the influences of other factors, namely the DNAm level changes in myeloid and lymphoid cells ($(P_m + \Delta k') \times \Delta\beta_m + (P_l - \Delta k') \times \Delta\beta_l$) and the cell-free DNA from disease tissue cells ($P_t \times (\beta_t - ((\beta_m + \Delta\beta_m) \times (P_m + \Delta k') + (\beta_l + \Delta\beta_l) \times (P_l - \Delta k')))$), are smaller than the influence of the proportion changes of cell-free DNA from myeloid and lymphoid cells ($\Delta k' \times (\beta_m - \beta_l)$), the DNAm alteration state of this locus in disease sample compared to normal control could be consistent with that in myeloid cells compared to lymphoid cells.

In peripheral whole blood/leukocytes and serum, if the alteration state of a locus in disease sample compared to normal control is observed to be consistent with that in myeloid cells compared to lymphoid cells, we consider that the DNAm alteration of this locus in disease sample is predominantly determined by the proportion changes of myeloid and lymphoid cells under the disease condition according to the above deduction, although other factors could affect the extent of the DNAm alterations.

RESULTS

The DNAm difference between myeloid and lymphoid cells.

With $FDR < 0.01$ and $\Delta\beta \geq 0.03$, 4385 and 7668 DM loci between myeloid and lymphoid cells were identified from the N_MTL1 and N_MTL2 datasets, respectively. Among the 4385 DM loci detected from the first dataset, 94.2% were also detected from the second dataset and all of them showed the same alteration states in the two datasets, indicating that the DM loci identified from both datasets were significantly consistent (binomial test, P -value $< 2.2 \times 10^{-16}$). Integrating the DM loci extracted from the two datasets, we obtained 7924 DM loci between myeloid and lymphoid cells, referred to as the M-L DM loci list.

Then, we detected the DM loci lists for each pair of myeloid cell subtype and lymphoid cell subtype from N_MTL2 dataset and compared these lists with the M-L DM loci list. The result showed that most of the DM loci in these lists were overlapped with M-L DM loci list and at least 98.2% of the overlapped DM loci had the same alteration states (binomial test, all P -value $< 2.2 \times 10^{-16}$; Supplementary Table S1). For example, for the 4551 DM loci detected between granulocytes and CD4+ T cells, 91.1% of them were also found to be differentially methylated between myeloid and lymphoid cells and 99.9% of them had the same alteration states in granulocytes compared to CD4+ T cells with their alteration states in myeloid cells compared to lymphoid cells (binomial test, P -value $< 2.2 \times 10^{-16}$). These results suggested that it should be reasonable to classify the peripheral leukocyte cells into two groups (myeloid and lymphoid cells) to analyse the influence of cell population shifts to the observed blood-borne DNAm alterations under the cancer or inflammation condition.

Blood-borne DNAm alterations for cancer. First, we analysed the DNAm alteration states of the DM loci observed in the peripheral whole blood of cancer patients. From the HNSCC_PB dataset for peripheral whole blood of HNSCC patients, we detected 546 DM loci in HNSCC with $FDR < 0.01$ and $\Delta\beta \geq 0.03$. Among these DM loci, 98.7% were included in the M-L DM loci list and their methylation alteration states in HNSCC samples (compared to normal controls) were consistent with their alteration states in myeloid cells (compared to lymphoid cells), which was unlikely to be observed by chance (binomial test, P -value $< 2.2 \times 10^{-16}$; Figure 1). Similarly, among the 839 DM loci detected from the

OVC_PB dataset for ovarian cancer peripheral whole blood, 98.8% were included in the M-L DM loci list and their methylation alteration states in ovarian cancer were consistent with their methylation alteration states in myeloid cells (binomial test, P -value $< 2.2 \times 10^{-16}$).

Then, we analysed DNAm alterations in cancer peripheral leukocytes using a list of 96 DM loci for pancreatic cancer (Pedersen *et al*, 2011) and a list of 36 DM loci for SCLC; Wang *et al*, 2010) extracted from the published literatures. Among the 96 DM loci for pancreatic cancer, 12 loci were measured in the platforms analysed in this study and all were included in the M-L DM loci list. All of their alteration states in pancreatic cancer were consistent with those in myeloid cells (binomial test, P -value = 2.44×10^{-4}). Similarly, among the 36 DM loci for SCLC, 11 loci were measured in the platforms analysed in this study and all were differentially methylated in myeloid cells compared to lymphoid cells with the same alteration states in SCLC (binomial test, P -value = 4.88×10^{-4}).

Finally, we examined the DNAm alterations observed in cancer serum compared to normal controls. Using the HNSCC_SR dataset, we detected 4608 DM loci in HNSCC serum, among which 94.7% were included in the M-L DM loci list and 99.9% of their methylation alteration states in HNSCC serum were consistent with their alteration states in myeloid cells (binomial test, P -value $< 2.2 \times 10^{-16}$).

Our data suggested that almost all aberrant DNAm loci observed in the peripheral whole blood/leukocytes and serum of cancer patients compared to normal controls were predominantly determined by the population shifts in myeloid and lymphoid cells, in the sense that their alteration states in disease samples compared

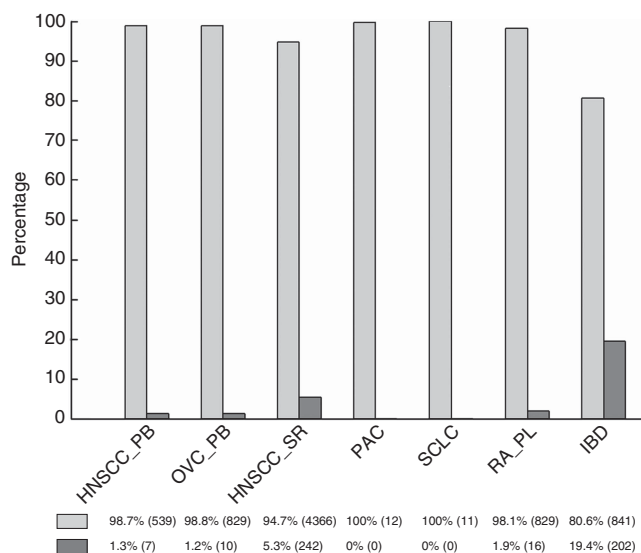


Figure 1. Overlap of the DM loci of cancer and inflammatory diseases with M-L DM loci. For each dataset, light and dark grey bars represent percentages of DM loci detected in this dataset that overlapped/did not overlap with the M-L DM loci list, respectively. DM loci numbers are depicted in brackets after the corresponding percentages. DM = differentially methylated; M-L DM loci = DM loci in myeloid cells compared to lymphoid cells; HNSCC_PB and OVC_PB = DM loci detected from peripheral whole blood of head and neck squamous cell carcinoma and ovarian cancer samples, respectively; HNSCC_SR = DM loci from the serum samples of head and neck squamous cell carcinoma; PAC and SCLC = DM loci from peripheral leukocytes of pancreatic and small-cell lung cancer samples, respectively; RA_PL and IBD = DM loci detected from peripheral leukocytes of rheumatoid arthritis samples and peripheral whole blood of inflammatory bowel disease samples, respectively.

to normal controls mainly reflect the DNAm difference between myeloid and lymphoid cells (see Methods). As the substantial correlations among multiple tests would potentially distort the estimation of the true proportion of DM loci (Leek and Storey, 2007), we further compared the DM loci lists detected from HNSCC_PB, OVC_PB and HNSCC_SR datasets using different thresholds of P -value with M-L DM loci list. The results showed that the DNAm alteration states of DM loci detected with different P -value thresholds also mainly reflected the DNAm difference between myeloid and lymphoid cells (Table 2). The results also showed that more stringent thresholds will reduce the number of detected DM loci overlapped with the M-L DM loci list. Thus, a few DNAm alterations detected from the peripheral whole blood/leukocyte and serum of cancer patients were not included in the M-L DM loci list with $FDR < 0.01$ (Figure 1). This could also be explained by the fact that the M-L DM loci list included only a portion of all DM loci between myeloid and lymphoid cells. For example, among the 242 DM loci detected in the HNSCC serum but not included in the M-L DM loci list, 58 loci tended to be significantly different (with an unadjusted P -value < 0.05) between myeloid and lymphoid cells in both the N_MTL1 and N_MTL2 datasets and all their methylation states in HNSCC serum were consistent with their methylation states in myeloid cells. Of particular interest, the alteration states of four DM loci observed in HNSCC serum were inconsistent with respect to their alteration states in myeloid cells. However, since all the differential DNAm loci were determined at a certain FDR control level, we could not exclude the possibility that this observation was introduced by the false positives for the cancer and/or for the M-L DM loci list.

Blood-borne DNAm alterations for inflammatory diseases.

Then, we examined whether the alteration states of blood-borne DNAm alterations observed for inflammatory diseases were consistent with their alteration states in myeloid cells compared to lymphoid cells (Figure 1). Using the RA_PL dataset for peripheral leukocytes of RA patients, we identified 841 DM loci in RA with $FDR < 0.01$ and $\Delta\beta \geq 0.03$, among which 98.1% DM loci were included in the M-L DM loci list and all methylation

alteration states in RA were consistent with their alteration states in myeloid cells (binomial test, P -value $< 2.2 \times 10^{-16}$). Similar result was observed for DM loci detected with low-stringency thresholds ($FDR < 0.05$ and $P < 0.05$; Table 2). For the 1117 DM CpG loci previously reported for IBD (Nimmo *et al*, 2012), 1043 were measured in all platforms analysed in this study, among which 80.6% were included in the M-L DM loci list and 98.8% of their methylation alterations in IBD were consistent with those in myeloid cells (binomial test, P -value $< 2.2 \times 10^{-16}$). There were 10 loci whose methylation alteration states in IBD were inconsistent with their alteration states in myeloid cells. However, we could not exclude the possibility that this observation was introduced by false positives for IBD and/or for the M-L DM loci list.

The above results clearly showed that almost all DNAm alterations observed in the peripheral leukocytes of RA and peripheral whole blood of IBD patients were predominantly determined by the shifted subpopulations of myeloid and lymphoid cells, in the sense that their alteration states in disease samples compared to normal controls mainly reflect the DNAm difference between myeloid and lymphoid cells. Although some DM loci observed for these inflammatory diseases were not included in the M-L DM loci list, they could be explained by the incompleteness of the M-L DM loci list. Notably, although no data were available for analysing DNAm alterations in inflammatory disease patients' serum, it is reasonable that DNAm alterations in inflammatory disease serum could also be determined by the shifted subpopulations of myeloid and lymphoid cells, which also undergo greater apoptosis in inflammatory diseases (Courtney *et al*, 1999).

Comparisons of blood-borne DNAm alterations in cancer and inflammatory disease.

As described above, blood-borne DNAm alterations observed in both cancer and inflammatory disease patients were determined by shifted populations of myeloid and lymphoid cells, suggesting that blood-borne DNAm alterations in cancer patients were consistent with the blood-borne alterations in inflammatory disease patients. Thus, we compared DNAm alterations observed in the peripheral whole blood (or peripheral leukocytes) from both patient types.

Among the 546 DM loci identified from the HNSCC_PB dataset, 75.6% were significantly altered in RA peripheral leukocytes and all methylation alteration states in HNSCC were consistent with those in RA (binomial test, P -value $< 2.2 \times 10^{-16}$). For the remaining 133 DM loci detected in HNSCC peripheral whole blood samples, 125 tended to be differentially methylated in RA peripheral leukocytes (unadjusted P -value < 0.05) and all of their alteration states in HNSCC were consistent with those in RA. Similarly, among the 839 DM loci identified from the OVC_PB dataset, 65.7% were significantly altered in RA peripheral leukocytes and all methylation alteration states in ovarian cancer were consistent with those in RA (binomial test, P -value $< 2.2 \times 10^{-16}$). Among the remaining 288 DM loci in ovarian cancer samples, 270 loci tended to be differentially methylated in the peripheral leukocytes of RA patients (unadjusted P -value < 0.05) and only one of these loci had an inconsistent alteration states in ovarian cancer and RA.

Among the 546 DM loci identified from the HNSCC_PB dataset, 61.3% were included in the DM loci list previously reported for IBD and all were similar with respect to alteration states in these two diseases (binomial test, P -value $< 2.2 \times 10^{-16}$). Among the 839 DM loci identified from the OVC_PB dataset, 48.3% were included in the IBD DM loci list and all except one were of the same alteration states in these two diseases (binomial test, P -value $< 2.2 \times 10^{-16}$). The low coverage of the DM loci of HNSCC and ovarian cancer by the IBD DM loci list could be due to the incompleteness of the IBD DM loci list. These data suggest

Table 2. The consistency of DM loci detected at different significant levels for cancer and inflammation diseases with M-L DM loci list

Dataset	DM loci ^a	Overlapped DM loci ^b	Consistent DM loci ^c (%)	Binomial P-value ^d
FDR < 0.05 and $\Delta\beta \geq 0.03$				
HNSCC_PB	614	606 (98.7%)	100	$< 2.2 \times 10^{-16}$
OVC_PB	867	847 (97.6%)	99.8	$< 2.2 \times 10^{-16}$
HNSCC_SR	5618	5181 (92.2%)	99.9	$< 2.2 \times 10^{-16}$
RA_PL	842	825 (98.0%)	100	$< 2.2 \times 10^{-16}$
P < 0.05 and $\Delta\beta \geq 0.03$				
HNSCC_PB	679	655 (96.5%)	99.8	$< 2.2 \times 10^{-16}$
OVC_PB	879	856 (97.4%)	99.9	$< 2.2 \times 10^{-16}$
HNSCC_SR	6113	5513 (90.2%)	99.9	$< 2.2 \times 10^{-16}$
RA_PL	843	825 (97.9%)	100	$< 2.2 \times 10^{-16}$

Abbreviations: DM = differentially methylated; FDR = false discovery rate; HNSCC = head and neck squamous cell carcinoma; M-L DM loci = DM loci in myeloid cells compared to lymphoid cells; OVC = ovarian cancer; PB = peripheral whole blood; PL = peripheral leukocytes; RA = rheumatoid arthritis; SR = serum.

^aThe number of DM loci identified from each dataset with different significant levels.

^bThe number of identified DM loci overlapped with M-L DM loci list and corresponding percentage.

^cThe percentage of overlapped DM loci having the consistent alteration states with those in myeloid and lymphoid cells.

^dThe binomial distribution P -value.

that DNAm alterations in the peripheral blood of HNSCC and ovarian cancer patients were consistent with alterations in the peripheral leukocytes of RA and peripheral whole blood of IBD patients.

DISCUSSION

Our results indicate that most, if not all, DNAm alterations observed in the peripheral whole blood/leukocytes of cancer patients are predominantly determined by the population shifts in myeloid and lymphoid cells, in the sense that their alteration states in cancer samples compared to normal controls mainly reflect the DNAm difference between myeloid and lymphoid cells. Our analyses indicate that DNAm alterations observed in the serum of cancer patients also mainly reflect the DNAm difference between myeloid and lymphoid cells. Therefore, the signals of circulating cell-free DNA from cancer cells could be largely masked by the signals of cell-free DNA from myeloid and lymphoid cells. This can explain previous reports that DM loci found in cancer tissues infrequently appear in serum (Ichikawa *et al*, 2004; Fujiwara *et al*, 2005; Van De Voorde *et al*, 2012). Therefore, preselecting candidate blood-borne DNAm biomarkers from DNAm loci previously found to be hypermethylated in cancer tissues is a misleading strategy (Ichikawa *et al*, 2004; Fujiwara *et al*, 2005; Van De Voorde *et al*, 2012). To avoid the influence of shifted populations of myeloid and lymphoid cells, a reasonable approach may be to focus on finding blood-borne DNAm biomarkers of cancer from those DNA loci that tend to have similar amounts of DNAm in myeloid and lymphoid cells. Unfortunately, by analysing the 242 DM loci in HNSCC serum detected from the HNSCC_SR dataset (but not included in the M-L DM loci list), we found no evidence for the feasibility of this approach. From these 242 DM loci, we selected 40 DM loci that were unlikely to be differentially methylated between peripheral myeloid and lymphoid cells according to the criterion that their unadjusted *P*-values in both the N_MTL1 and N_MTL2 datasets were >0.2 . Among these 40 DM loci, 16 were hypermethylated in the HNSCC serum samples compared to normal controls but none were hypermethylated in the HNSCC tissue samples compared to normal controls in the dataset for HNSCC from The Cancer Genome Atlas database (McLendon, 2008). According to this analysis, we could not determine whether these DM loci originated from circulating DNA released from cancer tissues, the clarification of which may require more data and experiment validation.

Our results also showed that DNAm alterations observed in the peripheral whole blood/leukocytes of patients with inflammatory diseases compared to normal controls are significantly consistent with those observed in cancer patients compared to normal controls. This phenomenon is likely caused by the similar proportion changes of myeloid and lymphoid cells under the disease conditions. In fact, not only the myeloid/lymphoid proportions, but also the leukocyte subtype proportions tended to have similar changes under cancer or inflammation condition. According to the estimation reported by previous studies (Houseman *et al*, 2012; Liu *et al*, 2013), the proportions of T cells, B cells and NK cells tended to decrease while monocytes and granulocytes tended to increase under both the cancer and inflammation conditions. Although we were unable to examine the consistency of the blood-borne DNAm alterations between particular types of cancer and their corresponding inflammatory diseases due to limited data sources, we hypothesise that they would be similar due to similar trends of population shifts of peripheral myeloid and lymphoid cells for both cancers (Kuss *et al*, 2004; Cho *et al*, 2009; Accomando *et al*, 2012; Houseman *et al*, 2012) and inflammatory diseases (Hanai *et al*, 2004;

Domagala-Kulawik *et al*, 2007; Zhang *et al*, 2011). Therefore, the blood-borne DNAm alterations detected from cancer samples compared to normal controls might be difficult to distinguish cancer from inflammatory diseases.

Hopefully, subtle differences of peripheral leukocyte subpopulation shifts between cancer and inflammatory patients may exist, which could lead to DNAm differences between cancer and inflammatory patients. For example, according to the estimated proportions of leukocyte subtypes in peripheral bloods of HNSCC, OVC and peripheral leukocytes RA patients (Houseman *et al*, 2012; Liu *et al*, 2013), we found that the average proportion of myeloid cells increased 8.0% and 9.5% in HNSCC and OVC patients, respectively, while increased 12.3% in RA patients compared to normal controls. To detect such subtle DNAm difference between cancer and inflammatory diseases, a direct comparison of cancer and inflammatory disease samples is needed. It is also possible that DNAm alterations occurring in specific leukocyte subtypes in cancer or inflammatory diseases could serve as diagnostic biomarkers to distinguish cancer from inflammatory diseases. Actually, it has been found that, compared to normal controls, expression of some interferon-stimulated genes (ISGs) tends to be significantly downregulated in peripheral CD4 + T, CD8 + T and B cells of patients with melanoma, breast and gastrointestinal cancers (Critchley-Thorne *et al*, 2007; Critchley-Thorne *et al*, 2009), whereas some of these ISGs tend to be significantly upregulated in peripheral CD4 + T, CD8 + T and B cells of patients with inflammatory diseases such as systemic lupus erythematosus (Becker *et al*, 2013). A complicated method has been proposed to identify DNAm alterations occurring in specific leukocyte subtypes of patients with a particular disease from peripheral whole blood/leukocyte samples by adjusting the influence of the leukocyte subpopulation shifts (Liu *et al*, 2013). Nevertheless, the estimation of the leukocyte proportions required by this method tends to be rough as it depends on the proper selection of cell-type-specific markers (Zhong *et al*, 2013). Obviously, it would be more efficient to measure and compare cell-type-specific DNAm profiles for cancer and inflammatory diseases to identify cancer-specific diagnostic biomarker.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 81372213, 81071646, 91029717).

REFERENCES

- Accomando WP, Wiencke JK, Houseman EA, Butler RA, Zheng S, Nelson HH, Kelsey KT (2012) Decreased NK cells in patients with head and neck cancer determined in archival DNA. *Clin Cancer Res* **18**(22): 6147–6154.
- Bahn AK (1969) Application of binomial distribution to medicine: comparison of one sample proportion to an expected proportion (for small samples). Evaluation of a new treatment. Evaluation of a risk factor. *J Am Med Womens Assoc* **24**(12): 957–966.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Edgar R (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* **37**(Database issue): D885–D890.
- Becker AM, Dao KH, Han BK, Kornu R, Lakhanpal S, Mobley AB, Li QZ, Lian Y, Wu T, Reimold AM, Olsen NJ, Karp DR, Chowdhury FZ, Farrar JD, Satterthwaite AB, Mohan C, Lipsky PE, Wakeland EK, Davis LS (2013) SLE peripheral blood B cell, T cell and myeloid cell transcriptomes display unique profiles and each subset contributes to the interferon signature. *PLoS One* **8**(6): e67003.

- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57: 289–300.
- Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E, Goldmann T, Seifart C, Jiang W, Barker DL, Chee MS, Floros J, Fan JB (2006) High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 16(3): 383–393.
- Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, Glockner S, Piantadosi S, Gabrielson E, Pridham G, Pelosky K, Belinsky SA, Yang SC, Baylin SB, Herman JG (2008) DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med* 358(11): 1118–1128.
- Cho H, Hur HW, Kim SW, Kim SH, Kim JH, Kim YT, Lee K (2009) Pre-treatment neutrophil to lymphocyte ratio is elevated in epithelial ovarian cancer and predicts survival after treatment. *Cancer Immunol Immunother* 58(1): 15–23.
- Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Kushi LH, Kwan ML, Wiencke JK (2010) Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS Genet* 6(7): e1001043.
- Courtney PA, Crockard AD, Williamson K, Irvine AE, Kennedy RJ, Bell AL (1999) Increased apoptotic peripheral blood neutrophils in systemic lupus erythematosus: relations with disease activity, antibodies to double stranded DNA, and neutropenia. *Ann Rheum Dis* 58(5): 309–314.
- Critchley-Thorne RJ, Simons DL, Yan N, Miyahira AK, Dirbas FM, Johnson DL, Swetter SM, Carlson RW, Fisher GA, Koong A, Holmes S, Lee PP (2009) Impaired interferon signaling is a common immune defect in human cancer. *Proc Natl Acad Sci USA* 106(22): 9010–9015.
- Critchley-Thorne RJ, Yan N, Nacu S, Weber J, Holmes SP, Lee PP (2007) Down-regulation of the interferon signaling pathway in T lymphocytes from patients with metastatic melanoma. *PLoS Med* 4(5): e176.
- Domagala-Kulawik J, Hoser G, Dabrowska M, Chazan R (2007) Increased proportion of Fas positive CD8+ cells in peripheral blood of patients with COPD. *Respir Med* 101(6): 1338–1343.
- Fujiwara K, Fujimoto N, Tabata M, Nishii K, Matsuo K, Hotta K, Kozuki T, Aoe M, Kiura K, Ueoka H, Tanimoto M (2005) Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. *Clin Cancer Res* 11(3): 1219–1225.
- García-Closas M, Gail MH, Kelsey KT, Ziegler RG (2013) Searching for blood DNA methylation markers of breast cancer risk and early detection. *J Natl Cancer Inst* 105(10): 678–680.
- Hanai H, Takeuchi K, Iida T, Kashiwagi N, Saniabadi AR, Matsushita I, Sato Y, Kasuga N, Nakamura T (2004) Relationship between fecal calprotectin, intestinal inflammation, and peripheral blood neutrophils in patients with active ulcerative colitis. *Dig Dis Sci* 49(9): 1438–1443.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13: 86.
- Ichikawa D, Koike H, Ikoma H, Ikoma D, Tani N, Otsuji E, Kitamura K, Yamagishi H (2004) Detection of aberrant methylation as a tumor marker in serum of patients with gastric cancer. *Anticancer Res* 24(4): 2477–2481.
- Koestler DC, Marsit CJ, Christensen BC, Accomando W, Langevin SM, Houseman EA, Nelson HH, Karagas MR, Wiencke JK, Kelsey KT (2012) Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. *Cancer Epidemiol Biomarkers Prev* 21(8): 1293–1302.
- Kuss I, Hathaway B, Ferris RL, Gooding W, Whiteside TL (2004) Decreased absolute counts of T lymphocyte subsets and their relation to disease in squamous cell carcinoma of the head and neck. *Clin Cancer Res* 10(11): 3755–3762.
- Langevin SM, Koestler DC, Christensen BC, Butler RA, Wiencke JK, Nelson HH, Houseman EA, Marsit CJ, Kelsey KT (2012) Peripheral blood DNA methylation profiles are indicative of head and neck squamous cell carcinoma: an epigenome-wide association study. *Epigenetics* 7(3): 291–299.
- Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 3(9): 1724–1735.
- Li L, Choi JY, Lee KM, Sung H, Park SK, Oze I, Pan KF, You WC, Chen YX, Fang JY, Matsuo K, Kim WH, Yuasa Y, Kang D (2012) DNA methylation in peripheral blood: a potential biomarker for cancer molecular epidemiology. *J Epidemiol* 22(5): 384–394.
- Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 31(2): 142–147.
- Lonneborg A, Aaroe J, Dumeaux V, Borresen-Dale AL (2009) Found in transcription: gene expression and other novel blood biomarkers for the early detection of breast cancer. *Expert Rev Anticancer Ther* 9(8): 1115–1123.
- Maier S, Dahlstroem C, Haefliger C, Plum A, Piepenbrock C (2005) Identifying DNA methylation biomarkers of cancer drug response. *Am J Pharmacogenomics* 5(4): 223–232.
- Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT (2011) DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. *J Clin Oncol* 29(9): 1133–1139.
- Mclendon R (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216): 1061–1068.
- Nimmo ER, Prendergast JG, Aldhous MC, Kennedy NA, Henderson P, Drummond HE, Ramsahoye BH, Wilson DC, Semple CA, Satsangi J (2012) Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. *Inflamm Bowel Dis* 18(5): 889–899.
- Pedersen KS, Bamlet WR, Oberg AL, de Andrade M, Matsumoto ME, Tang H, Thibodeau SN, Petersen GM, Wang L (2011) Leukocyte DNA methylation signature differentiates pancreatic cancer patients from healthy controls. *PLoS One* 6(3): e18223.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 7(7): e41361.
- Robertson KD (2005) DNA methylation and human disease. *Nat Rev Genet* 6(8): 597–610.
- Saito H, Takaya S, Osaki T, Ikeguchi M (2013) Increased apoptosis and elevated Fas expression in circulating natural killer cells in gastric cancer patients. *Gastric Cancer* 16(4): 473–479.
- Saito T, Dworacki G, Gooding W, Lotze MT, Whiteside TL (2000) Spontaneous apoptosis of CD8+ T lymphocytes in peripheral blood of patients with advanced melanoma. *Clin Cancer Res* 6(4): 1351–1364.
- Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M (2009) An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One* 4(12): e8274.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* 17(6): 520–525.
- Van De Voorde L, Speeckaert R, Van Gestel D, Bracke M, De Neve W, Delanghe J, Speeckaert M (2012) DNA methylation-based biomarkers in serum of patients with breast cancer. *Mutat Res* 751(2): 304–325.
- Wang L, Aakre JA, Jiang R, Marks RS, Wu Y, Chen J, Thibodeau SN, Pankratz VS, Yang P (2010) Methylation markers for small cell lung cancer in peripheral blood leukocyte DNA. *J Thorac Oncol* 5(6): 778–785.
- Yoshikawa T, Saito H, Osaki T, Matsumoto S, Tsujitani S, Ikeguchi M (2008) Elevated Fas expression is related to increased apoptosis of circulating CD8+ T cell in patients with gastric cancer. *J Surg Res* 148(2): 143–151.
- Zhang JY, Zou ZS, Huang A, Zhang Z, Fu JL, Xu XS, Chen LM, Li BS, Wang FS (2011) Hyper-activated pro-inflammatory CD16 monocytes correlate with the severity of liver injury and fibrosis in patients with chronic hepatitis B. *PLoS One* 6(3): e17484.
- Zhong Y, Wan YW, Pang K, Chow LM, Liu Z (2013) Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC Bioinformatics* 14: 89.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)