Article Title: Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma

Short Title: Global methylation analysis in myeloma

Authors: Brian A. Walker¹, Christopher P. Wardell¹, Laura Chiecchio², Emma M. Smith¹, Kevin D. Boyd¹, Antonino Neri³, Faith E. Davies¹, Fiona M. Ross², Gareth J. Morgan¹

Institutions:
¹Section of Haemato-Oncology, The Institute of Cancer Research, London, United Kingdom.
²Leukaemia Research Fund UK Myeloma Forum Cytogenetics Group, Wessex Regional Cytogenetic Laboratory, Salisbury, United Kingdom.
³Laboratorio di Ematologia Sperimentale e Genetica Molecolare, Università di Milano, Ospedale Maggiore IRCCS, Milan, Italy.

Corresponding Author: Dr Brian Walker, Section of Haemato-Oncology, The Institute of Cancer Research, 15 Cotswold Road, London, SM2 5NG, United Kingdom. Telephone +44(0)20 87224130, Fax +44(0)20 8722 4432, Email: brian.walker@icr.ac.uk

Scientific Category: Lymphoid Neoplasia
Abstract

We have used genome-wide methylation microarrays to analyze differences in CpG methylation patterns in cells relevant to the pathogenesis of myeloma plasma cells, including B cells, normal plasma cells, MGUS, presentation myeloma, and plasma cell leukemia (PCL). We show that methylation patterns in these cell types are capable of distinguishing non-malignant from malignant cells and that the main reason for this difference is hypomethylation of the genome at the transition from MGUS to presentation myeloma. In addition, gene-specific hypermethylation was evident at the myeloma stage. Differential methylation was also evident at the transition from myeloma to PCL with re-methylation of the genome, particularly of genes involved in cell-cell signaling and cell adhesion, which may contribute to independence from the bone marrow microenvironment. There was a high degree of methylation variability within presentation myeloma samples and this was associated with the cytogenetic differences between samples. More specifically, we found that methylation subgroups were defined by translocations and hyperdiploidy, with t(4;14) myeloma having the largest impact on DNA methylation. Two groups of hyperdiploid samples were identified, based on unsupervised clustering, which had an impact on overall survival. Overall, DNA methylation changes significantly during disease progression and between cytogenetic subgroups.
Introduction
Upon encounter with antigen, naive B cells undergo somatic hypermutation and class switch recombination in the germinal center, finally differentiating into plasma cells residing in the bone marrow.\textsuperscript{1} Multiple myeloma is a clonal malignancy of these plasma cells which develops as a consequence of a multistep transformation process. Insight into the molecular mechanisms underlying this transformation process can come from the study of the individual steps leading to myeloma which is known to evolve from a pre-malignant state, monoclonal gammopathy of undetermined significance (MGUS), and transforms into myeloma at a rate of 1\% per year.\textsuperscript{2, 3} Additional genetic events may transform the myeloma clone further to a more aggressive disease state known as plasma cell leukemia (PCL) where the clonal cells lose their dependency on the bone marrow microenvironment.

Genomic instability is a characteristic feature of myeloma cells where translocations involving the IGH locus and MMSET/FGFR3, CCND1, CCND3, MAF and MAFB occur, as well as numerous structural copy number alterations including del(1p), del(6q), del(8p), del(13q), del(16q), del(22) and gain of 1q.\textsuperscript{4-6} However, the mechanisms involved in the progression from MGUS to myeloma are incompletely understood as, although present at decreased frequencies, the genetic markers characteristic of myeloma such as IGH rearrangements, hyperdiploidy, and gains and losses of chromosomal regions are also present in MGUS.\textsuperscript{7, 8}

While there has been substantial work on the genetics of myeloma little is known about the epigenetic changes leading to disease progression. Changes in DNA methylation status are one of the key epigenetic features known to regulate gene expression. Methylation changes occur primarily at CpG dinucleotides which are present at a higher frequency in promoter regions as well as within repeat sequences and transposable elements.\textsuperscript{9} Hypomethylation in cancer cells mainly occurs within repeat sequences and transposable elements, whereas hypermethylation occurs in promoter regions, particularly of putative tumor suppressor genes.\textsuperscript{10} Such hypermethylation of DNA is linked with transcriptionally inactive heterochromatin, and is associated with methylated histone H3K9 residues.\textsuperscript{11-13}
With the exception of one recent study\textsuperscript{14} the epigenetic factors contributing to the pathogenesis of myeloma have been studied on a gene-by-gene basis and, using methylation specific PCR, several genes have been identified which are hypermethylated including \textit{VHL}, \textit{XAF1}, \textit{IRF8}, \textit{TP53}, \textit{CDKN2A}, \textit{CDKN2B}, \textit{DAPK}, \textit{SOCS1}, \textit{CDH1}, \textit{PTGS2}, \textit{CCND2} and \textit{DCC}.\textsuperscript{15-22} Promoter hypermethylation of \textit{CDKN2A} and \textit{TGFBR2} have been shown to correlate with poor prognosis in myeloma patients, although the prognostic value of \textit{CDKN2A} hypermethylation remains debatable.\textsuperscript{16, 23, 24}

In this study we used the Infinium array (Illumina) to analyze CpG island promoter methylation using normal plasma cells, MGUS, myeloma and PCL samples in order to identify methylation changes which may contribute to the pathogenesis of myeloma, or that could act as prognostic factors. Additionally, we utilized cytogenetic data available for the myeloma samples to identify methylation changes between known cytogenetic subgroups. These arrays have been used and validated by many groups and favorably correlate with whole-genome methylation sequencing technologies.\textsuperscript{25-28}

\textbf{Methods}

\textbf{Patient Samples and Clinical Data}\n
The MRC Myeloma IX trial recruited 1970 newly diagnosed patients and comprised two arms; the first for older and less fit patients and the second for younger fitter patients. The details of the trial have been published elsewhere but in summary younger, fitter patients were put on the intensive arm and received autologous transplantation following induction with cyclophosphamide, thalidomide and dexamethasone (CTD) or cyclophosphamide, vincristine, doxorubicin and dexamethasone (CVAD).\textsuperscript{29} The non-intensive arm consisted of older patients who were treated with either attenuated CTD (CTDa) or melphalan and prednisolone (MP). All patients were then randomized to thalidomide maintenance or no thalidomide maintenance. The trial was approved by the MRC Leukaemia Data Monitoring and Ethics committee (MREC 02/8/95, ISRCTN68454111).
Bone marrow aspirates were obtained after informed consent. Plasma cells (PC) from non-myeloma patients (normal plasma cell controls, n=3) and presentation myeloma samples (n=161) were selected to a purity of >90% using CD138 microbeads and magnet-assisted cell sorting (Miltenyi Biotech, Bisley, UK). In order to achieve a sufficient quantity of DNA some normal plasma cell control samples were pooled. MGUS samples (n=4) were analyzed by flow cytometry to determine the percentage of plasma cells within the leukocyte population (range 0.3-3.1%) and the percentage of those plasma cells with an abnormal phenotype (CD19-, CD56+, CD45-, range 80-100%). Plasma cells subsequently underwent cell selection using CD138 microbeads as above. Samples from PCL patients (n=7) were not CD138 selected but contained >90% plasma cell infiltration as determined by microscopy. PCL samples were genetically characterized by FISH, SNP 6.0 mapping array (Affymetrix), or U133 Plus 2.0 expression array. The seven PCL samples consisted of 3 t(4;14), 3 t(11;14), and 1 hyperdiploid (HRD) sample.

DNA was extracted using commercially available kits (RNA/DNA mini kit or Allprep kit, Qiagen, Crawley, UK) according to manufacturer’s instructions. DNA quality and quantity were determined on an ND-1000 Spectrophotometer (Nano-Drop Technologies, Wilmington, DE). Interphase FISH analysis was performed on purified PC using the micro-FISH technique and probes which have previously been documented. Briefly, probes to detect t(4;14) (n=15), t(6;14) (n=1), t(11;14) (n=35), t(14;16) (n=7), t(14;20) (n=3), del(1p32.3) (n=21), gain 1q (n=49), del(17p) (n=8) and hyperdiploidy (defined by gain of any 2 of chromosomes 5, 9 and 15, n=73) were used to identify abnormalities. Samples with a split IGH probe but no identified partner were termed Unknown translocation.

**Methylation Arrays**

500 ng of DNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research, CA) and subsequently processed for hybridization onto the Infinium humanmethylation27 BeadArray (Illumina) according to manufacturers protocols. This array interrogates 27 578 CpG dinucleotides encompassing 14 495 genes. Briefly, DNA was bisulfite treated, converting non-methylated C nucleotides to U (T),
whereas methylated C nucleotides remained unaffected. Bisulfite-treated DNA was subsequently amplified, fragmented and hybridized to locus-specific oligonucleotides on the BeadArray. C or T nucleotides were detected by fluorescence signal from single-nucleotide extension of the DNA fragments. Results were interpreted as a ratio (β value) of methylated signal (C) compared to the sum of methylated and unmethylated signal (C+T) for each locus, where a β-value of 0 represents fully unmethylated DNA and a value of 1 fully methylated DNA. Data have been submitted to GEO under accession number GSE21304.

Bisulfite PCR and Sequencing
100 ng of genomic DNA was bisulfate treated and purified using the Epitect bisulfate kit (Qiagen, Crawley, UK) following manufacturer’s instructions. Primers to amplify the regions surrounding methylation array probes were designed and are available in the Supplementary data. DNA was amplified using Platinum Taq DNA polymerase (Invitrogen, Paisley, UK), reactions purified and sequenced on a 3500 DNA capillary sequencer (Applied Biosystems, Warrington, UK). Sequences were analyzed using Sequencher 4.8 (Gene Codes, Ann Arbor, MI).

Expression Array Data
Expression array data are available for myeloma samples, have previously been published and are available under accession number GSE21349.5

Data Analysis
Data were analyzed in GenomeStudio using the methylation module (Illumina). Further analyses were performed using R and the LIMMA package.33, 34 Missing elements in the data were imputed using row means. Differential methylation between samples were identified using an empirical Bayes moderated t-test and the resulting p-values were adjusted using the Benjamini and Hochberg method.35 P-values below 0.05 were considered significant. Hierarchical clustering was performed using the Euclidean distance and Ward’s method.36 Cluster stability was ascertained with multiscale bootstrap analysis using the pvclust R package.37 Approximately unbiased p-values were calculated using 100,000 resamplings of the original data. To further investigate the relationships shown using hierarchical clustering, principal component analysis was performed on the methylation data.
Using a screeplot, the first three principal components were deemed to be significant and these data plotted against one another (see Supplementary Data).

Expression data were normalized using RMA (quantile normalized, median polish). Correlation between methylation and expression data was investigated using the Pearson correlation between the corresponding probes that mapped to the same gene symbol. Correlations were only considered if they were significantly different (p<0.05) from zero.

Results

Methylation Profiling of Different Disease Stages in Myeloma
Genome-wide methylation profiles were compared between normal B cells (n=6), normal plasma cells (n=3), MGUS samples (n=4), presentation myeloma (n=161), plasma cell leukemia (n=7) and human myeloma cell lines (HMCLs) (n=9). Using XY scatter plots significant variations in methylation profiles within the B cell subset of samples which were due to X-chromosome inactivation in females were noted. In order to remove these confounding data points we removed all probes present on the sex chromosomes leaving 26 486 probes on the array to interrogate.

Pair-wise comparisons on XY scatter plots were used to determine the variation in methylation within each cell type (Figure 1). Correlation coefficients were determined and the range and median values are shown in Figure 1 and detailed in Table 1. Median correlation coefficients >0.9 were seen in pre-malignant cell types such as B cells, normal plasma cells and MGUS plasma cells indicating a homogeneity of gene methylation between samples. However, in presentation myeloma, as well as in cell lines, the median correlation coefficients were <0.7 indicating a greater heterogeneity of methylation between samples of the same cell type. This heterogeneity may be the result of methylation variation between the different cytogenetic subgroups, and is investigated below.

The data for each probe were averaged over cell type and filtered to remove differences between the cell types characteristic of the multistep pathway seen in the development of myeloma which were not significant (p>0.05) or had an average
β-value <0.25 or >0.75 in both cell types. Overall methylation relationships were analyzed between cell types by cluster analysis using the Euclidean method (Figure 2A). The results of this analysis revealed that the overall methylation in these cell types can accurately distinguish between pre-malignant and malignant cells. MGUS samples cluster closely with normal plasma cells and B cells whereas presentation and PCL samples cluster closely with HMCLs. The large branch between non-malignant and malignant cell types is a consequence of the large difference in global methylation patterns between the two cell types, indicating an important role for methylation in the progression of MGUS to myeloma.

The number of probesets differing between component steps of the pathogenesis to myeloma is shown in Figure 2B. This analysis indicates that there are few methylation changes between normal plasma cells and the MGUS phenotype. In contrast, there are 3407 probes (1428 genes) which undergo hypomethylation and 87 probes which are hypermethylated from MGUS to presentation myeloma. However, it is the large number of probes which are hypomethylated that are the main cause for the distinct methylation differences between non-malignant and malignant myelomatous cell types. The probes are designated as being either within a CpG islands or not, and of the 3407 which undergo hypomethylation only 655 (19.2%) are within a CpG island, whereas 48 (58.5%) of the 82 which undergo hypermethylation are within a CpG island. As global hypomethylation of cancers is known to occur outside of CpG islands this is indicative of both global hypomethylation and gene specific methylation at the transformation of MGUS to myeloma.

At the transition from myeloma to PCL the main changes are hypermethylation of genes (2151 probes, 1802 genes). Of these 2151 probes, 1412 are not within a CpG island meaning that there is methylation change both within and outside CpG islands. Interestingly, 1168 of these probes (1068 genes) were previously de-methylated at the MGUS to myeloma transition consistent with re-methylation of previously de-methylated genes occurring at the transition to myeloma.
Gene Ontology (GO) analysis of the 82 probes (77 genes) hypermethylated at the transition to myeloma indicates that 3 main groups of genes affected are: regulation of developmental processes, cell cycle processes and regulation of transcription. The genes include transcription factors or genes which regulate transcription (ACVR1, ARID3A, BRCA2, C19orf33, CALCA, CBX4, FOXD2, GATA4, HIPK3, HOXB8, HOXD11, ID4, IRF7, LDB1, NCOR2, ONECUT2, RAB37, RUNX2, ZIC1, ZNF385, ZNF560), as well as regulators of cell cycle (ACVR1, AIF1, BCL2, BRCA2, CDKN2B, GAS2L1, ID4, MPHOSPH9, PKMYT1).

Of the 2151 probes (1802 genes) hypermethylated at the myeloma to PCL transition 739 are annotated as occurring within CpG islands. GO term analysis of these 739 probes reveals that cell-cell signaling (44 genes, \(p=5.28 \times 10^{-7}\)), cell development or differentiation (34 genes, \(p=3.28 \times 10^{-6}\)) and cell adhesion molecules (45 genes, \(p=1.2 \times 10^{-5}\)) are significantly enriched.

**Global Methylation Differences in Cytogenetic Subgroups**

When the dataset is analyzed using an unsupervised clustering approach per sample, rather than per cell type, there are 3 main clusters evident: non-malignant cells, HMCLs and t(4;14) samples, and other myeloma samples (Figure 3). The clustering was confirmed by principal component analysis (Supplementary data). The samples within the main myeloma group can be divided into 5 clades, and when the samples are annotated according to FISH results these clades are clearly defined by cytogenetic abnormalities. Clades B and C are mainly hyperdiploid samples, but clade C is more closely related to translocation samples. Clades D through to G are predominantly samples with translocations with clades D and F consisting of t(11;14) samples and clade E of t(14;16) samples. Translocation/Cyclin D and UAMS expression-based classification\(^40\, 41\) of these sub-clusters shows that clade D consists of CD-2 samples whereas clade F can be split into 2 distinct subclusters consisting of those in CD-2 (left branch) and those solely containing CD-1 samples (right branch). Clades G and H are on a separate branch from the majority of myeloma samples and consist of t(4;14) samples and HMCLs, respectively. Other abnormalities such as del(1p), gain 1q, del(13q), del(16q), del(17p) and del(22q) are not associated with nor define specific methylation subgroups nor drive clustering of the samples. Therefore, the main currently known
cytogenetic abnormalities which affect methylation in myeloma samples are the 
translocations (t(4;14), t(11;14) and t(14;16)) and hyperdiploidy. PCL samples did 
not segregate together, but remained within their respective cytogenetic clades, 
extcept for one outlying t(4;14) PCL sample which clustered with the non-malignant 
cell types.

Based on the above discussion it seems that the observed heterogeneity in global 
methylation within presenting myeloma samples seem to be due to the presence of 
different cytogenetic subgroups within the sample set. To identify the methylation 
differences driving the clustering we first split the presentation myeloma samples 
according to the IgH translocation, comparing each translocation group (t(4;14) 
n=15; t(11;14) n=35; t(14;16) n=7; t(14;20) n=3, unknown translocations n=15) to 
samples with no split IGH locus (n=66), as determined by FISH. Methylation β-
values were averaged across the samples within each group and analyzed as 
before. In this analysis the biggest differences were seen in the t(4;14) comparison 
with 2503 probes (9.4%) with increased methylation in the t(4;14) group compared to 
those with no split IGH locus and 302 probes with decreased methylation (Table 3). 
Fewer changes were seen when the t(11;14) (98 hyper- and 320 hypo-methylated), 
t(14;16) (26 hyper- and 19 hypo-methylated), t(14;20) (10 hyper- and 1 hypo-
methylated), and unknown translocations (no differences) were compared against 
those with no split IGH locus.

GO Term analysis of the 2503 probes (1881 genes) hypermethylated in the t(4;14) 
samples indicates methylation of genes involved in cell adhesion (147 genes, 
p=7.8x10^{-22}) and cell-cell signaling (128 genes, p=1.41x10^{-19}). Within the t(4;14) 
sample group several genes of interest were hypermethylated, some of which were 
validated by bisulfite specific PCR (Supplementary data). These include APC, 
adenomatous polyposis coli gene, which is a Wnt signaling pathway antagonist as 
well as being involved in cell adhesion, transcriptional activation and apoptosis. In 
the t(4;14) subset methylation of APC has a β-value of 0.36-0.41 compared to 0.03- 
0.07 in samples with no translocation, indicating it is hemi-methylated. PAX1, paired 
box gene 1, was solely methylated in t(4;14) samples and is a known methylation 
marker in ovarian cancer. Also, SOCS2 (0.45 vs. 0.14) and CDKN2A (0.7 vs. 0.34)
are hypermethylated in t(4;14) samples, but were not differentially methylated between MGUS and myeloma. To investigate this difference we separated the data by cell type and, within the myeloma group, by translocation. This analysis revealed that CDKN2A has significantly more methylation in the t(4;14) samples (p=0.0003) compared to samples with no split IGH, but that myeloma samples as a whole do not have significantly more methylation of CDKN2A than MGUS samples (p=0.518) (FIGURE 4). As such, hypermethylation of CDKN2A is significantly prognostic within the myeloma group (comparing β-values <0.3 vs. >0.3, p=0.03), but this is due to its association with the poor prognostic t(4;14) subgroup. This observation is different to the situation at CDKN2B, which lies adjacent to CDKN2A in the genome, is fully methylated in all myeloma cytogenetic subgroups and is significantly altered at the transition from MGUS to myeloma.

We investigated the effect of methylation of the genes in the t(4;14) samples further by comparing the data to gene expression data. t(4;14) expression data were compared to samples with no split IGH locus to generate a list of differentially expressed genes. From this analysis, 353 expression probesets were differentially expressed with the corresponding gene differentially methylated, of which 333 had lower expression in t(4;14) samples with increased methylation (Supplementary data). The genes with the highest expression fold changes are shown in Table 3 along with the corresponding methylation changes. C20orf103, which has similarity to LAMP domain proteins, was most differentially expressed with a 6.6-fold drop in expression in t(4;14) samples and a corresponding increase in methylation from 0.231 to 0.472. CD79A was also under-expressed in t(4;14) samples with an increase in methylation. This molecule has been found to have loss of protein expression in a subset of myeloma samples which also have low cyclin D1 expression.42 These are likely to be t(4;14) samples, which are cyclin D2 positive, in which methylation of CD79A has resulted in loss of protein in the cells. Other genes which are under-expressed in t(4;14) samples include GLTSCR2 (glioma tumor suppressor candidate) and SOCS2 (suppressor of cytokine signaling). Conversely, genes hypo-methylated in t(4;14) samples with increased expression include DNMT3A (DNA methyltransferase 3A), responsible for de novo DNA methylation, and IRS2 (insulin receptor 2) which is a tyrosine kinase receptor and mediates PI3K
signaling. However, the gene with the largest fold change was CTHRC1 which is implicated in promoting cell migration, osteoblastic bone formation, and activation of Wnt signaling pathways.43, 44

As t(4;14) samples most closely resemble HMCLs, with respect to DNA methylation, we investigated whether or not all HMCLs have a t(4;14) methylation profile. We discovered that t(4;14) HMCLs do have hypermethylation of genes which are hypermethylated in t(4;14) samples, but additionally at selected loci non-t(4;14) HMCLs also had hypermethylation. For example, C20orf103 hypermethylation is specific to t(4;14) myeloma samples and t(4;14) HMCLs, whereas hypermethylation of CD79A is specific to t(4;14) myeloma samples and all HMCLs, irrespective of translocation (see Supplementary data). This indicates that all HMCLs acquire hypermethylation of genes, in a similar fashion to t(4;14) myeloma samples and PCL samples, but methylation of some genes remain t(4;14)-specific.

Samples with cytogenetic abnormalities were also compared against those without the same abnormality. The abnormalities examined were del(1p32.3), gain 1q, del(13q), del(16q), del(17p), del(22q), hyperdiploidy, and any split IGH (Figure 3, Table 2). In comparison to translocation subgroups these cytogenetic abnormalities are associated with far fewer methylation changes, indicating that these abnormalities are not significantly associated with methylation. The largest methylation changes were noted in the hyperdiploid comparison where 134 probes were hypomethylated and 194 were hypermethylated compared to non-hyperdiploid samples. Of interest in the non-hyperdiploid hypomethylated gene list was CCND1, in which 6 probes show a decrease in methylation. However, this difference in methylation did not correlate with a difference in expression of CCND1. When split by translocation group the decrease in methylation was not limited to the t(11;14) subgroup, which over-express CCND1, but was present in all the major translocation groups. In addition, hypermethylation of CCND1 within the hyperdiploid samples did not correlate with translocation/cyclin D (TC) classification status, indicating that methylation of CCND1 at these CpG sites is not linked to expression of the gene.

From unsupervised clustering of all samples we found 2 distinct groups of hyperdiploid samples, designated clades B and C (Figure 3). There was no
disparity between the two groups with respect to the known cytogenetic markers, such as gain 1q or del(13q) which have been used by others to delineate hyperdiploid groups.\textsuperscript{4} When the overall survival of samples within clades B and C are analyzed we found a significant difference ($P=0.03$, median OS 44.8 vs. >70 months; Figure 5) indicating methylation may have both clinical and biological effects within the HRD patients. A comparison between these two clades on the entire dataset gives 3174 differentially methylated probes, but the majority of these have a difference <0.2, leaving 209 probes with a statistically significant difference in methylation. Of these, 11 are more heavily methylated in clade B, which has the poorer OS, and include \textit{CDKN2A} and \textit{CDKN2B} (cell cycle inhibitors) and \textit{MAPT} (microtubule associated protein).

\textbf{Changes in methylation from myeloma to PCL}  
In order to more clearly delineate the changes in methylation pattern occurring from myeloma to PCL we compared samples with the same translocation in the 2 disease states. When comparing t(4;14) samples at myeloma ($n=15$) and PCL ($n=2$) stages we identified 618 probes (566 genes) as being significantly differentially methylated (Supplementary data). These probes were exclusively hypermethylated in t(4;14) PCL compared to t(4;14) myeloma. When the same comparison was performed between t(11;14) myeloma ($n=35$) and t(11;14) PCL ($n=3$) we identified 566 probes (532 genes) which were differentially methylated, of which 560 were hypermethylated in PCL compared to myeloma. There were 71 genes commonly hypermethylated in both t(4;14) and t(11;14) PCL samples. Although the numbers of PCL samples are limited the consensus interpretation of these analyses is that there is an increase in methylation of CpG dinucleotides in the promoters of genes at the transition from myeloma to PCL. Pathway analysis of the genes in which methylation levels increase indicates that cytokine-cytokine receptor interaction and Jak/STAT signaling pathways are affected.

\textbf{Discussion}  
In this study we used a genome-wide array approach to interrogate the methylation status of over 27 000 CpG sites. Using a selection of cell types relevant to the multistep pathogenesis of myeloma we have been able to determine the methylation
changes which occur from normal plasma cells, through MGUS to presentation myeloma and plasma cell leukemia. We describe a clear distinction in methylation pattern between non-malignant cells (B cells, normal plasma cells and MGUS cells) compared with malignant plasma cells (presentation myeloma, PCL and HMCLs). We also go on to show that the major differences in methylation profile are found at the transition of MGUS to myeloma and myeloma to PCL.

At the transition from MGUS to myeloma the key feature is an overwhelming loss of methylation. Such global hypomethylation is associated with genome instability in many cancer cell types including colorectal, gastric, breast and chronic lymphocytic leukemia.45-48 Genome hypomethylation is frequently linked to altered chromatin structure, changes in DNA methyltransferase activity, loss of imprinting and increased frequencies of copy number abnormalities. The resulting aberrant transcription and chromosomal instability within clones is likely to contribute to disease progression and is one of the critical differences distinguishing MGUS from myeloma. These results are not unique to the experimental approach used in this experiment and are consistent with a previous analysis of non-CpG element (LINE-1, Alu and SAT-α) methylation levels which show decreased methylation levels in myeloma compared with controls.14

We also identified gene-specific hypermethylation at the transition of MGUS to myeloma involving 77 genes. Pathway analysis of the genes affected demonstrates involvement of developmental, cell cycle, and transcriptional regulatory pathways. The genes involved include CALCA, ONECUT2, GATA4, and CDKN2B, but not CDKN2A or CDH1, all of which are known to be methylated in other cancer types.49-51 The analysis of differentially methylated genes at the transition from MGUS to myeloma did not identify genes which have been shown to be methylated previously using methylation specific PCR. However, upon inspection of raw data we did find hypermethylation of some of these genes including CDKN2A, CDH1 and DCC in myeloma samples, but the spread of data points across the samples resulted in a p-value >0.05 (Supplementary Data).
At the transition from myeloma to PCL, rather than finding further hypomethylation as may have been anticipated, we found further gene-specific hypermethylation with 1802 genes showing an increase in methylation status. In particular we show re-methylation of genes involved in cell signaling and cell adhesion pathways which would be consistent with a mechanism whereby adhesion to the specialized bone marrow niche is impaired leading to bone marrow independent growth, allowing the tumor to enter the circulation and proliferate more freely. However, these data are based on a limited PCL sample cohort and require further investigation.

It is now widely accepted that there are two etiological subgroups of myeloma defined by the presence of either an aberrant class switch recombination event or hyperdiploidy. The relationship between these etiological subgroups and methylation status is important to understand. From unsupervised clustering of the 161 presentation myeloma samples it was clear that several independent methylation profiles exist within multiple myeloma: a t(4;14) group, 2 separate t(11;14) groups, and 2 separate hyperdiploid groups. These findings are in contrast to other chromosomal abnormalities such as del(1p32.1), gain 1q, del(13q), del(16q), del(17p), or del(22) which did not affect clustering of the samples. The most distinct of these methylation profiles belonged to the t(4;14) cytogenetic subgroup which showed more frequent hypermethylation of genes compared to the other subgroups. Cases with a t(4;14) over-express 2 potential oncogenes, MMSET and FGFR3, of which MMSET is of particular interest as it encodes a histone methyl transferase (HMT) which is known to methylate H3K36 and H4K20 residues and act as a transcriptional repressor.\textsuperscript{52, 53} Pathway analysis of the genes hypermethylated in t(4;14) myeloma indicates a similar phenotype to that of PCL samples. The similarity in methylation profiles between t(4;14) myeloma and PCL suggest that methylation may significantly contribute to the more aggressive clinical phenotype seen in both disease subtypes.

There were a limited number of genes in the t(4;14) subgroup whose methylation status change correlated with gene expression changes. This limited correlation may reflect additional influences from other inactivating methods such as deletions, mutations (through nonsense mediated decay), and upstream activation or silencing of transcription factors. This is especially true of cell cycle inhibitors such as
CDKN2A and CDKN2B which are also known to be deleted in myeloma samples.\textsuperscript{5} In particular we have observed deletions of CDKN2C, located at 1p32, which do not correlate completely with loss of gene expression. This is, however, mostly seems to reflect the fact that expression of this gene is lost in >95% of myeloma samples, presumably through alternate mechanisms.\textsuperscript{54} We believe that the correlation between methylation and gene expression will be equally complex.

An interesting observation in this study is our description of two specific subgroups of hyperdiploid myeloma based on their methylation profile. The TC classification of myeloma, while being a significant step forward, arguably does not deal adequately with the hyperdiploid group. In particular while the translocation subgroups have a distinct clinical outcome the hyperdiploid cases, amounting to 50% of the total, are apparently homogeneous in terms of their clinical outcome. Other groups have delineated groups of hyperdiploidy using mapping or expression array data. These analyses have been able to separate hyperdiploidy into groups based on the presence of 1q+,11+, NF-κB deregulation, proliferation or changes in the expression of cancer testis antigens.\textsuperscript{4, 55} Here we show that it is possible to split hyperdiploid samples into 2, based on their methylation profiles, and that each of these groups has a significant difference in OS. These 2 groups are independent of cyclin D expression levels, cytogenetic abnormalities and of presenting clinical features. Interestingly, we did not find a difference in the methylation status of NF-κB, proliferation or cancer testis antigens between the 2 groups, which may have led to the expression changes seen by other groups.

It is important to understand and develop models of how methylation changes may mediate the progressive transformation process from MGUS to myeloma. While it is clear from this analysis that genome-wide hypomethylation occurs at the transition from MGUS to myeloma, it is not so clear when gene-specific hypermethylation occurs. Hyperdiploidy and the main translocation groups (with the exception of t(4;14)) are present at similar frequencies in MGUS and presentation myeloma. As these cytogenetic abnormalities are the main defining feature of the methylation subgroups it may mean that these methylation groups also exist in MGUS cells, but we have not been able to analyze sufficient cases to demonstrate this adequately.
At the transition from MGUS to presentation myeloma secondary hits occur which result in genome hypomethylation. Whether these hits are mutations of genes controlling DNA methylation, such as DNA methyltransferases, or activation of transposable elements remains to be determined.
Acknowledgements

Research grants and financial support were received from the Leukaemia Research Fund, Cancer Research UK, the Bud Flanagan Research Fund, the United Kingdom Department of Health, Myeloma UK, Associazione Italiana Ricerca sul Cancro and the Biological Research Centre of the National Institute for Health Research at the Royal Marsden Hospital.

We also thank the staff at the Haematological Malignancy Diagnostic Service, Leeds, the LRF UK Myeloma Forum Cytogenetics Group, Salisbury, the Clinical Trials Research Unit, Leeds, UK, and the Cambridge Genomic Services, University of Cambridge, UK for processing of methylation arrays.

Authorship Contributions

BAW designed and performed research, analyzed data and wrote the paper
CPW analyzed data
LC performed research and analyzed data
EMS performed research
KDB performed research
AN provided biomaterial
FED designed research
FMR designed and performed research and analyzed data
GJM designed research and wrote the paper

Disclosure of Conflicts of Interest

The authors have no conflicts of interest to disclose.
References


Table 1: Variation in global methylation within cell types. Pair-wise XY scatter plots were used to generate Pearson correlation coefficients ($r^2$) between samples of the same cell type.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Range</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>B cell (6)</td>
<td>0.9716-0.99</td>
<td>0.9833</td>
</tr>
<tr>
<td>PC (3)</td>
<td>0.935-0.9397</td>
<td>0.9384</td>
</tr>
<tr>
<td>MGUS (4)</td>
<td>0.8828-0.9730</td>
<td>0.9351</td>
</tr>
<tr>
<td>Presentation Myeloma (161)</td>
<td>0.0816-0.9611</td>
<td>0.6731</td>
</tr>
<tr>
<td>PCL (7)</td>
<td>0.5439-0.8541</td>
<td>0.6456</td>
</tr>
<tr>
<td>HMCL (9)</td>
<td>0.2884-0.8834</td>
<td>0.472</td>
</tr>
</tbody>
</table>
Table 2: Number of probesets significantly altered in methylation status between samples with and without cytogenetic abnormalities. Methylation is relative to the control group.

<table>
<thead>
<tr>
<th>Cytogenetic abnormality</th>
<th>Hypomethylated</th>
<th>Hypermethylated</th>
</tr>
</thead>
<tbody>
<tr>
<td>t(4;14)</td>
<td>302</td>
<td>2503</td>
</tr>
<tr>
<td>t(14;16)</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>t(14;20)</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>t(11;14)</td>
<td>320</td>
<td>98</td>
</tr>
<tr>
<td>Unknown translocation</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hyperdiploidy</td>
<td>134</td>
<td>194</td>
</tr>
<tr>
<td>Gain 1q</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Del(1p)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Del(13q)</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>Del(16q)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Del(22q)</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>Del(17p)</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3: Genes in t(4;14) samples which are both differentially methylated and differentially expressed compared to samples with no split IGH locus. Expression values are log transformed and methylation numbers are β-values.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr.</th>
<th>Probe</th>
<th>t(4;14)</th>
<th>No split</th>
<th>Fold Change</th>
<th>P value</th>
<th>Probe</th>
<th>CpG island</th>
<th>t(4;14)</th>
<th>Difference</th>
<th>P value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C20orf103</td>
<td>20</td>
<td>219463_at</td>
<td>8.038</td>
<td>10.772</td>
<td>-6.656</td>
<td>0.008991</td>
<td>cg09119967</td>
<td>FALSE</td>
<td>0.472</td>
<td>0.231</td>
<td>0.241</td>
<td>0.001135</td>
</tr>
<tr>
<td>CD79A</td>
<td>19</td>
<td>1555779_a_at</td>
<td>7.168</td>
<td>8.408</td>
<td>-2.362</td>
<td>0.006872</td>
<td>cg04790874</td>
<td>FALSE</td>
<td>0.484</td>
<td>0.222</td>
<td>0.262</td>
<td>4.11E-05</td>
</tr>
<tr>
<td>FAM49A</td>
<td>2</td>
<td>208092_s_at</td>
<td>6.461</td>
<td>7.775</td>
<td>-2.487</td>
<td>0.009265</td>
<td>cg10106284</td>
<td>FALSE</td>
<td>0.287</td>
<td>0.167</td>
<td>0.12</td>
<td>0.033294</td>
</tr>
<tr>
<td>GLTSCR2</td>
<td>19</td>
<td>234339_s_at</td>
<td>10.806</td>
<td>12.216</td>
<td>-2.658</td>
<td>1.05E-07</td>
<td>cg16791686</td>
<td>TRUE</td>
<td>0.307</td>
<td>0.159</td>
<td>0.148</td>
<td>3.36E-05</td>
</tr>
<tr>
<td>GPX1</td>
<td>3</td>
<td>200736_s_at</td>
<td>9.846</td>
<td>11.311</td>
<td>-2.760</td>
<td>0.000267</td>
<td>cg06613840</td>
<td>TRUE</td>
<td>0.442</td>
<td>0.177</td>
<td>0.265</td>
<td>2.46E-08</td>
</tr>
<tr>
<td>GRM8</td>
<td>7</td>
<td>216992_s_at</td>
<td>4.346</td>
<td>4.896</td>
<td>-1.464</td>
<td>0.002383</td>
<td>cg09868882</td>
<td>FALSE</td>
<td>0.532</td>
<td>0.267</td>
<td>0.265</td>
<td>4.34E-08</td>
</tr>
<tr>
<td>MAB21L1</td>
<td>13</td>
<td>206163_at</td>
<td>3.443</td>
<td>4.138</td>
<td>-1.619</td>
<td>0.037692</td>
<td>cg05093686</td>
<td>TRUE</td>
<td>0.538</td>
<td>0.207</td>
<td>0.271</td>
<td>4.86E-07</td>
</tr>
<tr>
<td>MBP</td>
<td>18</td>
<td>209072_at</td>
<td>6.539</td>
<td>7.032</td>
<td>-1.407</td>
<td>0.000475</td>
<td>cg12559007</td>
<td>TRUE</td>
<td>0.841</td>
<td>0.411</td>
<td>0.43</td>
<td>2.61E-08</td>
</tr>
<tr>
<td>NME4</td>
<td>16</td>
<td>212739_s_at</td>
<td>8.007</td>
<td>8.658</td>
<td>-1.570</td>
<td>0.012902</td>
<td>cg18676162</td>
<td>TRUE</td>
<td>0.321</td>
<td>0.178</td>
<td>0.143</td>
<td>0.027662</td>
</tr>
<tr>
<td>OSBPL10</td>
<td>3</td>
<td>231650_s_at</td>
<td>5.979</td>
<td>6.995</td>
<td>-2.023</td>
<td>0.002011</td>
<td>cg15840985</td>
<td>TRUE</td>
<td>0.329</td>
<td>0.170</td>
<td>0.159</td>
<td>0.00022</td>
</tr>
<tr>
<td>RPS2</td>
<td>16</td>
<td>212433_x_at</td>
<td>13.275</td>
<td>14.107</td>
<td>-1.781</td>
<td>0.000313</td>
<td>cg18279742</td>
<td>TRUE</td>
<td>0.344</td>
<td>0.205</td>
<td>0.139</td>
<td>0.041014</td>
</tr>
<tr>
<td>SOCS2</td>
<td>12</td>
<td>203373_at</td>
<td>7.734</td>
<td>8.405</td>
<td>-1.593</td>
<td>0.041561</td>
<td>cg04797323</td>
<td>TRUE</td>
<td>0.450</td>
<td>0.140</td>
<td>0.31</td>
<td>0.000257</td>
</tr>
<tr>
<td>ACADVL</td>
<td>17</td>
<td>200710_at</td>
<td>10.904</td>
<td>10.024</td>
<td>1.840</td>
<td>0.000336</td>
<td>cg06630241</td>
<td>TRUE</td>
<td>0.637</td>
<td>0.285</td>
<td>0.352</td>
<td>7.18E-10</td>
</tr>
<tr>
<td>CTHRC1</td>
<td>8</td>
<td>225681_at</td>
<td>12.349</td>
<td>8.766</td>
<td>11.984</td>
<td>4.57E-05</td>
<td>cg19188612</td>
<td>TRUE</td>
<td>0.620</td>
<td>0.840</td>
<td>-0.22</td>
<td>0.00053</td>
</tr>
<tr>
<td>DNMT3A</td>
<td>2</td>
<td>222640_at</td>
<td>6.557</td>
<td>6.273</td>
<td>1.218</td>
<td>0.037539</td>
<td>cg21629895</td>
<td>FALSE</td>
<td>0.339</td>
<td>0.491</td>
<td>-0.152</td>
<td>0.042813</td>
</tr>
<tr>
<td>GALNT1</td>
<td>18</td>
<td>201724_s_at</td>
<td>8.838</td>
<td>8.011</td>
<td>1.774</td>
<td>0.041426</td>
<td>cg05714729</td>
<td>FALSE</td>
<td>0.212</td>
<td>0.383</td>
<td>-0.171</td>
<td>0.034844</td>
</tr>
<tr>
<td>GNPTG</td>
<td>16</td>
<td>224887_at</td>
<td>9.518</td>
<td>8.917</td>
<td>1.516</td>
<td>0.008847</td>
<td>cg18146152</td>
<td>TRUE</td>
<td>0.534</td>
<td>0.748</td>
<td>-0.214</td>
<td>0.000131</td>
</tr>
<tr>
<td>IRS2</td>
<td>13</td>
<td>209184_s_at</td>
<td>7.470</td>
<td>6.309</td>
<td>2.236</td>
<td>4.05E-05</td>
<td>cg14341579</td>
<td>TRUE</td>
<td>0.037</td>
<td>0.285</td>
<td>-0.248</td>
<td>0.011963</td>
</tr>
<tr>
<td>LRIG1</td>
<td>3</td>
<td>211596_s_at</td>
<td>6.472</td>
<td>5.568</td>
<td>1.872</td>
<td>0.009466</td>
<td>cg26131019</td>
<td>TRUE</td>
<td>0.057</td>
<td>0.310</td>
<td>-0.253</td>
<td>0.001657</td>
</tr>
<tr>
<td>Gene</td>
<td>Array IDs</td>
<td>Log2 Fold Change</td>
<td>Adj Fold Change</td>
<td>P-Value</td>
<td>Expression</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>----------</td>
<td>------------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LRP12</strong></td>
<td>8</td>
<td>220253_s_at</td>
<td>4.404</td>
<td>3.418</td>
<td>1.981</td>
<td>1.74E-08</td>
<td>cg09531892 TRUE 0.224 0.429 -0.205 0.03994 Low density lipoprotein-related protein 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>219631_at</td>
<td>5.758</td>
<td>4.230</td>
<td>2.883</td>
<td>8.52E-07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PELI1</strong></td>
<td>2</td>
<td>218319_at</td>
<td>12.686</td>
<td>11.809</td>
<td>1.836</td>
<td>0.033075</td>
<td>cg15309578 FALSE 0.299 0.421 -0.122 0.006695 Pellino homolog 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>232213_at</td>
<td>9.231</td>
<td>8.255</td>
<td>1.966</td>
<td>0.034372</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PTPRA</strong></td>
<td>20</td>
<td>213795_s_at</td>
<td>7.332</td>
<td>6.753</td>
<td>1.493</td>
<td>0.02016</td>
<td>cg03115886 FALSE 0.443 0.601 -0.158 0.025021 Protein tyrosine phosphatase receptor type A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>218319_at</td>
<td>12.686</td>
<td>11.809</td>
<td>1.836</td>
<td>0.033075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PTPRCAP</strong></td>
<td>11</td>
<td>204960_at</td>
<td>9.872</td>
<td>8.598</td>
<td>2.419</td>
<td>0.004979</td>
<td>cg05751148 FALSE 0.111 0.319 -0.208 0.036243 Protein tyrosine phosphatase receptor type C associated protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20792833</td>
<td>9.872</td>
<td>8.598</td>
<td>2.419</td>
<td>0.004979</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SBNO1</strong></td>
<td>12</td>
<td>218737_at</td>
<td>5.164</td>
<td>4.624</td>
<td>1.454</td>
<td>0.003605</td>
<td>cg04398275 FALSE 0.438 0.700 -0.262 2.95E-06 Strawberry notch homolog 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>213795_s_at</td>
<td>7.332</td>
<td>6.753</td>
<td>1.493</td>
<td>0.02016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SNRK</strong></td>
<td>3</td>
<td>209481_at</td>
<td>8.785</td>
<td>8.049</td>
<td>1.666</td>
<td>0.009873</td>
<td>cg040098913 FALSE 0.171 0.314 -0.143 0.031077 SNF related kinase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>227670_at</td>
<td>5.386</td>
<td>4.892</td>
<td>1.409</td>
<td>0.02804</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ZNF75A</strong></td>
<td>16</td>
<td>227670_at</td>
<td>5.386</td>
<td>4.892</td>
<td>1.409</td>
<td>0.02804</td>
<td>cg02825709 TRUE 0.322 0.575 -0.253 0.000649 Zinc finger protein 75A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>209481_at</td>
<td>8.785</td>
<td>8.049</td>
<td>1.666</td>
<td>0.009873</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For personal use only. www.bloodjournal.org From bloodjournal.org by guest on January 22, 2018. For personal use only.
Figure 1: XY scatterplots show the overall methylation differences between and within cell types. A, Two MGUS samples are very similar. B, A myeloma sample shows hypomethylation of probes compared to an MGUS sample. C, Two myeloma samples show heterogeneity of methylation. D, A PCL sample shows hypermethylation compared to a myeloma sample. E, Two PCL samples have equally heterogeneous methylation profiles. F, Summary of R² correlation values of samples within each cell type.

Figure 2: Hierarchical clustering of overall methylation in cell types can distinguish pre-malignant (green clade) and malignant phenotypes (red clade) (A). The number of probes differentially methylated in sequential steps of myeloma pathogenesis (B). Hypo, hypomethylated; Hyper, hypermethylated at transition.

Figure 3: Unsupervised hierarchical clustering of samples reveals discrete methylation groups based on cytogenetic abnormalities. Distinct clades are highlighted and labeled A-H. Samples are color coded with cytogenetic data: cell type, translocation, TC and UAMS classifications and colors are as per the key. Other cytogenetic data are coded as red for present, green for absent, white for no data. Heatmap key indicates beta value methylation level. UAMS subgroups: CD-1/CD-2, cyclin D1 subgroups; HY, hyperdiploid; LB, low bone disease; MF, MAF; MS, MMSET/FGFR3; PR, proliferation. TC subgroups: D1, cyclin D1; D2, cyclin D2; D3, cyclin D3; MAF, MAF/MAFB; MMSET, MMSET/FGFR3.

Figure 4: Boxplots of methylation beta values indicate t(4;14) specific hypermethylation of PAX1 (A), CDKN2A (B), APC (C), SOCS2 (D).

Figure 5: Two groups of hyperdiploid sample are defined by methylation status which has an impact on OS.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma

Brian A. Walker, Christopher P. Wardell, Laura Chiecchio, Emma M. Smith, Kevin D. Boyd, Antonino Neri, Faith E. Davies, Fiona M. Ross and Gareth J. Morgan