

Genome-wide DNA Methylation in Chronic Myeloid Leukaemia

Alexandra Bazeos

Imperial College London
Department of Medicine
Centre for Haematology

Thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy of Imperial College London
2015

Abstract

Epigenetic alterations occur frequently in leukaemia and might account for differences in clinical phenotype and response to treatment. Despite the consistent presence of the *BCR-ABL1* fusion gene in Philadelphia-positive chronic myeloid leukaemia (CML), the clinical course of patients treated with tyrosine kinase inhibitors (TKI) is heterogeneous. This might be due to differing DNA methylation profiles between patients. Therefore, a validated, epigenome-wide survey in CML CD34⁺ progenitor cells was performed in newly diagnosed chronic phase patients using array-based DNA methylation and gene expression profiling. In practice, the CML DNA methylation signature was remarkably homogeneous; it differed from CD34⁺ cells of normal persons and did not correlate with an individual patient's response to TKI therapy. Using a meta-analysis tool it was possible to demonstrate that this signature was highly enriched for developmentally dynamic regions of the human methylome and represents a combination of CML-unique, myeloid leukemia-specific and pan-cancer sub-signatures. The CML profile involved aberrantly methylated genes in signaling pathways already implicated in CML leukaemogenesis, including TGF-beta, Wnt, Jak-STAT and MAPK. Furthermore, a core set of differentially methylated promoters were identified that likely have a role in modulating gene expression levels. In conclusion, the findings are consistent with the notion that CML starts with the acquisition of a *BCR-ABL1* fusion gene by a haematopoietic stem cell, which then either causes or cooperates with a series of DNA methylation changes that are specific for CML.

Declarations

All clinical diagnoses referred to in this thesis were made by other physicians and not myself. Dr. Rob Lowe performed all bioinformatic analyses. Unless otherwise stated or referenced, all work presented in this thesis is my own.

Alexandra Bazeos

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they no do use it for the commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

On December 24th 2013, the day my dearest friend and mentor Professor John Goldman died, his daughter Cassie gave me her father's MD thesis "Leukaemic Stem Cells". On the first page in faded typewriter print was the epigraph "*Felix, qui potuit rerum cognoscere causas*" literally translated "Happy is he who has been able to understand the causes of things".

This is certainly true; the complexities of the epigenetic code in chronic myeloid leukaemia had provided the backdrop for hours of animated debate, usually late into the night with John diligently commenting on yet another paper I had presented him with. His enthusiasm was infectious, his intellect superhuman and his ability to interpret scientific data extraordinary. Moreover, John guided me, placed opportunities in front of me and showed me the doors that might be useful to open. I know with certainty that I will never have the privilege of meeting anyone like him again.

I am grateful to my supervisors Professors Vardhman Rakyen, Jane Apperley and Letizia Foroni. Vardhman, thanks for allowing me to spend so much time with your group at the Blizard and equipping me with the molecular techniques necessary to carry out this study. Letizia, without your practical support, I am not sure that this thesis would have ever reached its conclusion.

I would especially like to thank my friend and collaborator Dr. Rob Lowe for not only providing his bioinformatic expertise but also for contributing many excellent ideas during our time working together. It is true to say that without his analyses (and re-analyses), endless assistance and kindness; my project would genuinely not have been possible.

I must also express my greatest appreciation and respect to the patients many alive, some no longer, who in the face of adversity donated the blood and bone marrow samples that form the basis of this work. Thanks also, to my medical and nursing staff colleagues at The Catherine Lewis Centre, Hammersmith Hospital.

This work was generously supported by the Medical Research Council's Clinical Research Training Fellowship and Leuka (Registered charity no. 1154856).

Finally, I acknowledge my close friends and amazing family whose boundless love, encouragement and faith will forever inspire my human experience. Dad, Mum, Anna and Sophia, as long as we have each other, everything is possible.

I dedicate my thesis to John Goldman's delectable grandchildren Bodecia (aged 7), Juno (aged 7), Zephyrus (aged 6) and Penn (aged 6).

1 INTRODUCTION.....	13
1.1 Milestones in CML	13
1.2 Aetiology and epidemiology of CML.....	18
1.3 Clinical aspects of CML	19
1.4 Treatment and treatment related outcomes	25
1.5 Resistance to imatinib and other TKIs	30
1.6 Molecular biology of CML.....	34
1.7 Epigenetics.....	41
1.8 Hypothesis and aims of this thesis	57
2 MATERIALS AND METHODS	58
2.1 Materials.....	58
2.2 Methods	65
3 IDENTIFICATION OF A UNIQUE DNA METHYLATION SIGNATURE IN CML CELLS.....	94
3.1 Background	94
3.2 Hypothesis	97
3.3 Objectives	97
3.4 Results.....	98
3.5 Discussion	182
4 THE EFFECTS OF <i>BCR-ABL1</i> CONDITIONAL EXPRESSION ON THE GENOME-WIDE DNA METHYLATION PATTERNS OF U937P210BCR- ABL/C6 CELLS	187

4.1	Introduction	187
4.2	Hypothesis	189
4.3	Aims	189
4.4	Methods and overall experimental design	190
4.5	Results.....	194
4.6	Discussion	202
5	INVESTIGATING THE IMPACT OF ABERRANT DNA METHYLATION ON GENE EXPRESSION IN CML.....	205
5.1	Introduction	205
5.2	Hypothesis	208
5.3	Aims	208
5.4	Methods	209
5.5	Results.....	214
5.6	Discussion	227
6	CONCLUSION AND FUTURE DIRECTIONS.....	233
6.1	Conclusion	233
6.2	Future work.....	234

Table 1.1 Definition of accelerated and blast crisis according to the World Health Organisation (WHO) ^[35] and European LeukemiaNet (ELN) ^[36] expert panel	22
Table 1.2 Definitions of cytogenetic and molecular responses to treatment.....	27
Table 1.3 ELN Recommendations for the management of CML	29
Table 2.1 Fluorescent-labeled antibodies to characterise and enumerate haematopoietic stem and progenitor cells	71
Table 2.2 Immunophenotypic surface markers of haematopoietic stem and progenitor cells	72
Table 3.1 Biological and Clinical Characteristics of Hammersmith Hospital Patients (Discovery Cohort).....	101
Table 3.2 Healthy Control Characteristics.....	102
Table 3.3 Biological and Clinical Characteristics of SPIRIT 2 Patients (Validation Cohort)	103
Table 3.4 Confusion matrix (contingency table) for leave-one-out cross-validation procedure (repeated 1000 times per sample) on training set (Hammersmith Hospital).	140
Table 3.5 Contingency table showing the 759-feature classifier performance for prediction of response to TKI therapy, in the SPIRIT 2 validation set.	141
Table 3.6 Confusion matrix (contingency table) for leave-one-out cross-validation procedure (repeated 1000 times per sample) on training set (Hammersmith Hospital) using revised ELN criteria for response.....	145
Table 3.7 Contingency table showing the 581-feature classifier performance for prediction of response to TKI therapy, in the SPIRIT 2 (validation) cohort using revised ELN criteria for response.....	145
Table 3.8 CpGs showing >20% hypomethylation differences in CML-CCyR samples with respect to healthy control bone marrow (FDR < 0.01).....	173
Table 3.9 CpGs showing >20% hypermethylation differences in CML-CCyR samples with respect to healthy control bone marrow (FDR < 0.01)	177
Table 4.1 Culture conditions of U937p210BCR-ABL1/c6 cells	191
Table 5.1 KEGG functional annotation of genes aberrantly expressed in CD34 ⁺ CML-CP cells	220

Table 5.2 Differentially methylated CML-CP promoter CGIs demonstrating anti-correlation with gene expression changes (Spearman's correlation coefficient = -0.70, p-value = 1.4e-12).	226
---	------------

Figure 1.1 Typical blood film appearance of chronic phase CML	20
Figure 1.2 The Philadelphia translocation	34
Figure 1.3 The breakpoints in the <i>ABL</i> and <i>BCR</i> genes and structure of <i>BCR-ABL1</i> mRNA transcripts.....	36
Figure 1.4 Model of disease progression in CML	40
Figure 1.5 DNA methylation patterns in healthy and cancer cells	48
Figure 2.1 Gating strategy for haematopoietic stem and progenitor cells.....	73
Figure 2.2 Overview of the Illumina Infinium methylation assay	79
Figure 3.1 Overall strategy for CML patient study	100
Figure 3.2 Bisulfite conversion efficiency using array-based control probes for Hammersmith cohort and healthy samples	106
Figure 3.3 Density plots of the beta-values for the two Infinium assay types considered (red: Infinium I; blue: Infinium II; black: all probes) from Hammersmith cohort and health control samples.....	109
Figure 3.4 X-chromosome beta-value density plot in male and female study samples.....	110
Figure 3.5 Dendrogram showing unsupervised hierarchical clustering of CD34 ⁺ cell methylation profiles in CML samples from the Hammersmith cohort and healthy controls following HM450 preprocessing	115
Figure 3.6 Scatter plot displaying CML-CP differentially methylated probes (DMPs) represented as beta values (measured on a 0 - 1 scale).	119
Figure 3.7 Box plot of an exemplar CML-CP differentially methylated probe (DMP).....	120
Figure 3.8 The relative genomic distribution of CML-CP DMPs (compared to healthy controls).	121
Figure 3.9 Identification of aberrant co-methylation in CpGs flanking CML DMPs	122
Figure 3.10 Bisulfite conversion efficiency of SPIRIT 2 validation cohort.....	125
Figure 3.11 ComBat uncorrected and corrected dendrograms showing hierarchical clustering of all CML-CP samples from Hammersmith Hospital Patients (Discovery Cohort) and SPIRIT 2 Patients (Validation Cohort).....	127

Figure 3.12 Across batch and within batch comparison of separately normalised beta-values. .	129
Figure 3.13 Scatter plot of beta-value correlation between healthy controls	130
Figure 3.14 Across-batch validation of the 4,070 Hammersmith CML-CP DMPs.....	131
Figure 3.15 Marmal-aid heat map of the 4,070 CML-CP DMPs (vertical axis) separated into 6 clusters (horizontal axis) based on unsupervised hierarchical clustering.	136
Figure 3.16 Heat map representation (scale by row). 756 probes (759 features including age, sex and Sokal score) probe DNA methylation classifier Responder vs. Non-responder using Random Forest machine learning	139
Figure 3.17 FACS plots of CD34 ⁺ haematopoietic stem (HSC) and progenitor cell (HPC) populations from 3 healthy subjects following G-CSF mobilisation	148
Figure 3.18 FACS plots of CD34 ⁺ haematopoietic stem (HSC) and progenitor cell (HPC) populations from 4 CML-CCyR patients on imatinib therapy following G-CSF mobilisation	149
Figure 3.19 FACS plots of CD34 ⁺ haematopoietic stem (HSC) and Progenitor cell (HPC) populations from 4 untreated CML-CP patients (R, responders)	150
Figure 3.20 FACS plots of CD34 ⁺ haematopoietic stem (HSC) and Progenitor cell (HPC) populations from 4 untreated CML-CP patients (NR, non-responders)	150
Figure 3.21 The proportion of FACS sorted CD34 ⁺ stem and progenitor cell subsets in CML patients (CML-CP-R/NR and CML-CCyR) and healthy controls	151
Figure 3.22 Restoration of CML DMPs at the time of CML-CCyR.....	154
Figure 3.23 Probes representing methylation patterns unique to CML-CCyR samples.....	156
Figure 3.24 Probes representing methylation patterns common to all CML samples.....	157
Figure 3.25 Drug induced and residual disease DMPs in CML-CCyR samples.....	158
Figure 3.26 Dendrogram showing hierarchical clustering of CML-CCyR samples with other blood cell types using Marmal-aid.....	164
Figure 3.27 FACS plots comparing CD14 expression on 'CD14-like' CML-CCyR cells and healthy control samples	165
Figure 3.28 Cluster dendrogram comparing CD34 ⁺ cell methylation differences in G-CSF mobilised and bone marrow harvested samples	166
Figure 3.29 Cell-type composition methylation differences using Marmal-aid	168

Figure 3.30 Scatter plot illustrating the beta values of the CML-CCyR bone marrow DMPs at diagnosis.....	180
Figure 3.31 Scatter plot showing CpGs with >20% methylation differences in CML-CCyR-BM samples.....	181
Figure 4.1 Western blotting showing BCR-ABL1 induction in c6 cells upon doxycycline stimulation	192
Figure 4.2 Experimental design.....	193
Figure 4.3 <i>BCR-ABL1</i> mRNA expression ratio (%) by QT-PCR.....	194
Figure 4.4 Western blotting showing BCR-ABL1 (OC20), p-BCR-ABL1 (p-Tyr [4G10]) and p-CrkL protein expression. Lamin B1 provided a loading control.....	196
Figure 4.5 unsupervised hierarchical clustering (scaled by row) of probes showing variable DNA methylation β values of more than 10% in U937, c6 and CD14+ human monocytes.....	199
Figure 4.6 Unsupervised hierarchical cluster dendrogram on the top 100 DMPs.....	200
Figure 4.7 Comparison of the average DNA methylation levels (β -value) (y-axis) across the top 100 c6-On (+Dox) and c6-Off (-Dox) DMPs.....	201
Figure 5.1 Illumina whole-genome gene expression direct hybridization assay	210
Figure 5.2 HT-12 array pre-processing steps.....	212
Figure 5.3 Unsupervised hierarchical cluster dendrogram of HT-12 gene expression data	216

1 INTRODUCTION

1.1 Milestones in CML

1.1.1 19th Century

In 1845 John Bennett and Rudolf Virchow reported the first clinical descriptions consistent with diagnoses of CML [1, 2]. Six weeks apart, they independently published post mortem findings of patients who had become unwell over a number of months developing massive spleens, lymphadenopathy and abnormal consistency of their blood, which Virchow referred to as ‘weisses blut’, or white blood to reflect the pus-like collection present in the blood vessels.

By the close of the 19th Century, Otto Naegeli had identified the myelocyte as the granulocyte precursor [3] and from a therapeutic standpoint, treatment to address the symptoms associated with leukaemia were being trialed in the form of toxic arsenical preparations. The first documented use of ‘Fowler’s solution’ (potassium arsenite) in CML was by Arthur Conan Doyle during 1882 [4]. Conan Doyle used high doses of arsenic combined with iodine plus potassium chlorate and found this to be ‘remarkably efficacious’ for some months as palliative chemotherapy.

1.1.2 20th Century

The pathogenesis of CML had long been debated until 1960 when Peter Noel and David Hungerford discovered the very first somatic chromosome aberration in cells cultured from the blood of seven patients [5]. This ‘minute chromosome’ was subsequently found to be a truncated form of chromosome 22 and named the ‘Philadelphia chromosome’ or ‘Ph¹’ in recognition of the city where it has been detected and 1 to represent the starting point for future discoveries. However, further cytogenetic abnormalities were never found so the superscript was later lost.

The application of trypsin Giemsa banding (G-banding) in 1971 was a key innovation in cytogenetics for the mapping of chromosome abnormalities in haematological

malignancies. In 1972, Janet Rowley, a geneticist in Chicago was able to see that an abnormally elongated chromosome 9 accompanied the shortened chromosome 22 previously described [6]. The Ph chromosome was not the result of a deletion rather there existed a reciprocal translocation between the long arms of chromosomes 9 and 22 t(9;22) (q34;q11).

Further important breakthroughs in the molecular characterisation of CML include the demonstration by De Klein and colleagues in 1982 that c-Abl, the human homologue of the murine v-Abl oncogene usually located on chromosome 9, was translocated to chromosome 22 [7]. Subsequently, the reciprocal genetic material from a region of chromosome 22 to chromosome 9 was identified and named the 'breakpoint cluster region' (BCR) gene, giving rise to the expectation that BCR and ABL were associated in a way that led to the activation of ABL, a gene encoding a protein tyrosine kinase [8].

This was confirmed over the next few years when first Eli Canaani and his group detected messenger RNA (mRNA) BCR-ABL1 fusion gene transcripts in the cells of patients with CML [9] followed by Ben-Neriah, Witte and others in David Baltimore's laboratory identifying the p210 bcr-abl protein tyrosine kinase product with constitutive enzymatic activity [10].

Proof that p210^{BCR-ABL1} protein alone was sufficient to cause CML came from George Daley and David Baltimore in Boston. In 1988 they developed an *in vitro* transforming assay using retroviral constructs expressing the fusion gene in a bone marrow-derived interleukin 3-dependent Ba/F3 cell line [11]. This was followed in 1990 by the demonstration that a murine myeloproliferative disease resembling CML could be induced in lethally irradiated syngeneic mice receiving p210^{BCR-ABL1}-transduced bone marrow cells which could then be serially passaged to syngeneic recipients [12].

During the first half of the 20th century treatment approaches directed at CML were limited. Splenic irradiation provided symptom control but the effects were short-lived. Other therapies included, benzene and radioactive phosphorus as myelosuppressive treatments, splenectomy that frequently led to catastrophic perioperative haemorrhage and later leukapheresis, a specific type of apheresis, still available today as a mechanical means of cytoreduction.

The first cytotoxic drug used in CML was busulfan, an oral alkylating agent with a selective myelosuppressive mode of action. Introduced in 1953, busulphan offered a survival advantage to patients and largely replaced radiotherapy by 1960 [13]. Hydroxycarbamide, a ribonucleotide reductase inhibitor succeeded busulfan as the preferred first-line treatment in patients with CML from the late 1960s [14]. Patients treated with hydroxycarbamide (hydroxyurea, Hydrea®) experienced less therapy-related toxicities and benefitted from improved progression free survival. Furthermore, although both agents could induce haematological remissions in more than 70% of all CML patients, neither had the capacity to reduce the burden of Ph-positive cells in the bone marrow.

From 1983 until 2000, the majority of patients presenting in the early chronic phase of disease (CML-CP) received interferon alpha (IFN- α)-based therapy as the non-transplant treatment of choice [15]. The immunomodulatory effects of IFN- α conferred further survival benefits over conventional chemotherapy and resulted in the majority achieving a complete haematological remission including a fraction of patients who demonstrated attainment of durable cytogenetic responses.

Bone marrow and peripheral blood haematopoietic stem cell transplantation (HSCT) became a reality from the late 1970s when for the first time eligible patients were offered the hope of consistent cure through eradication of the Ph-positive clone.

Initial efforts focused on autografting patients in advanced phase disease following dose-intensive chemoradiotherapy with the intention of restoring chronic phase haematopoiesis [16]. With advances in the understanding of tissue matching and human histocompatibility typing (human leucocyte antigens, HLA), allogenic haematopoietic stem cell transplantation (allo-HSCT) programmes became feasible. By 1986 John Goldman in London and Don Thomas in Seattle had both successfully cured CML patients by means of HLA-identical sibling allo-HSCT with long-term, disease-free survival rates of 60–80% [17].

The first series of HLA-matched unrelated allo-HSCT for CML were disappointing. However, with the expansion of national donor transplant registries and improved donor/patient selection, the outcome of HLA-matched unrelated allo-HSCT became comparable to those of matched sibling procedures [18].

The demonstration that BCR-ABL1 constitutive kinase activity was the critical factor in cellular transformation provided the rationale for developing molecules to target and block this oncoprotein [19]. By 1998 the introduction of the orally administered ABL tyrosine kinase inhibitor, imatinib mesylate (formerly STI571) completely changed the disease landscape of CML. Developed by Ciba-Geigy (now Novartis) and Brian Druker in Portland, Oregon, imatinib entered phase I clinical trials in 1998 and phase II a year later [20].

1.1.3 21st Century

The efficacy of imatinib was remarkable and the drug was approved for use in May 2001 prior to the open label, phase III International Randomized Study of Interferon versus STI571 (IRIS) read out. The IRIS study recruited 1106 newly diagnosed chronic phase patients with CML who were randomised to receive imatinib or INF- α plus cytarabine.

At 19 months median follow-up, 95% of patients on imatinib compared with 56% receiving IFN- α had achieved complete haematological responses ($p < 0.001$). Furthermore, the estimated 12-month complete cytogenetic response rates were 68% versus 7% in the two respective groups with major molecular remissions of 40% for imatinib treated patients but just 2% for those treated with IFN- α /cytarabine ($p < 0.001$) [21].

The 5-year IRIS study update provided further evidence for the substantial benefits gained by patients receiving up front imatinib therapy. Survival was calculated to be 89% overall and 100% for those achieving a 3-log reduction in transcript levels by month 12, emphasising the importance of early optimal responses [22]. However, responding patients taking longer to achieve deep remissions had a low probability of relapse (less than 1% per year) by the 5th year on treatment with imatinib.

Despite the incredible success of imatinib, approximately 35% of patients starting therapy fail, owing to drug resistance or intolerable side effects. For some, but not all of these individuals, rapid responses have been achieved using more potent, second and third generation tyrosine kinase inhibitors (TKIs). To date, it is unclear whether the medium term benefits seen with nilotinib, dasatinib, bosutinib or ponatinib will

translate into overall survival improvement for all or selected subsets of patients whose imatinib treatment fails.

Over the next decade, study data will inform the most effective use of TKIs in the upfront setting to improve the longer-term outcome of newly diagnosed patients with CML. Furthermore, elucidating the biological mechanisms that underlie clinical heterogeneity will undoubtedly facilitate selection of the best treatment for a given patient.

The increasing prevalence of CML also highlights the need for strategies to eradicate CML and to produce molecular cures (allowing discontinuation of therapy) as opposed to functional ones (maintenance of close to normal lifespan with continuation of TKI therapy).

1.2 Aetiology and epidemiology of CML

Chronic myeloid leukaemia (CML) is a clonal, myeloproliferative disorder (MPD) with a consistent worldwide incidence of 1 to 2 per 100,000 annually [23-25]. Women are less frequently affected than men and show slightly higher survival rates. CML is rare in children, representing less than 5% of all paediatric blood cancers but in the western world constitutes 15% of all adult leukaemia cases with a median age at onset of 60-65 years. The age-adjusted incidence of CML is estimated to be between 10-15 cases/ 10^6 per year and while this remains unchanged, the prevalence is steadily rising to reflect the positive impact of TKI therapy on the survival of patients. The estimated prevalence of CML in the United States was approximately 70,000 in 2010, is projected to reach 112,000 in 2020 and will plateau to 35 times its incidence at 200,000 by 2050 [26].

The initiating event of CML is a chromosomal translocation t(9; 22), giving rise to the Philadelphia chromosome (Ph, 22q-) that results in the BCR-ABL1 fusion gene. The resulting oncoprotein is a continuously activated tyrosine kinase that alters a number of signal transduction pathways resulting in growth, survival and differentiation of haematopoietic cells. The aetiology of CML is unknown and there exists little evidence for genetic predisposition. One genome-wide association study involving Korean and European cohorts identified two novel loci, 6q25.1 and 17p11.1 that appeared to be linked to the development of the disease [27]. However, the lack of disease concordance between monozygotic twin pairs together with the fact that offspring of affected individuals do not have a higher incidence than the general population, suggest that CML is an acquired rather than inherited disorder [28].

There is a causal link between the development of CML and exposure to acute, high-dose ionizing radiation, as was the case for survivors of the atomic bomb detonations in Hiroshima and Nagasaki [29, 30], but the role of other environmental factors such as infectious agents and chemical compounds like benzene cannot definitively be ruled out [31].

1.3 Clinical aspects of CML

1.3.1 Natural History

Untreated, CML typically evolves through three phases. Most patients (~95%) present in the chronic phase (CP) of disease and often remain stable for 3-6 years before spontaneously progressing to blast crisis (BC), either directly or through a transitory 'accelerated' phase (AP). CML blast crisis is an aggressive form of acute leukaemia lasting 2-6 months and is invariably fatal. In keeping with the primitive stem cell origins of the disease, transformation to the advanced phase may be myeloid, lymphoid or undifferentiated/mixed. Myeloid blastic transformation constitutes up to two thirds of cases and carries a marginally worse prognosis than lymphoid [32].

1.3.2 Clinical presentation

In the developed world, up to one third of patients newly diagnosed with CML are asymptomatic and identified incidentally during routine blood testing. The remaining present with a vague history of insidious, non-specific symptoms including weight loss, fatigue, weakness, early satiety or low grade fevers [33].

Rarely, hyperleukocytosis ($\text{WBC} > 200 \times 10^9/\text{L}$) results in microvascular obstruction which clinically manifests as leukostasis, a medical emergency requiring leukapheresis to rapidly reduce the peripheral WBC count and risk of organ dysfunction or death.

Splenomegaly and purpura are the most common physical signs at presentation. Hepatomegaly is usually only detectable when patients present with more advanced phase and likewise lymphadenopathy is a feature of blastic transformation.

1.3.3 Histopathology and laboratory aspects

A diagnosis of CML is based on the histopathologic findings in the peripheral blood and bone marrow cells. The workup consists of a full blood count (FBC) with differential, a peripheral blood smear, a mandatory bone marrow aspirate and trephine biopsy together with bone marrow cytogenetics to demonstrate the presence of the Ph chromosome and real-time quantitative reverse transcriptase polymerase chain reaction (RQ-PCR) for BCR-ABL1 mRNA transcripts.

The typical blood count at diagnosis in chronic phase CML shows a leukocytosis and anaemia together with thrombocytosis, although the platelet count can vary and be normal or low. The blood film shows a neutrophilia, basophilia and usually eosinophilia. There is a marked increase in myeloid precursors including eosinophil and basophil myelocytes, promyelocytes and blast cells (Figure 1.1).

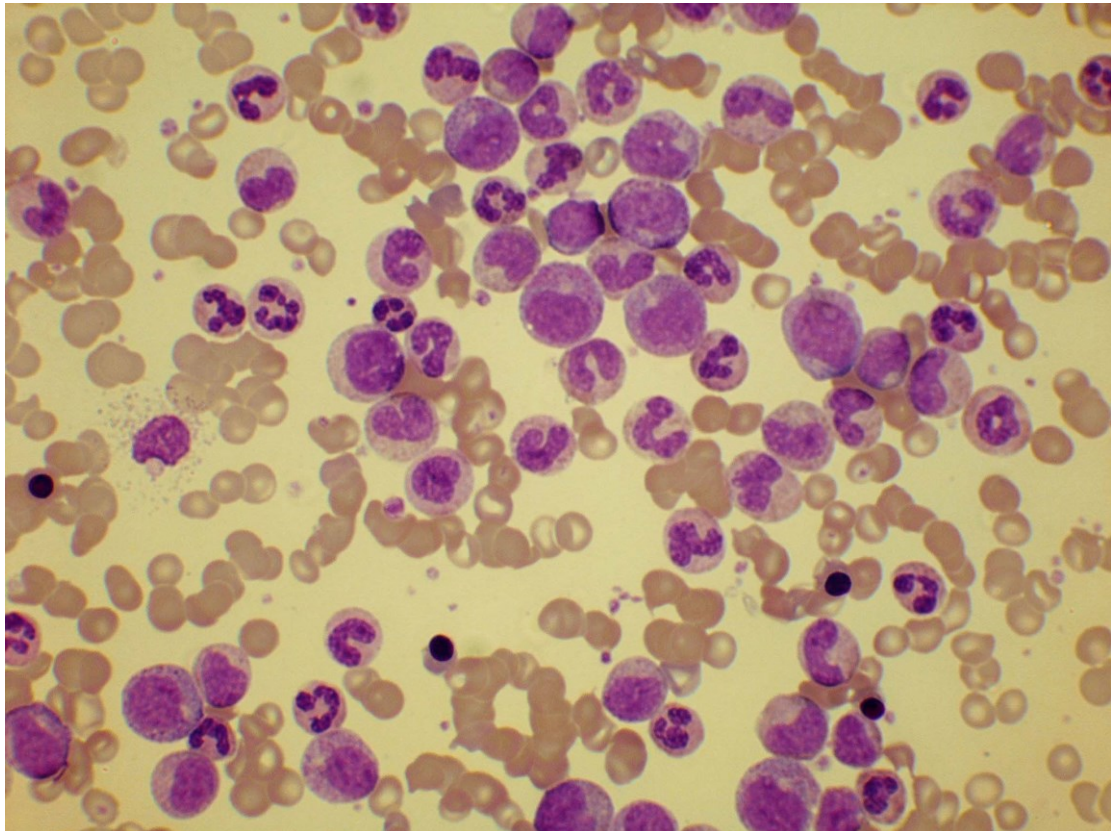


Figure 1.1 Typical blood film appearance of chronic phase CML

Courtesy of Professor Barbara J. Bain, St Mary's Hospital, Imperial College London

The bone marrow aspirate shows marked hypercellularity, with expansion of the granulocytic lineage and a myeloid:erythroid ratio of up to 25:1. Megakaryocytes are usually increased in number but are smaller in size than normal with nuclear hypolobation. Mild fibrosis is often seen in the reticulin stain. The above findings are reflected in the trephine biopsy, which apart from loss of fat cells retains normal bone marrow architecture.

Cytogenetic assessment of the peripheral blood or bone marrow aspirate detects t(9;22)(q34;q11.2) in 95% of all newly diagnosed CML cases. The technique is highly specific and can identify additional chromosomal abnormalities in the leukemic clone (e.g. trisomy 8 or isochromosome 17) that might impact on prognosis and guide future treatment choices. In the 5% of patients who lack the Ph chromosome but do have a suspected diagnosis of CML, half will have a cryptic BCR-ABL1 fusion gene on an apparently 'normal' chromosome 22 resulting in a falsely negative karyotype.

Fluorescence in situ hybridization (FISH) studies on peripheral blood can provide a rapid and sensitive alternative for detecting the presence of the BCR-ABL1 as well as more cryptic and complex BCR-ABL variants without the prerequisite for culture preparations and dividing cells in metaphase. Conversely, FISH does not provide the global genomic information gleaned from conventional chromosome banding analysis described above.

In addition, BCR-ABL messenger RNA (mRNA) transcripts from the blood or bone marrow can be detected by real-time, quantitative reverse transcriptase polymerase chain reaction (RQ-PCR). RQ-PCR is an extremely sensitive assay developed to monitor therapy response by measuring minimal residual disease. Although RQ-PCR is valuable at baseline as an additional means of disease confirmation its greatest value at diagnosis relates to characterising the chimeric transcript type prior to starting therapy [34].

1.3.4 Disease staging

Although the vast majority of patients present in early chronic phase, around 10-15% of newly diagnosed CML cases meet the criteria for advanced phase disease according to the World Health Organization (WHO) Classification of Tumours (2008) and European Leukaemia Net (ELN) guidelines (Table 1.1).

Similar to those with a known diagnosis of CML at the time of disease transformation, the management of *de novo* advanced phase patients is challenging and tends to be complicated by suboptimal responses to TKIs, resistance to chemotherapy and consideration for the need to proceed with allo-HSCT in suitable candidates.

Table 1.1 Definition of accelerated and blast crisis according to the World Health Organisation (WHO)^[35] and European LeukemiaNet (ELN)^[36] expert panel

	Accelerated phase		Blast crisis	
	WHO	ELN	WHO	ELN
Extramedullary involvement	-	-	Yes (except spleen)	Yes (except spleen)
Spleen size	Increasing on treatment	-	-	-
Platelet count (x10⁹/L)	<100 ¹ or > 1000 ²	<100 ¹	-	-
WBC count	Increasing ²	-	-	-
Blast cells in peripheral blood or bone marrow (%)	10-19	15-29	≥20 ³	≥30
Basophils in peripheral blood (%)	≥20	≥20	-	-
Additional cytogenetic aberrations	Clonal evolution on treatment	Clonal evolution on treatment	-	-

¹Unrelated to therapy side effects; ²unresponsive to therapy; ³or large foci of blast cells in the bone marrow or spleen

1.3.5 Risk stratification models in CML

Various attempts have been made to establish criteria at the time of diagnosis with prognostic (disease-related) and predictive (treatment-related) impact. In 1984, Sokal and colleagues applied multivariate analyses and regression calculations to 813 busulfan treated patients with chronic phase CML in an attempt to derive pre-treatment characteristics that would risk stratify according to their time to death [37]. Based on an algorithm that uses age, spleen size and blast cell percentage as continuous variables as well platelet numbers greater than $700 \times 10^9/L$, it was possible to segregate patients into low-, medium- and high-risk groups. To date the Sokal score remains the most widely applied relative risk tool in clinical studies.

The Hasford or the 'Euro' index was proposed in 1998 and developed using 1,573 INF- α treated CML patients across 12 European centres [38]. An extension to the Sokal system, the Hasford index incorporates the proportion of peripheral blood eosinophils and basophils to the relative risk calculation of each patient [39].

Remarkably the Sokal and Hasford classification systems have stood the test of time and gone on to demonstrate a high degree of reproducibility in a number of further studies involving INF- α based therapy and most recently imatinib [22].

However, both should be interpreted with care when applied to a patient on an individual patient basis since those enrolled prospectively on clinical trials often have higher performance status as per eligibility criteria and neither can be extrapolated for use in patients beyond early chronic phase. Furthermore, for an accurate score to be calculated all parameters must be obtained at diagnosis, prior to any treatment including short term hydroxycarbamide therapy.

While the Sokal and Hasford scoring systems were developed prior to the availability of TKIs, the EUropean Treatment Outcome Study (EUTOS) score was proposed after the introduction of imatinib and validated prospectively using data from 2,060 patients [40]. Using just two pre-treatment variables namely peripheral blood basophil percentage and spleen size it was possible to separate patients into low and high risk groups translating clinically to progression-free survival rates of 90% and 82% respectively after 5 years on imatinib treatment. Despite the higher predictive power of this prognostic index, the EUTOS score has not been widely used to risk stratify

patients in CML clinical trials and where it has been applied, results have so far been variable. Furthermore, the validity of EUTOS in the second generation TKI setting also remains to be tested [41].

1.4 Treatment and treatment related outcomes

1.4.1 Imatinib

Imatinib (IM) is a 2-phenylaminopyrimidine derivative with activity against ABL and Abelson-related gene (ARG) as well as members of the class III family receptor tyrosine kinases including c-KIT, platelet-derived growth factor receptor (PDGFR), and macrophage colony stimulating factor receptor (CSF-1R) [42, 43].

Preclinical studies demonstrated that IM competitively inhibits the ATP binding pocket of the SH1 kinase domain to stabilise the inactive, non-ATP-binding configuration of BCR-ABL1 and block the downstream effector, signal transduction pathways involved in oncogenesis.

1.4.2 Predicting response to imatinib/TKIs

Many individuals diagnosed in early CML-CP will respond to TKI therapy and have an excellent outlook, though a smaller but significant proportion fail to respond optimally and need treatment with alternative TKIs or proceed to transplant.

Despite the great success of imatinib and the second generation TKIs, there are no consistent biological or molecular biomarkers to detect progression or predict with reasonable accuracy how a patient with chronic phase CML (CML-CP) will fair in the medium to long term. In fact, prognostic factors at baseline have not advanced since busulfan and hydroxycarbamide were in widespread use.

The disease phase at diagnosis and consequently the time point in the disease when therapy is started, still remain the most significant factors in determining outcome. Patients in early CML-CP respond more consistently than those in accelerated phase (CML-AP), who in turn respond more reliably than those in blast crisis (CML-BC), where clinical benefit from TKIs is temporary.

Among the patients presenting in early chronic phase, the Sokal score remains a valuable clinical tool in predicting response where Sokal high-risk patients demonstrate lower rates and slower times to attaining complete cytogenetic responses. Since progression free survival (PFS) is strongly related to depth of response, patients in the high-risk category have poorer outcomes. Here, increasing the dose of imatinib

or switching therapy to second generation TKIs can improve responses but the differences remain significant.

This provides evidence for intrinsic heterogeneity existing in leukemic cells reflecting unknown differences [44]. When these are combined with host factors such as an individual patient's pharmacodynamic profile or immunological response to leukaemia cells, it's clear the situation is complex.

Overall, 10-15% of CML patients who commence imatinib in CML-CP display primary drug resistance and do not respond from the outset, failing to reach a complete haematological response (CHR) by 3 months (2-3% patients) or major cytogenetic response (MCyR) by month 6 (approximately 15%) [22]. This proportion increases when individuals are diagnosed in more advanced phase disease and was seen more frequently in CML-CP patients who received IM after IFN- α failure or intolerance [45]. The biological and molecular mechanisms underlying primary resistance are unclear but hardly ever due to kinase domain mutations.

Secondary or acquired resistance is more common and refers to the scenario where an initial haematological or cytogenetic response is obtained but subsequently lost accompanied by rising BCR-ABL transcripts with or without progression to advanced phase disease. Altogether, the incidence of acquired resistance in patients who are diagnosed and start IM therapy in the chronic phase is approximately 30-35% and declines with time. Notably, patients in CCyR for longer than 5 years have a favourable prognosis and carry a low risk of relapse [22, 46]. Like primary resistance, secondary resistance was diagnosed more frequently in CML-CP patients with prior exposure to IFN- α and is considerably higher in accelerated phase (73%) and blast crisis (95%).

1.4.3 Monitoring response to therapy

Response to treatment is a powerful prognostic factor and can be defined by laboratory criteria at three levels; haematological, cytogenetic and molecular. The normalisation of the peripheral blood counts, absence of immature granulocytes and lack of palpable spleen marks a haematological response. Cytogenetic response is considered complete in the absence of detectable Ph positive cells by conventional cytogenetics and partial when between 1-35% are found. Major cytogenetic responses

comprise complete and partial responses (Table 1.2). A major molecular response (MMR) measured by real time quantitative (RQ-PCR) is usually defined as a three-log reduction in BCR-ABL transcript levels or <0.1% on the converted international scale (IS).

Ph-positive marrow metaphases (%)	Designated response
0	Complete cytogenetic response (CCyR)
1-35	Partial cytogenetic response (PCyR)
36-65	Minor cytogenetic response
66-95	Minimal cytogenetic response
>95	None
Percentage based on an absolute minimum of 20 assessable metaphases. CCyR and PCyR may be grouped together as 'Major cytogenetic responses' (MCyR).	
Ratio of <i>BCR-ABL/ABL</i> (%)	
<0.1	Major molecular response (MMR)
CCyR usually corresponds to approximately a two-log reduction in <i>BCR-ABL</i> transcript levels or 1% on the international scale (IS). MMR is usually defined as a three-log reduction in transcript levels or 0.1% on IS	

Table 1.2 Definitions of cytogenetic and molecular responses to treatment

1.4.4 European LeukemiaNet recommendations

One major objective in the management of CML is to recognise and manage therapy failure at the earliest opportunity in a patient's disease course in order to prevent progression. In 2005, the European LeukemiaNet (ELN) assigned a series of criteria for assessing prognosis using IM therapy response as a function of time. An expert panel defined failure and suboptimal response for early CML-CP patients on standard dose (400 mg per day) IM [36]. Although these original recommendations were not based on published evidence, subsequent analyses confirmed that the guidelines were valid [47]. The ELN recommendations have since evolved and most recently updated in 2013 [48] (Table 1.1).

Responses are defined as 'optimal' or 'failure'. An optimal response indicates satisfactory ABL kinase domain inhibition and is associated with durable responses and favourable long-term outcomes. In the IRIS study, achievement of MMR one year following the start of TKI therapy was associated with 100% progression free survival at 5 years [22].

Patients in the ELN therapy 'failure' category are at high risk of transformation to advanced phase disease. Here, changing to an alternative second line TKI at the earliest opportunity is crucial to limit the risk of progression and death. Between optimal and failure, an intermediate category exists. Previously referred to as 'suboptimal' and since labeled 'warning', a warning flag at a given time point informs the clinician that although ongoing benefit might be gained from current TKI therapy, long term value is equivocal and outcome less favorable so careful monitoring is essential to allow expedient changes where necessary.

More recently, molecular response at 3 months has been shown to accurately predict longer term outcome. Marin *et al* showed that *BCR-ABL1* transcripts >10% taken 3 months after the start of therapy is the most accurate way of identifying those patients with CML-CP destined to fare less well, thus allowing early clinical intervention [49].

Table 1.3 ELN Recommendations for the management of CML

Time on treatment (months)	Optimal	Warning	Failure
Baseline		High risk patient ¹ and/or Major CCA ²	
3	BCR-ABL ^{IS} ≤10% and/or Ph ⁺ ≤35% (PCyR)	BCR-ABL ^{IS} >10% and/or Ph ⁺ 36-95%	No CHR and/or Ph ⁺ >95%
6	BCR-ABL ^{IS} <1% and/or Ph ⁺ 0% (CCyR)	BCR-ABL ^{IS} 1-10% Ph ⁺ 1-35%	BCR-ABL ^{IS} >10% and/or Ph ⁺ >35%
12	BCR-ABL ^{IS} ≤0.1% (MMR)	BCR-ABL ^{IS} 0.1-1%	BCR-ABL ^{IS} >1%*Ph ⁺ >0%
At anytime	MMR or better	CCA	Loss of any previously attained response confirmed on repeat testing and/or KD mutations and/or CCA

The response definitions apply to all patients (CML-CP, CML-AP, and CML-BC) receiving imatinib OR any other TKI as first-line OR second-line (in cases of prior drug intolerance).

¹Assessed by Sokal-, Euro- or EUTOS-Score; ²major clonal chromosome abnormality (CCA) include trisomy 8, 2nd Ph⁺, isochromosome 17, and trisomy 19; ^{IS} On the International Scale; KD= Kinase Domain (not ABL1 polymorphisms).

Adapted from Baccarani et al [48]

1.5 Resistance to imatinib and other TKIs

Resistance to imatinib and other TKIs can be classified in a number of ways including whether the mechanisms involved are BCR-ABL dependent or independent. BCR-ABL1-dependent factors include the emergence of leukemic clones containing BCR-ABL tyrosine kinase domain (KD) mutations and overexpression of the BCR-ABL oncoprotein resulting from BCR-ABL1 gene amplification. Mechanisms occurring independently of BCR-ABL1, are less well defined, and in practice are likely to be multifactorial. They include genetic polymorphism in genes that are involved in pharmacokinetic and pharmacodynamic processes, such drug transporters and activation of oncogenic pathways downstream of BCR-ABL1.

1.5.1 BCR-ABL-dependent mechanisms of IM resistance

1.5.1.1 *ABL kinase domain (KD) point mutations*

ABL1 KD mutations occur in up to 50% of patients who develop secondary resistance to IM while they are very rare in patients with primary resistance [50, 51]. Clinically relevant mutations can either disrupt key contact points between BCR-ABL1 and IM or drive the active conformation of BCR-ABL1 to which it cannot bind resulting in the reactivation of BCR-ABL1 kinase activity within Ph⁺ cells [52]. In excess of 100 different mutations have been described throughout the KD with a tendency for clustering within sub-regions including the imatinib-binding site, the ATP binding loop (P-loop, 45-50%), the activation loop (A-loop, 5%), catalytic domain (20%) and the carboxy terminal (16%) [53]. Not all mutations result in the same biochemical and clinical sequelae; and while some lead to highly resistant phenotypes, others remain comparatively sensitive to IM *in vitro* and clinical resistance may potentially be overcome by IM dose increase [54, 55].

The T315I mutation is located within the IM binding site at the ‘gatekeeper position’ and not only prevents the hydrogen bonding but also allosterically obstructs imatinib drug binding. This leads to markedly lower (200-fold) biochemical sensitivity to IM and well as nilotinib and dasatinib. Survival of patients with the T315I mutation mostly depends on the stage of the disease but even for CML-CP patients, the

emergence of this mutation profoundly compromises overall survival and PFS [56, 57].

1.5.1.2 BCR-ABL1 overexpression

Genomic amplification of *BCR-ABL1* leading to over expression of the BCR-ABL protein has been investigated as a process that could account for a proportion of IM resistant cases. The persistent activity of BCR-ABL1 kinase is crucial for CML establishment, maintenance and disease progression when transcript levels always rise at the time of CML-BC. BCR-ABL1 confers escape from apoptosis, encourages growth factor independence, promotes genomic instability and supports clonogenicity in a dose dependent manner [58]. Although the mechanisms supporting *BCR-ABL1* gene amplification are not fully defined, the process is likely to be driven by the proliferative advantage gained from increased kinase activity. Here, excessive transcripts levels and resultant target molecules means that standard therapeutic doses of imatinib may be ineffective or suboptimal and possibly drive the expansion of IM-resistant mutant subclones [59].

1.5.2 BCR-ABL independent mechanisms of IM resistance

1.5.2.1 Drug transporters

The probability of achieving an optimal treatment response in CML is most likely to be influenced by the variable capacity of a patient's leukemia cells to accumulate intracellular imatinib (IM). The transport of IM into cells is an active process mediated by the human organic cation transporter (hOCT1) protein, a membrane associated ATP-dependent protein channel [60].

High *hOCT1* expression in blood-derived white blood cells (WBC) collected at diagnosis predicted for a 3-log reduction in *BCR-ABL1* transcript levels (major molecular response, MMR) and also for a greater degree of transcript reduction in patients who had achieved complete cytogenetic responses (CCyR) [61]. Investigators also showed that elevated *hOCT1* mRNA expression in diagnostic peripheral blood mononuclear cells (PBMC) predicted for a better response to IM [62]. Interestingly, dasatinib and nilotinib do not require active hOCT1 transport into leukemic cells since cellular influx of these agents is by the passive route [63, 64].

In contrast to drug influx pump proteins, multi drug resistance to the TKIs including the second generation agents could also be mediated by energy depended drug efflux pumps of which the P-glycoprotein (Pgp) (encoded for by the *ABCB1/MRD1* gene) is best characterized [60]. IM and the other TKIs are substrates of Pgp and Pgp might lead to decreased intracellular IM drug concentrations, eventually resulting in therapy failure.

Intracellular levels of IM are lower in Pgp-expressing cells and similarly, IM resistant CML cell lines exhibit *ABCB1* gene overexpression [65]. From a translational perspective, clinical studies have so far failed to convincingly demonstrate a similar association. Inhibition of *ABCB1* does not appear to enhance the effect of IM against BCR-ABL1 activity. Furthermore, although *ABCB1* mRNA expression can be high in CML progenitor cells, the role this efflux pump plays within the primitive CD34⁺ compartment is unlikely to be significant.

1.5.2.2 Imatinib drug levels and therapy adherence

Like many oral medications, variations in the extent to which patients take imatinib/TKIs as prescribed together with the absorption, metabolism, plasma protein binding, and passage into the target cell all contribute to the overall drug plasma levels.

Although most patients are commenced on a standard IM dose of 400 mg daily, a number of studies have identified a relationship between adequate trough plasma drug levels (>1000 ng/mL) and improved long term outcomes [66]. Conversely, suboptimal imatinib pre-dose (trough) plasma concentrations can correlate with a lower probability of achieving a CCyR and higher rates of discontinuation [67].

Low trough imatinib levels could be an indication to escalate the prescribed IM dose with the objective of improving responses in individual patients but the value of this approach has not been validated and may have a negative impact on a patients overall adherence to therapy. One study by Marin and others showed that adherence in patients who had their imatinib dose increased was significantly lower than in the patients who remained on a dose of 400 mg per day [61].

The same study showed that the overall achievement of a molecular response was related to imatinib adherence where the probability of MMR for patients with

adherence rates $\leq 90\%$ was 13.9%, compared with 93.7% for the those taking more than 90% of their prescribed doses correctly ($P < 0.001$). Similarly, the probability of a 4-log reduction was 4.3% versus 76% ($P < 0.001$), and the probability of CMR was 0% versus 43.8% ($P = 0.002$).

1.5.2.3 Alternative signaling pathways

BCR-ABL1 activates multiple signal transduction pathways involved in growth, survival and differentiation of haematopoietic cells including Ras, MAPK, STAT, JNK, PI-3 kinase, NF-kB and c-MYC [68]. However, clonal evolution and acquired TKI resistance could be mediated via alternative, BCR-ABL-independent signal transduction pathways including the structurally related, non-receptor SRC family kinases (e.g. Lyn, Fyn, Hck and Lck).

LYN kinase is activated indirectly by BCR-ABL1 upon activation of JAK2 kinase [69]. BCR-ABL has been shown to co-immunoprecipitate with and activate Hck and Lyn in myeloid cells [70] [71]. Over expression and activation of *LYN* was reported in IM-resistant CML cell lines and LYN suppression by a Src kinase inhibitor resulted in decrease proliferation and increased survival of the IM-resistant but not the sensitive parental cell line. Moreover, BCR-ABL1-independent activation of LYN has been identified in BCR-ABL1 unmutated patients with acquired resistance to IM [72]. Similarly, in CML cells driven by high Lyn expression, inhibition of BCR-ABL kinase by IM alone was insufficient to fully block BCR-ABL-mediated signaling [73].

1.6 Molecular biology of CML

1.6.1 The Philadelphia chromosome

The $t(9;22)(q34;q11)$ translocation juxtaposes the 3' end of c-ABL from the long arm of chromosome 9 to the 5' region of the breakpoint cluster region (BCR) gene on the long arm of chromosome 22, resulting in one shortened chromosome 22 (22q-), the Ph chromosome and one elongated chromosome 9 (9q+). The molecular consequence of the Ph chromosome is the generation of *BCR-ABL1* oncogene that encodes the constitutively active protein kinase BCR-ABL1 (Figure 1.2).

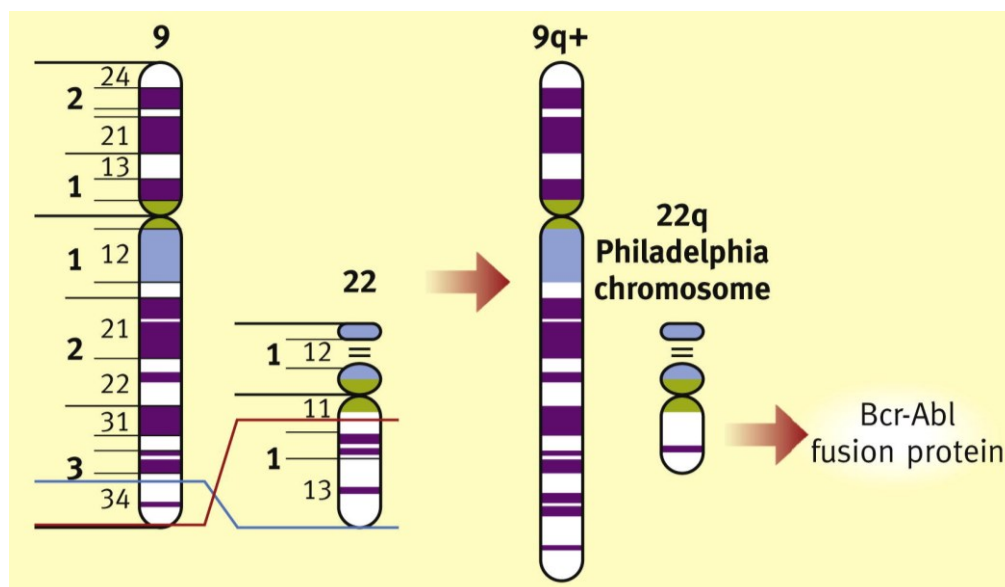


Figure 1.2 The Philadelphia translocation

Breakpoints in the long arm of chromosome 9 (q34) and chromosome 22 (q11) results in the reciprocal translocation leading to an elongated chromosome 9q+ and a shortened 22q-, the Philadelphia (Ph) chromosome. As a result, a *ABL-BCR* chimeric gene is formed on the derivative chromosome 9q and *BCR-ABL1* on the derivative chromosome 22. *Figure courtesy of Dr Hughes de Lavallade, Kings College London [74]*

1.6.2 The *BCR-ABL1* gene

The *ABL1* gene on chromosome 9 encodes a non-receptor protein-tyrosine kinase, c-ABL consisting of 11 exons including 2 alternative 5' first exons, and a first intron of approximately 250 kilobases (kb). On the Ph chromosome, the DNA sequences adjacent to the 5' of ABL1 are derived from chromosome 22 sequences, and were originally named the breakpoint cluster region, 'BCR'. Subsequent characterisation around this locus established that the bcr region was located within a 25 exon cytoplasmic phosphoprotein-coding gene which became named the *BCR* gene. The 5 exons spanning the bcr region were originally referred to as exons b1-b5. More latterly they became relabelled as exons 12-16 of the major breakpoint cluster region (M-bcr) [75].

The breakpoints within *ABL1* map upstream of the second exon, usually between exons 1b and 1a and depending on the location, alternative splicing produces fusion transcripts with either e13a2 (b2a2, according to the original BCR exon nomenclature) or e14a2 (b3a2) junctions both of which give rise to a 210 kDa fusion protein; p210 BCR-ABL1 (p210^{BCR-ABL1}) (Figure 1.3). In the majority of Ph⁺ B-ALL and rarely in CML the *BCR* breakpoint lies between exons e2' and e2 (designated the minor breakpoint cluster region m-bcr) to yield an e1a2 transcript which is translated into a 190 kDa protein, p190^{BCR-ABL1} [76, 77]. Occasionally breaks in the micro bcr region (μ -bcr) 3' of e19 generates e19a2 transcripts and a 230 kDa protein, p230^{BCR-ABL1} which can be detected in some patients with chronic neutrophilic leukemia [78].

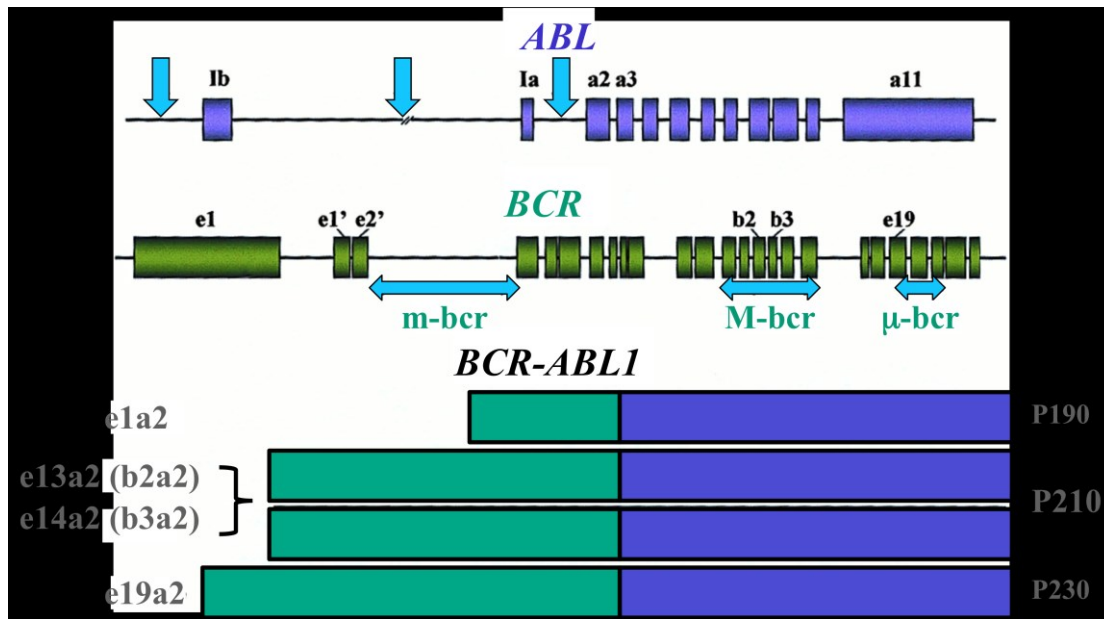


Figure 1.3 The breakpoints in the *ABL* and *BCR* genes and structure of *BCR-ABL1* mRNA transcripts.

(Adapted from Deininger, M.W.(2000) [76])

1.6.3 BCR-ABL1 kinase signalling pathways

The constitutive tyrosine activity of BCR-ABL1 initiates several downstream signalling cascades including RAS/MAPK, PI3K/AKT and JAK/STAT the net effect of which is a number of proliferative and anti-apoptotic cellular responses.

The phosphorylation of BCR tyrosine residue 177 (Tyr177) of is crucial for BCR-ABL1-mediated activity [79]. As a result, this generates a high affinity binding site for the SH2 domain of growth factor receptor-bound protein 2 (GRB2) which consequently recruits son of sevenless (SOS), a guanine–nucleotide exchanger of RAS along with GRB2-associated binding protein 2 (GAB2) [80]. The resulting GRB2/GAB2 complex is essential to fully activate the RAS/ERK pathway, engage SHP2 and recruit phosphatidylinositol 3-kinase (PI3K)/Akt [81, 82].

1.6.4 The CML stem cell

Early evidence that chronic myeloid leukemia (CML) was a clonal disease arising from haematopoietic stem cells (HSC) came in the early 1950s when Dameshek showed that bone marrow cells from CML patients displayed high proliferative capacity and had the ability to differentiate down all myeloid lineages [83]. This was subsequently followed by the demonstration that the Ph chromosome and *BCR-ABL1* gene were present in these multiple myeloid lineages [84]. The functional existence of leukaemia stem cells (LSC) was first highlighted in acute myeloid leukaemia (AML) when it was shown that only the primitive $CD34^{+}CD38^{-}$ subpopulation of cells were capable of initiating disease in non-obese diabetic/severe combined immunodeficiency (NOD/SCID) mice recipients [85].

1.6.4.1 Quiescence and persistence of CML stem cells

Initial support for the notion that CML LSCs had repopulating capacity, came from clinical observations that non-irradiated peripheral white blood cells collected from patients with CML were capable of short term engraftment and generating Ph^{+} progeny in severely neutropenic recipients [86]. Later, this was explained by the presence of a high proportion of mobilised LSCs in the peripheral blood of CML donors consisting of two groups; a small, highly quiescent population of primitive Ph^{+} cells in G0 and downstream progenitors with proliferative potential and sensitivity to differentiation cues [87, 88]. Most recently, functional studies such as long-term culture-initiating cell (LTC-IC) assays have detected the persistence of $BCR-ABL1^{+}$ LSCs cells with self-renewal and long-term repopulating capacity in CML patients on TKI therapy in deep molecular remissions [89, 90]. It is the survival of these LSCs that are thought to be responsible for the relapse of chronic phase disease in around 50% of patients withdrawing from TKIs in the controlled clinical trial setting following attainment of molecular remission [91]. Accordingly, understanding the extreme resistance exhibited by $BCR-ABL1^{+}$ LSCs to chemotherapy, radiation and TKIs is essential for delivering effective therapies to this elusive compartment to fully eradicate disease.

1.6.4.2 CML stem cell maintenance pathways

In one study Corbin and colleagues used phenotypic cell surface markers to identify populations enriched for stem cell activity and investigate imatinib sensitivity in the

primitive stem versus committed progenitor cell compartments [92]. It was possible to demonstrate that although imatinib inhibited BCR-ABL1 to the same extent between the two subgroups, in the presence of exogenous cytokines the viability and survival of quiescent cells was independent of BCR-ABL1 activity.

Several embryonic development signaling pathways have been associated with stem cell maintenance in leukaemia pathogenesis where they confer growth and survival advantages to LSCs.

The Wnt/ β -Catenin pathway regulates stem cell pluripotency and cell fate decisions during development and is aberrantly regulated in CML [93, 94]. Using a conditional mouse model, Zhao and others demonstrated that the absence of β -catenin reduced the severity of murine CML disease as a result of impaired LSC self-renewal capability [95]. Conversely, increased levels of β -catenin and the transcriptional co-activator and DNA-binding protein lymphoid enhancer factor/T-cell factor LEF/TCF can be detected in granulocyte-macrophage progenitors (GMPs) isolated from patients in blasts crisis where they display increased self-renewal capabilities and stemness *in vitro*, as assessed by serial replating assays, compared with normal counterparts [96].

The Hedgehog (Hh) pathway is essential for normal embryonic development and contributes to adult tissue renewal and maintenance [97]. In haematopoietic stem cells (HSC), this pathway appears to be important for short term functioning and stress-induced responses [98].

Hh proteins (ligands) initiate signaling through binding to the patched-1 (PTCH1) receptor. Smoothened (Smo), also a transmembrane receptor is negatively regulated by PTCH1. The binding of Hh to PTCH results in the accumulation of SMO and downstream signal transduction including the release and activation of the transcriptional effector proteins GLI [99].

Several reports have identified gene mutations and abnormal activation of the Hh pathway in CML. In a CML murine model, BCR-ABL1 positive cells displayed SMO overexpression both at the mRNA and protein levels. This appeared to be relatively oncogene independent since treatment with imatinib or nilotinib led to only minor reductions in *SMO* and *GLI* transcript levels [100].

1.6.5 Mechanisms of disease evolution and progression in CML

The processes underlying transformation to advanced phase disease (CML-AP and CML-BC) are complex and only partially understood due to the lack of suitable animal models able to reproduce disease evolution and progression from chronic phase. While the Ph chromosome is likely to be the initiating event in CML, the transition to blast crisis appears to involve many other chromosomal, genetic and epigenetic changes [75, 101, 102]. In CML-CP, LSCs are found in the self-renewing Lin⁻CD34⁺CD38⁻ population but during advanced phase disease, committed Lin⁻CD34⁺CD38⁺ bone marrow granulocyte-macrophage progenitors (GMPs) acquire self-renewal properties in vitro and gain LSC status (Figure 1.4) [96].

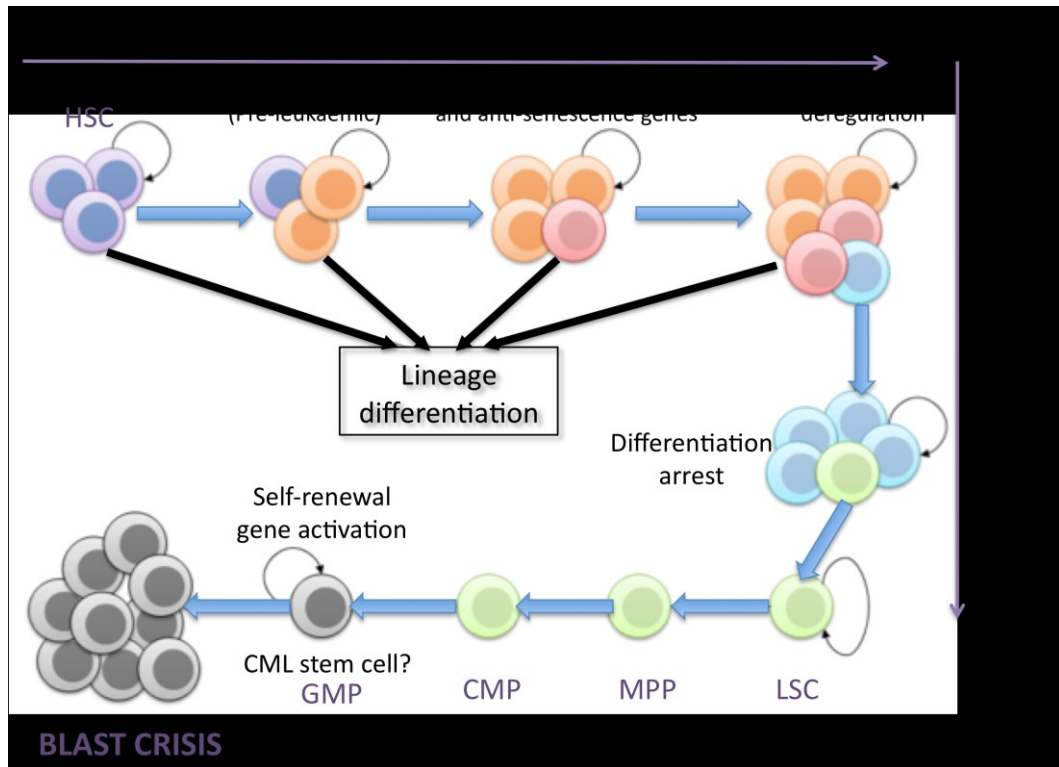


Figure 1.4 Model of disease progression in CML

Leukaemia stem cells (LSC) are likely to originate from self-renewing haematopoietic stem cells (HSC). *BCR-ABL1* expression is probably the initiating pro-leukaemic event in HSCs. Through the progressive activation of anti-apoptotic genes e.g. *BCL2* and proliferative genes e.g. *MYC*, cells gain clonal advantage and fail to undergo programmed cell death. Loss of tumour suppressor genes e.g. *TP53* and *CDKN2A/p16* confers further loss of molecular homeostasis and the acquisition of additional self-renewal capacity giving rise to the LSC. Differentiation arrest occurs as a result of Ikaros family zinc finger 1 (*IKZF1*) or *CEBPA* suppression in LSCs as well as re-activation of developmental self-renewal genes e.g. *CTNNB1* encoding β -catenin protein. The additional accumulation of genetic and/or epigenetic aberrations in multipotent progenitor cells (MPP), followed by common myeloid progenitor cells (CMP) eventually leads to CML-BC at the level of the granulocyte macrophage progenitors (GMP).

(Adapted from Quintas-Cardama et al [103]).

1.7 Epigenetics

Conrad Waddington first used the term *epigenetics* in 1942 for ‘the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being’ [104]. Today, the meaning of the word epigenetics is defined as heritable traits that are not due to changes in the DNA sequence and therefore fundamentally reversible. More broadly epigenetics encompasses the processes that alter differential gene expression [105, 106].

The human genome comprises 3 billion base pairs, two copies of which make up the 2 meters of DNA that is condensed and packaged in to a nucleus just a few microns in diameter. All cells within an organism have identical DNA, epigenetic regulators and transcription factors arranged into accessible and closed regions of the genome to guarantee the precise transcriptional program in any given cell type. This degree of organisation is vital for retaining cell identity and is necessary for crucial processes such as development [107], differentiation [108], rejuvenation [109], proliferation and repair [110].

Epigenetic mechanisms include DNA methylation, nucleosome remodeling, post-translational modifications of histones and (post-) transcriptional regulation by non-coding RNA (ncRNA).

1.7.1 DNA methylation, an historical perspective

Over time, it was clear that in addition to yeasts and fungi, mitotic inheritance in the somatic cells of higher organisms was important. Various aspects of development required further clarification including how terminally differentiated cells could maintain their individual phenotypes during cell division, how an undifferentiated stem cell could divide to produce one differentiated cell, and another undifferentiated stem cell and how one X chromosome is randomly inactivated in the early development of female mammals in each cell, whilst the other remained active.

During the 1960s there became a renewed interest in genetic assimilation, the mechanisms through which environmentally induced phenotypic traits become genetically fixed and thereby removing the need for the original stimulus.

In 1969 Griffith and Mahler hypothesised that DNA methylation could play a key biological role in the molecular heritability of gene activity from a study looking at long-term memory function [111]. Two independent papers followed in 1975 both outlining a model for controlling gene expression through the enzymatic methylation of cytosine in DNA [112, 113]. A third paper published in the same year on DNA methylation postulated that there existed enzymes in eukaryotic organisms to restrict unmodified DNA [114].

Since epigenetic mechanisms are important for normal development, it is now widely apparent that epigenetic dysregulation is a feature of many diseases [115, 116] and abnormal DNA methylation patterning is a recognised hallmark of most human cancers [117-119].

1.7.2 DNA methylation basic function

DNA methylation is one of the most widely studied and well-defined epigenetic modifications. It is a process by which a methyl group ($-\text{CH}_3$) is covalently added to DNA, more specifically the 5-carbon of the cytosine ring resulting in 5-methylcytosine (5-mC). In humans, 5-mC constitutes 1.5% of all genomic DNA and occurs mostly in the context of paired symmetrical methylation of a CpG dinucleotide, where cytosines precede guanine nucleotides [120-122].

More recently, genome-wide assays have identified the presence of non-CpG methylation predominantly in embryonic and pluripotent stem cells [123] but also in certain differentiated cell types including human skeletal muscle [124], brain [125], and T-lymphocyte cells [126]. The purpose of non-CpG methylation is yet to be fully defined and may include transcriptional repression but will not be considered further in this work.

CpG sites are not distributed equally throughout the genome rather there are CpG-dense areas ~1 kb in length called CpG islands (CGI), which span the 5' end of the regulatory/promoter region and are associated with approximately half of all genes in the human genome [127]. CGI tend to remain unmethylated to permit gene expression whether or not the gene is transcriptionally active [120, 121]. In contrast, the vast majority (80%) of the 28 million CpG dinucleotides within the genome, lie outside

promoter regions and are heavily methylated resulting in tight compacted chromatin and transcriptional repression.

Permanent 5mC-mediated gene silencing contributes to genomic stability through imprinting, X-chromosome inactivation and suppression of mobile transposable elements [128, 129].

1.7.2.1 DNA methyltransferases

The existence of proteins possessing methyltransferase activity predicted by Riggs and Holliday and Pugh in 1975 was confirmed by experiments establishing the clonal inheritance of methylation patterns [130, 131] and through the cloning of the first ever mammalian DNA methyltransferase, Dnmt1 [132].

The DNA methyltransferases (DNMTs) catalyse the transfer of a methyl group from S-adenosylmethionine (SAM) to DNA. In mammals three major DNMTs are present; DNMT1, DNMT3a and DNMT3b [133].

DNMT1 is the most abundant DNA methyltransferase and under normal conditions is responsible for the post-replicative maintenance of established patterns of DNA methylation through preferentially binding hemi-methylated DNA at CpG sites [134]. DNMT1 is indispensable for the survival of differentiated cells [135]. This is supported by the finding that homozygous null deletions of mouse DNMT1 are lethal early in life and result in an 80% reduction of global genomic methylation in embryos [136].

The DNMT3 enzymes are a family of DNA methyltransferases consisting of three members (DNMT3a, 3b, and 3L) that can methylate hemimethylated and unmethylated GpCs at the same rate. The architecture of DNMT3 enzymes are closely related to that of DNMT1 with 10 conserved amino acids in the C-terminal catalytic domain [137]. DNMT3a and 3b establish *de novo* DNA methylation patterns and have overlapping functions during early embryogenesis. Both are highly expressed in undifferentiated ES cells but are subsequently downregulated following differentiation [138]. Both are capable of facilitating methylation-independent gene repression and in addition, DNMT3a can interact with the heterochromatin protein 1 (HP1) and Methyl-CpG-binding domain (MBD) protein [139].

Although Dnmt1 is up to 100-fold more active on hemi-methylated substrate *in vitro*, it can be activated to methylate unmodified DNA through a process which appears to require the cooperation of Dnmt3a [140]. Similarly, the *de novo* function of Dnmt3a/b enzymes are not absolute and can adapt the role of maintenance DNMTs. Human cancer cells deficient in DNMT1 only show a 20% reduction in CpG methylation [141].

The remaining DNMT3 group member, DNMT3L is catalytically inactive but essential for directing genomic imprinting and is most active during gametogenesis. Loss of DNMT3L causes bi-allelic gene expression as a result of an anomalously active maternal allele [142]. Furthermore DNMT3L interacts and co-localizes in the nucleus with DNMT3a and 3b to increase their catalytic activity 15-fold [143].

Despite sharing a similar sequence homology with the other DNMTs, much remains to be understood about the biological function of DNMT2 [144]. DNMT2 displays very weak DNA methyltransferase activity but has been shown to possess a novel tRNA methyltransferase action which may be necessary for structure folding, stability [145] and protection against ribonuclease cleavage [146].

1.7.2.2 DNA demethylation

Although global methylation marks are faithfully maintained during somatic cell divisions, great changes to methylation patterning occurs at certain points in mammalian development which are essential for cellular programming and differentiation. Waves of demethylation occur at the time of fertilization followed by *de novo* methylation after implantation of the early embryo [147, 148]. Similarly, modulation of DNA methylation takes place during lineage-specific differentiation of haematopoietic multipotent progenitors [149].

Demethylation of DNA can either be achieved passively simply through the absence of methylation by DNMT1 during replication rounds or actively by way of removal of 5-mC. Three major classes of demethylation enzymes are thought to exist: ten-eleven translocation (TET) protein family members (TET1, TET2 and TET3), activation-induced cytidine deaminase (AID), and thymine DNA glycosylase (TDG).

The TET protein family members TET1, TET2 and TET3 demonstrate the ability to convert 5mC into 5-Hydroxymethylcytosine (5hmC) *in vitro* as well as *in vivo* with

the highest levels in mammalian brain tissue [150]. The wider role of 5-hmC is still uncertain but there is some evidence to show that 5-hmC plays a part in DNA demethylation and is predominately located within promoter regions where it is associated with transcriptionally activated genes [151, 152].

A number of mechanisms have been postulated for the demethylation of 5hmC. The first is that this base undergoes deamination by the AID family of cytidine deaminases to form 5-hydroxymethyluracil (5hmU) followed by base excision repair (BER) [153]. Alternatively, TET family proteins can further oxidise 5hmC to 5-fC (5-formylcytosine), and 5-fC to 5-caC (5-carboxylcytosine) followed by TDG-mediated base excision repair [154]. In support of this, TDG is vital for active DNA demethylation by linked deamination-base excision repair [155]. In addition, reduction of TDG results in the accumulation of 5-caC in mouse ESCs [156].

1.7.2.3 Differentially methylated regions (DMRs)

DNA methylation variation at any CpG site is referred to as a methylation variable position (MVP) and is often considered to be the epigenetic counterpart of a single nucleotide polymorphism (SNP). Where MPVs are detected in 1 kb (up to 1 Mb) clusters at multiple adjacent CpG sites, they are known as differentially methylated regions (DMR) and form the basis for comparative epigenome studies of differential methylation status between organisms or samples. These samples can include different cell or tissue-types from the same individual, the same cell at different time points, or cells from different subjects during health and/or disease. Furthermore, DMRs can be separated by type according to the regions of interest including tissue-specific (tDMR), cancer-specific (cDMR), reprogramming-specific (rDMR), imprinting-specific (iDMR), and aging-specific DMRs (aDMR). DMRs can be identified in the context of genetic imprinting since the presence of methylation is in accordance with either the maternal or the paternal chromosome with the methylated allele usually the silenced one. Epigenetic lesions are present when there are DMRs between the methylation patterns in the parental chromosome and the offspring [157].

1.7.2.4 Chromatin state

Chromatin is the state in which DNA and histone proteins are packaged within the nucleus of mammalian cells, the landscape of which ultimately determines the cell function and fate. Chromatin remodeling is the dynamic modification of chromatin

conformation by different classes of regulators such as histone-modifying enzymes and ATP-dependent remodeling complexes to control gene expression [158, 159].

The nucleosome is the basic unit of chromatin and is composed of an octamer of the four core histones (two copies each of H3, H4, H2A, H2B) around which ~147 base pairs of superhelical DNA are wrapped [160, 161]. These core particles are connected by stretches of 'linker DNA' approximately 80 base pairs (bp) long. The linker histone H1 (and its isoform H5) is involved in nucleosome stabilisation and chromatin compaction through the binding of histone H1 to the nucleosome and the linker DNA entering and exiting the core particle [162, 163].

Nucleosomes are folded through a series of successively higher order structures to eventually form the chromosome which not only physically compacts DNA but also creates an additional layer of regulatory control to maintain correct gene expression. The degree of chromatin compaction and precise nucleosome positioning can be influenced by a wide-array of cellular processes [164]. Euchromatin is associated with an open chromatin conformation and a state permissive of transcription while heterochromatin is tightly packed, inactive and hostile to protein binding [165].

1.7.2.5 Histone modifications

A histone modification is a covalent post-translational modification (PTM) to the amino acid residues on the unstructured 'tail' of histone proteins. A wide variety of biochemical modifications exist including acetylation, methylation, ubiquitination, phosphorylation, citrullination and ribosylation [166]. These modifications orchestrate various biological outcomes by either altering nucleosome-DNA and nucleosome-nucleosome interactions or by providing binding surfaces for downstream epigenetic 'readers' [167-169]. Modifications such as acetylation and phosphorylation lower the electrostatic interaction between histone tails and are predicted to relax the core-DNA binding depending on the location of the modification within the core [170]. Some modifications have been shown to correlate with gene silencing while others correlate with transcriptional activation. Combinations of these modifications are thought to constitute a phenomenon known as the 'histone code' which serves to recruit other proteins by specific recognition of the modified histone [167, 171].

For example, two histone modifications; trimethylation of H3 lysine 4 (H3K4me3) at the 5' proximal promoter region [172, 173] and trimethylation of H3 lysine 36

(H3K36me3) throughout the gene body [174, 175] are general characteristic of euchromatin and associated with active transcription. Conversely, the silent heterochromatic state is associated with low levels of acetylation and high levels of three histone modifications; trimethylation of H3 lysine 27 (H3K27me3) deposited by the polycomb complex PRC2 [176], Di- and tri-methylation of H3 lysine 9 (H3K9me2/3) which serves as a binding site for heterochromatin protein 1 (HP1), which recruits further repressive marks [177] and trimethylation of H4 lysine 20 (H4K20me3) [178].

1.7.3 DNA methylation and cancer

Epigenetic alterations play an important role in many cancers, [116]. Similar to genetic mutations, ‘epimutations’ can activate oncoproteins, inhibit tumour suppressor genes and propagate genomic instability. These alterations are associated with aberrant patterns of gene expression that are heritable and passed to daughter cells during mitosis. In contrast to genetic mutations, the DNA sequence remains entirely unaltered (providing an epimutation has not emerged secondary to a genetic one) [179]. Most importantly, since they are potentially reversible, epimutations are attractive drugable targets in cancer treatment.

Abnormal DNA methylation patterns are evident in almost all human cancers [180, 181] with virtually every type showing loss of methylation within CpG-depleted regions and gains in CGIs in or around 5’ gene promoter regions [182, 183] (Figure 1.5).

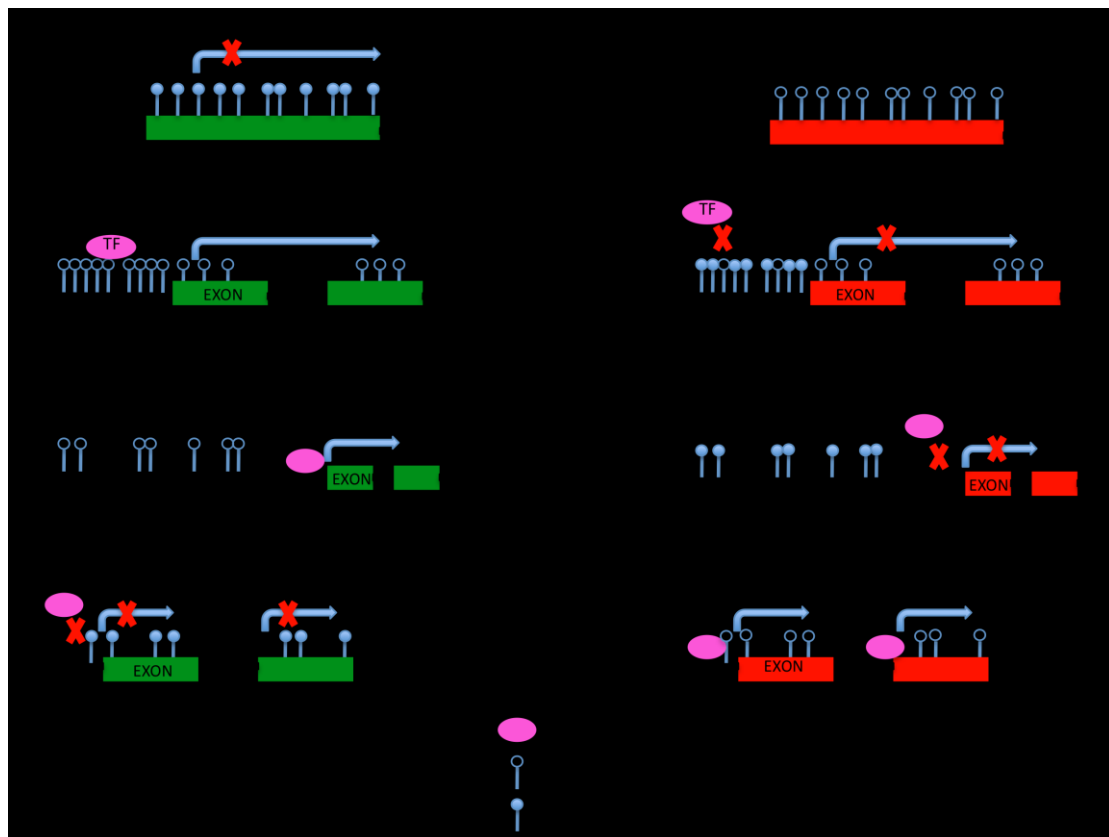


Figure 1.5 DNA methylation patterns in healthy and cancer cells

Repetitive sequences (e.g. LINE family member, L1) are usually methylated at CpG dinucleotides in normal cells with global loss in cancer cells. This leads to chromosomal instability and activation of transposable elements (**top**). CGI-associated promoters remain unmethylated in normal cells to permit expression of key regulatory genes including tumour suppressors (e.g. *p16INK4a*, *p15INK4b*). In cancer cells CGI-associated promoters can become hypermethylated leading to transcriptional repression (**second**). Similar patterns are seen in CpG island shores, located upstream of promoter regions affecting genes including *HOX2A* and *GATA2* (**third**). CpGs located in gene bodies are methylated in normal cells and reversed in cancer cells, giving rise dysregulated transcription at several incorrect sites (**bottom**).

1.7.3.1 DNA hypomethylation in cancer

Global genomic hypomethylation is evident in regions of repetitive DNA sequences [184] as well as inter- and intragenic regions [185], the consequences of which include loss of oncogene and/or imprinted gene repression [186, 187], genomic instability giving rise to mitotic recombination or chromosomal translocations [129, 188] and retrotransposon activation [128].

Coupled with this, the overall degree of hypomethylation in many cases is related to the disease aggressiveness. A number of studies have shown that as cancer progresses, so too does the global loss of methylation. Whether this is the cause of increasing genomic instability or a consequence of deteriorating cellular regulation is difficult to prove [189].

To add to the complexity, studies comparing genome-wide methylation in normal and transformed cells have found that site-specific hypomethylation occurs preferentially in gene-poor regions of the cancer methylome [182]. LINE1 is a highly repetitive retrotransposon that comprises up to 15% of the human genome with a number of regions capable of transposition or transcriptional interference of nearby genes [190, 191]. LINE1 hypomethylation was detected in a number of human cancers including chronic lymphocytic leukemia [192], bladder [193] and prostate carcinomas [194]. Similarly, the HERV-K family endogenous retroviral class of human retrotransposons may become hypomethylated and some of the long terminal repeats may initiate reporter gene expression [195]. Cancer-associated hypomethylation of HERV-K *gag* sequences was found in one small testicular carcinoma study and correlated with the expression of Gag at the protein level [196]. In addition, in one study looking at breast cancer, high transcript numbers of HERV-K *env* transcripts were detected in the majority of cancer samples but not in comparative healthy breast tissue [197].

It is well established that aberrant hypomethylation can disrupt genomic imprinting leading to biallelic expression, which is thought to be an early step in neoplastic transformation [198]. Hereditary Beckwith-Wiedemann syndrome is associated with loss of imprinting (LOI) of the insulin-like growth factor gene (*IGF2*) and is associated with a high risk of cancer [199]. Indeed, Loss of imprinting of the *IGF2* gene has also been found in many other types of malignancies, including Wilms tumour [200] breast [201] and ovarian [202] cancer and strikingly, can be detected

in 10% of outwardly healthy individuals at risk of bowel cancer [203, 204]. Although more recently the latter has been refuted [205].

1.7.3.2 DNA hypermethylation in cancer

The overall reduction of methylation levels in cancer cells is accompanied by hypermethylation at specific sites in the genome namely CpG islands within gene promoter regions. While the initiating events that drive the promoter hypermethylation phenotype in transformed cells are not entirely clear, the results can be catastrophic and negatively impact on genes involved in a number of cellular networks including the cell cycle [206], apoptosis [207], repair [208], angiogenesis [209] and adhesion [210].

It is still unclear what prompts methylation to take place at previously unmethylated sites in cancer cells and why this varies both inter- and intratumorally. It seems that the promoter CGIs of some genes such as *P16^{INK4a}*, *P15^{INK4a}* and *TP53* are susceptible to methylation across many cancer types [211], while others display promoter hypermethylation with high frequency in some cancers only e.g. *GSTP1* in 90% of prostate cancers but essentially remains unaltered in other malignancies [212, 213]. This may reflect variable dependency certain tumour types have on particular genes for growth and survival or could be related to the chromosomal location or position of CGIs within a particular nucleotide sequence that allows them to undergo large-scale epigenetic dysregulation [181].

At this time it is still impossible to distinguish between primary ‘driver’ and secondary ‘passenger’ methylation events. Whether aberrant methylation originates from the over activity of a methylating factor or failure of a demethylating process remains to be proven.

1.7.3.3 The role of DNA methyltransferases in cancer

It has been known for some time that an overall increase in DNMT activity occurs during tumorigenesis, which is surprising given that the cancer methylome undergoes an overall reduction of its 5-mC content in comparison to the normal cell from which it originated.

One possibility for this is that an alteration in global chromatin structure causes failure of DNMTs to recognise or access the ‘correct’ sequences for methylation,

whilst CGIs become methylation permissive and in turn attracts further DNMTs to the area. In an attempt to restore normal methylation patterns, DNMT gene transcription increases but fails to reverse the disordered process.

On the one hand it has been clearly demonstrated in *Dnmt1*-hypomorphic mice that sufficient *DNMT1* expression is a prerequisite for intestinal polyp formation and disruption of the DNMT1 gene significantly reduces the incidence of tumours [214]. Conversely, human colorectal cancer cells (HCT116) lacking DNMT1 show a marked reduction in cellular DNMT activity and growth, but only a modest decline in genome-wide DNA methylation levels. In particular, aberrant CGI methylation of tumour suppressor gene promoters such as *p16INK4a* remained, suggesting cooperation from the other methyltransferases DNMT3A and DNMT3B [134, 141, 215].

In addition to possessing methyltransferase activity, the DNMTs likely advocate transcriptionally repressive chromatin by means of their regulatory, non-catalytic N-terminal domain, interacting with histone deacetylases (HDACs) [216-219].

1.7.3.4 The role of histone deacetylases (HDACs) in cancer

Histone deacetylases (HDAC) catalyse the hydrolytic removal of acetyl groups from amino acid residues on histone tails. In the broadest terms (dependent somewhat on the amino acid residue and its placement along the histone tail), recruitment of these enzymes decreases levels of histone acetylation leading to compact chromatin and transcriptional repression [220]. Perturbed activity has been associated with tumorigenesis and even cancer metastases [221, 222].

Recently, certain ‘histone code’ patterns of acetylation and deacetylation together with other post-translational modifications have been used to distinguish malignant cells from their unaffected counterparts, adding a further layer of complexity to the epigenetic changes that occur in cancer [223]. For example, epigenetic inactivation of tumor-suppressor genes by promoter CGI hypermethylation is associated with a general combination of histone marks including deacetylation of histones H3 and H4, loss of H3K4me3 and gains of H3K9 and H3K27me3 (by the histone methylating, polycomb group repression complex, PRC2) [116, 224].

Furthermore, certain histone marks have been used to stratify the risk of disease recurrence in certain cancer types e.g. acetylation of H3K18 coupled with histone H3K4me2 reduces the risk for prostate cancer relapse [225].

Even in the absence of GCI hypermethylation, the cyclin-dependent kinase p21^{WAF1}, a cell-cycle checkpoint inhibitor with tumour suppressor activity is frequently silenced. In many cases and tumour types this is through hypoacetylation of histones H3 and H4 [226]. Furthermore, treatment with HDAC inhibitors results in a reduction of tumor-cell growth, an increase in the acetylation and subsequent upregulation of gene expression [227].

In addition, other studies demonstrate that HDACs not only regulate the expression of numerous other genes through interaction with various transcription factors including E2f/retinoblastoma protein (pRb) complex [228], Stat3, p53 [229, 230] and NF-κB [231] but also deacetylate non-histone proteins involved in cancer development and progression [232, 233].

1.7.4 Epigenetic aberrations in CML

The advent of high-throughput and more affordable next-generation technologies have led to a rise in the number of DNA methylation studies in both acute and chronic, lymphoid and myeloid malignancies. Underlying most clinical investigations is the aim of better understanding the mechanisms underlying disease pathogenesis as well as identifying biomarkers of disease aggressiveness or response to therapy.

Compared to other leukaemias, there are relatively fewer comprehensive reports of DNA methylation alterations in CML and the majority of these have adopted a targeted approach to address questions relating to small numbers of preselected genes. Moreover, most of these have addressed possible differences between disease phases.

Ben-Neriah and colleagues provided the first description of *ABL1* hypermethylation in CML and showed that promoter CGI methylation was indicative of clinical progression [234]. Using a methylation-sensitive restriction endonuclease approach on peripheral blood and bone marrow material it was possible to demonstrate that patients with early CML-CP lacked promoter methylation at the *ABL1* locus, while the majority of cell lines and primary CML-BC samples underwent marked hypermethylation, with CML-AP samples showing patterns intermediate to these

extremes. The same group later validated these data in a follow-up study with larger patient sample numbers [235, 236]. One further observation that emerged from this was that patients with longstanding CML-CP disease displayed partial allele-specific *de novo* methylation of the *ABL1* promoter nested within the chimeric oncogene similar to those patterns seen in accelerated phase patients. Furthermore, exposure to IFN- α but not hydroxycarbamide therapy resulted in a degree of reversal, regardless of cytogenetic response status. This indicates that perhaps *ABL1* promoter hypermethylation is not informative as a predictive response biomarker for this therapy, a conclusion made by other groups [237].

Strathdee and others analysed 378 primary samples from patients with different lymphoid and myeloid malignancies including CML for evidence of HOX family gene silencing by DNA methylation [238]. The HOX proteins are homeodomain-containing transcription factors and key regulators of embryonic development and cellular differentiation, capable of binding to enhancers where they either activate or repress genes accordingly [239]. Using combined bisulfite restriction enzyme analysis (COBRA) with pyrosequencing, a total of eight *HOXA* and B cluster genes were quantitatively assessed for methylation in 44 CML-CP and 23 CML-BC peripheral blood samples. *HOXA5* and *HOXA4* but no *HOXB* family genes were found to be hypermethylated in a significant proportion of the chronic phase samples. Moreover, the frequency and degree of *HOXA5* and *HOXA4* hypermethylation was markedly elevated at the time of blast crisis with 100% of myeloid blast crisis samples exhibiting *HOXA5* hypermethylation. Furthermore, re-expression of *HOXA5* in CML-BC cells caused induction of granulocytic differentiation markers leading to the hypothesis that loss of expression through hypermethylation may play an important role in disease progression in CML [238].

Promoter hypermethylation associated with a negative outcome on prognosis in CML has also been demonstrated with calcitonin 1 (*CALC1*) gene [240], *HIC1* [241], the oestrogen receptor gene [242], the preferentially expressed antigen in melanoma (*PRAME*) [243] and the cyclin-dependent kinase inhibitors *p16INK4a*, *p15INK4b* and *p14ARF* [244, 245].

Using a methylated CpG island recovery assay (MIRA) in combination with low-density genome-wide CpG island arrays, Dunwell *et al.* identified two genes;

TFAP2A and *EBF2* encoding sequence specific DNA binding transcription factors, that both showed a statistically significant increase in promoter methylation in CML-BC samples (n= 8) compared to CML-CP samples (n=55) [246].

Jelinek *et al.* studied the methylation status of 10 genes from the peripheral blood mononuclear cells (PMBCs) of 120 patients with CML at different time points and in various phases using bisulfite pyrosequencing and the LINE-1 repetitive element, as a surrogate marker of global genomic methylation [247]. The average proportion of methylated genes was higher in CML-AP than CML-CP, with significant methylation increases seen in 5 of the genes studied including *CDKN2B*, *OSCP1*, *PGRA*, *PGRB* and *TFAP2E*. Abnormal methylation of the pro-apoptotic, SRC suppressor gene, *PDLIM4*, was associated with shortened survival independent of CML phase and imatinib responsiveness.

The gene encoding Bcl-2-like protein 11 (*BCL2L11*, or '*BIM*') an apoptotic activator which interacts with BCL2, BCL2L1/BCL-XL, and MCL1, crucial for the regulation of HSC survival has been shown to be the subject of post-transcriptional, proteasome-dependent degradation by ERK1/2 phosphorylation [248, 249]. Furthermore, *BIM* up-regulation in response to IM accounts for drug selective cytotoxicity, while epigenetic down-regulation of *BIM* expression by promoter hypermethylation is associated with reduced optimal IM responses in CML and unfavourable outcomes [250-252]. More recent studies have identified two possible epigenetic mechanisms underlying *BIM* promoter hypermethylation namely the loss-of-function of Tet methylcytosine dioxygenase 2 (TET2) (catalyses the conversion of 5mC to 5hmC) [253] and abnormal DNMT1 recruitment [254]. Re-expression of *BIM* by decitabine (5-aza-2'deoxyctidine), potentially supports the use of epigenetic modulating agents in CML. Decitabine is a cytosine nucleoside analogue which binds DNA during cell replication and in so doing covalently 'traps' DNMTs leading to DNA-protein adducts. This trapping results in both the degradation of available cellular DNMT as well as activation of DNA damage response (DDR) mechanisms causing global hypomethylation and ultimately apoptosis [255].

To date, two phase II studies have been conducted to determine the safety and efficacy of decitabine in CML. Issa *et al* demonstrated clinical activity in a subset of IM-refractory or -intolerant patients while Oki and colleagues combined decitabine

with IM in CML-AP and myeloid -BC [254, 256, 257]. While decitabine was well tolerated, patient numbers were small in both trials and data were mixed in these ultra-high risk groups.

The Phosphatase and Tensin homolog (PTEN) is a tumor-suppressor gene that negatively regulates the PI3K pathway including the serine/threonine kinases Akt/protein kinase B (PKB) to limit pro-survival and -proliferation signaling [258]. Loss of PTEN protein is associated with poor prognosis in a variety of cancer types [259-261]. During CML-CP pathogenesis, PTEN has been shown to undergo non-genomic loss of function by a number of mechanisms including promoter hypermethylation through increased expression and recruitment of the DNMT3A and EZH2 complex formation [262]. Nishioka *et al* showed that IM resistance in CML can in part be due to epigenetic silencing of PTEN by abnormal promoter hypermethylation. They were able to demonstrate the potential role of hypomethylating agents in the treatment of patients with chemoresistance [263].

Although the potential role of HDACs has been studied extensively in acute leukemia, there are few reports addressing their aberrant activity in CML. Similarly, microRNA (miR) expression has not been widely studied in CML though more recently data is accumulating. miR are small non-coding RNA molecules that bind the 3'UTR of mRNAs in a sequence-specific manner to induce mRNA cleavage, increase decay, or repress translation [264]. miRs are emerging as major contributors in complex molecular signaling pathways [265, 266] and have been shown to be important for both normal haematopoietic cell development as well as leukaemogenesis [267-269] where they behave as tumor suppressors or oncogenes [270].

The expression patterns of miR-96, miR-150 and miR-151 are different in CD34⁺ CML bone marrow cells compared with their normal counterparts [271] and BCR-ABL1 kinase activity, which increases in CML-BC, appears to be responsible for miR deregulation in CML progenitors. For example the MYC- and BCR-ABL1-regulated polycistronic miR-17-92 clusters are upregulated early in CML, but not during blast crisis [272].

The upstream region of pri-mir-203, the gene encoding miR-203, is heavily methylated in Ph⁺ cells, resulting in its reduced expression [266]. Correspondingly, induction of miR-203 by DNMT inhibitors directly down regulates BCR-ABL1 upon

interaction with its 3'UTR and increases apoptosis. Similarly, miR-328 levels are low in CML-BC and increase upon imatinib treatment [269].

CML represents a paradigm in the therapeutic targeting of human cancers and is arguably one of the most widely studied malignancies in terms of its underlying pathobiology. It is clear that despite the consistent finding of BCR-ABL1, individual patients differ substantially in their disease course and response to treatment. These differences may relate to variations intrinsic in the leukemia cell, the host genome or epigenome, but likely a combination of these. Elucidating the mechanisms that underlie this heterogeneity is of fundamental importance, because it may facilitate selection of the best treatment for a given patient and even provide insights into the process of malignant transformation and disease progression in other forms of leukaemia or less well-characterized cancers.

1.8 Hypothesis and aims of this thesis

The introduction has speculated that aberrant epigenetic reprogramming and in particular abnormal DNA methylation might contribute to CML leukaemogenesis or underlie the variations in outcome and responses to imatinib.

1.8.1 Hypotheses

1. Global DNA methylation is altered in CML and associated with changes in gene expression
2. The degree of aberrant DNA methylation correlates with CML disease activity
3. Distinct DNA methylation profiles underlie CML clinical responsiveness to imatinib therapy.

1.8.2 Aims

The aims of the work presented in this thesis were to perform an integrated epigenomic study on CD34+ cells from CML patients to see whether the DNA methylation and transcription profiles would reflect disease-specific differences in Ph+ CML and to investigate the extent to which epigenetic profiling could be associated with response to TKI therapy.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

DNase I (Calbiochem)

Trypan blue solution 0.4% (Sigma-Aldrich)

Hank's balanced salt solution (HBSS) (Gibco)

RPMI 1640 medium (Gibco)

Lymphoprep (Axis-Shield)

3% acetic acid (Sigma-Aldrich)

Molecular grade ethanol (Fisher Scientific)

Phosphate buffered saline (PBS) (Gibco)

β -mercaptoethanol (Gibco)

RNaseZap Decontamination Solution (Life Technologies)

Doxycycline 1 μ g/mL (Sigma-Aldrich)

Imatinib mesylate (Novartis)

Marvel milk powder (Nestle)

10X Tris Buffered Saline with Tween 20 (TBS-T) (Cell Signaling Technology)

Agarose (Invitrogen)

Ethidium Bromide solution (10mg/mL) (Promega)

Water (Molecular Biology/PCR Grade) (Fisher Scientific)

2.1.2 Composition of buffers, solutions and media

Sterile, high-performance liquid chromatography (HPLC)-grade water was used when required to prepare all solutions

Cell freezing medium

Filter sterilised using vacuum bottle top filters:

- 50% fetal bovine serum (FBS) (Biosera)
- 40% RPMI 1640 medium (Gibco)
- 10% dimethyl sulfoxide (DMSO) (Sigma-Aldrich)

Cell thawing solution

Filter sterilised using vacuum bottle top filters:

- 50% FBS (Biosera)
- 50% RPMI 1640 medium (Gibco)
- Preservative-free heparin sodium 10 IU/mL (CP Pharmaceuticals)
- DNase I 1000 IU/mL (Calbiochem)

MACS buffer

Filter sterilized:

- Phosphate buffered saline (PBS) (Gibco)
- 0.5% bovine serum albumin (BSA) (PAA laboratories)
- 2 mM ethylenediaminetetraacetic acid (EDTA) (Promega)

Cell culture medium

Filter sterilized:

- Hygromycin B (447 U/mL) (Invitrogen)

- Geneticin (0.5 mg/mL) (Invitrogen)
- RPMI 1640 (Gibco)
- 1% L-glutamine (Invitrogen)
- 1% penicillin (Invitrogen)
- 1% streptomycin (Invitrogen)
- 10% Tet System Approved FCS (Clontech Laboratories)

Cell lysis buffer

- 50 mM Tris-HCl (pH 7.5) (Promega)
- 1 mM EDTA (Promega)
- 1 mM ethylene glycol tetraacetic acid (EGTA) (Merck Millipore)
- 1% (v/v) Triton X-100 (Promega)
- 10 mM b-glycerophosphate (Merck Millipore)
- 50 mM NaF (Sigma-Aldrich)
- 1 mM sodium orthovanadate (Sigma-Aldrich)
- 5 mM sodium pyrophosphate (Sigma-Aldrich)
- 100 nM okadaic acid (Abcam Biochemicals)
- 0.27 M sucrose (Merck Millipore)
- 1 mM dithiothreitol (DTT, Clelands reagent) (Thermo Scientific)
- 1X protease inhibitor cocktail (PIC) (Roche)

Loading buffer

- 2.3% sodium dodecyl sulphate (SDS) (Sigma-Aldrich)
- 10 mM DTT

- 0.01% bromophenol blue (Tocris)
- 10% glycerol (Sigma-Aldrich)
- 62.5 mM Tris (pH 6.8) (Sigma-Aldrich)

10X Running buffer

- 0.25 M Tris
- 1.92 M glycine
- 1% SDS

1X-running buffer was prepared by mixing 100 mL 10X buffer with 900 mL HPLC water and the pH adjusted to 8.5.

10X Tris-borate-EDTA (TBE) buffer

The following salts were dissolved in 1 L of HPLC water:

- 108 g Tris base
- 55 g boric acid (VWR International)
- 9.3 g EDTA

1X TBE buffer was prepared by mixing 100 mL 10X preparation with 900 mL HPLC water.

2.1.3 Enzymes, antibodies and kits

CliniMacs CD34 antibody (Miltenyi Biotec)
Dead Cell Removal Kit (Miltenyi Biotec)
MACS Cell Separation MS and LS Columns (Miltenyi Biotec)
Pre-Separation Filters (30 µm) (Miltenyi Biotec)
Antibody-capture CompBead particles (BD Biosciences)
Fluorescent-labeled antibodies for flow cytometry (see Table 2.1)
Anti-c-Abl antibody (Ab-3) (1:1000) (Merck Millipore)
Anti-Phosphotyrosine antibody (clone 4G10) (1:500) (Merck Millipore)
Anti-Phospho-CrkL antibody (Tyr207) (1:1000) (Cell Signaling Technology)
Anti-Lamin B1 antibody (Merck Millipore)
Horseradish peroxidase (HRP)-conjugated secondary antibodies (1:2000) (Dako)
Enhanced chemiluminescence (ECL) detection kit (Amersham Pharmacia Biotech)
QIAamp DNA Blood Mini Kit (Qiagen)
RNase A (Qiagen)
RNeasy Plus Micro Kit (Qiagen)
RNeasy Plus Mini Kit (Qiagen)
QIAshredder Homogenizer (Qiagen)
RNA Analysis ScreenTape for use with the TapeStation system (Agilent Technologies)
EZ DNA Methylation Kit (Zymo Research)
EZ-96 DNA Methylation Kit (Deep-Well format) (Zymo Research)
Illumina TotalPrep RNA Amplification Kit (Ambion)
Infinium HumanMethylation450 BeadChip Kit (Illumina)
HumanHT-12 v4 Expression BeadChip Kit (Illumina)
Ion AmpliSeq Custom Panel designed online using Ion AmpliSeq Designer (Ion Torrent, Life Technologies)
Ion AmpliSeq Library Kit 2.0 (Ion Torrent, Life Technologies)
KAPA Library Quantification Kit (Kapa Biosystems)
Ion Xpress Barcode Adapters 1–96 Kits (Ion Torrent, Life Technologies)
Agencourt AMPure XP Kit (Beckman Coulter)
Ion PGM Template OT2 200 Kit (Ion Torrent, Life Technologies)

Ion 318 Chip Kit v2 (Ion Torrent, Life Technologies)
Ion PGM Sequencing 200 Kit v2 (Ion Torrent, Life Technologies)
Qubit dsDNA BR and HS Assay Kits (Life Technologies)

2.1.4 General laboratory equipment

Centrifuge (5427 Refrigerated) (Eppendorf)
Mini-centrifuge (Benchmark Scientific)
Large bench-top centrifuge (Hettich Zentrifugen)
NanoDrop 2000 spectrophotometer (Thermo Scientific)
Qubit 2.0 Fluorometer (Life Technologies)
2200 TapeStation Instrument (Agilent Technologies)
Mini- and MidiMACS Separator (Miltenyi Biotec)
MACS MultiStand (Miltenyi Biotec)
Cryo-freezing containers (Nalgene, Thermo Scientific)
DynaMag-96 Side Magnet (Life Technologies)
pH meter (Jenway)
Kodak Biomax MR film (VWR)
Transilluminator (Uvitec Cambridge)
Light microscope (Nikon)
Haemocytometer and cover slips (Hawksley)
Platform rocker (Grant Bio)
Heating block (Grant Instruments)
FACSCanto II analyser (BD Biosciences)
G-Storm (GS4) thermal cycler (Thermal Cycler Systems)
7500 Real-Time PCR System (Applied Biosystems)

2.1.5 Genomic equipment

iScan (Illumina)
Ion OneTouch 2 System [including the Ion OneTouch 2 Instrument and OneTouch ES (enrichment system)] (Ion Torrent, Life Technologies)
Ion Personal Genome Machine (PGM) System (Ion Torrent, Life Technologies)

2.1.6 Bioinformatic software, scripts and statistical packages

Excel 2010 for Mac (Microsoft)

GraphPad Prism (GraphPad Software)

BD FACSDiva software v6.1.3 (BD Biosciences)

FlowJo v7.5 software

UCSC genome browser

Ensembl genome browser

GenomeStudio (Illumina)

R Statistical package

Bioconductor packages minfi and lumi

Marmal-Aid (Dr Rob Lowe)

Torrent Suite, Ion Reporter Software (Life Technologies)

The Cancer Genome Atlas (TCGA) (Broad Institute)

Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.7

[National Institutes of Health (NIH)]

2.2 Methods

2.2.1 Collection of human primary cell samples

This study obtained full approval from the Local Research Ethics Committee and written informed consent was obtained from all subjects. The majority of samples used were stored, frozen material collected and saved in the CML Biobank at the John Goldman Centre for Cellular Therapy, Hammersmith Hospital (Human Tissue Authority (HTA) licensing number 1118). However, patients newly diagnosed with CML and healthy donors were also included. Given that preliminary studies using colony-replating assays showed that the viability of stored frozen material was comparable with fresh cells, all cells were frozen prior to CD34⁺ enrichment to remove the possibility of processing bias. Whether the practice of freezing affects epigenetic processes such as DNA methylation and gene expression, is not entirely clear, but unlikely [273, 274].

Bone marrow aspiration and stem cell collection procedures were performed using standard operating procedures (SOPs) approved by the Human Tissue Authority (HTA) and Joint Accreditation Committee-ISCT & EBMT (JACIE), accreditation process. SOPs are available but not included.

2.2.1.1 Cytogenetic analysis and molecular monitoring

Sequential conventional cytogenetic analysis by metaphase karyotyping was performed on at least 20 metaphases in all patients with CML at the time of diagnosis and at 6 monthly intervals until the attainment of CCyR. This work was carried out in Dr Alistair Reid's laboratory, in the Cytogenetic unit of Imperial Molecular Pathology Laboratory, Hammersmith Hospital, London UK.

BCR-ABL1 transcripts were measured in the blood at 6- to 12-week intervals from diagnosis using real-time quantitative polymerase chain reaction (RTq-PCR) in Prof Letizia Foroni's laboratory, Molecular Unit of Imperial Molecular Pathology Laboratory, Hammersmith Hospital, London UK. Results were expressed as log₁₀ reductions from a standardized baseline according to the international scale (IS).

2.2.1.2 Peripheral blood mononuclear/progenitor cell collection by leukapheresis

Leukapheresis was performed on the Spectra Optia Apheresis System by specialist nurses in the Department of Clinical Haematology at Imperial College Healthcare NHS Trust, according to the SOP.

The Spectra Optia Apheresis System spins and separates blood by centrifugal force and gravity. Dual needle venous access was required for this procedure. Pumps are activated to harvest the desired component while the remaining products are combined and returned to the patient/donor circulation via the second venous line.

The target cell collection was 10×10^8 nucleated cells/kg in newly diagnosed patients with CML, 2×10^6 CD34⁺/kg in patients with CML at the time of CCyR and $5-12 \times 10^6$ CD34⁺/kg for healthy controls.

2.2.1.3 Granulocyte-colony stimulating factor (G-CSF) mobilisation of peripheral blood stem/progenitor cells from CML patients and healthy donors

This procedure was performed according to Imperial College Healthcare NHS Trust Clinical & Investigative Sciences Group (Clinical and Laboratory Haematology) Quality Management Policy and Procedures: Mobilisation of PBPC: G-CSF Alone.

Stem/progenitor cells were mobilised into the circulating peripheral blood by Filgrastim (G-CSF). Filgrastim was given at a dose of approximately 10µg/kg/day subcutaneously for at least 4 days (while continuing imatinib therapy in the patient group) with apheresis scheduled for the 5th day of G-CSF treatment where the estimated CD34⁺ cell count was achieved. In those patients with a suboptimal CD34⁺ cell estimation, harvest was delayed until day 6 while G-CSF therapy continued. If the target cell dose was not obtained in a single procedure, up to two additional harvests were attempted (usually on day 6 and 7). In some patients, stem cell harvest was repeated approximately 6 months later and only after imatinib therapy had been discontinued for 2-3 weeks prior to mobilisation, using higher doses of G-CSF.

2.2.1.4 Bone marrow aspirate

This procedure was performed by appropriately trained medical staff in the Department of Clinical Haematology at Imperial College Healthcare NHS Trust, according to the SOP; Imperial College Healthcare NHS Trust Clinical & Investigative Sciences Group (Clinical and Laboratory Haematology) Quality

Management Policy and Procedures: Performing and Reporting Bone Marrow Biopsy (Aspirates and Trepines)

Briefly after explaining the procedure, the benefits and potential risks, written informed consent was obtained. With the patient placed in the left lateral position, the posterior iliac crest was located and sufficient subcutaneous local anaesthetic (2% Lidocaine) infiltrated. Further Lidocaine was administered to the periosteum covering an area of 2-3 cm diameter and sufficient time given to ensure good anaesthesia. The skin was punctured with the aspirate needle and advanced until the posterior iliac crest was reached. The needle was carefully passed through the cortical bone until reaching the marrow cavity at which point the stylet was removed and replaced with a pre-heparinised 20 mL syringe. 20 mL of marrow were aspirated and immediately placed into standard ethylenediaminetetraacetic acid (EDTA) phlebotomy tubes. The aspiration needle was removed and pressure applied to the wound until haemostasis achieved.

2.2.2 Clinical sample processing

2.2.2.1 Cryopreservation of primary cell samples

The cell suspension was centrifuged at approximately 1,800 rpm for 5 minutes in 50 mL falcon tube and the supernatant removed without disturbing the loose cell pellet using an Eppendorf pipette controller. Up to 2×10^7 CD34⁺ selected cells or 10^9 mononuclear cells (MNC) were resuspended in cell freezing medium (stored at 4°C), gently mixed to ensure a homogeneous cell suspension and aliquoted into 1.8 ml sterile cryogenic storage vials before placing on ice for 5 minutes. The cell-containing cryovials were placed inside cryo-freezing containers (Mr. Frosty) and first cooled at -80°C overnight to provide a controlled temperature reduction, then transferred to liquid nitrogen for long-term cryogenic storage.

2.2.2.2 Recovering frozen primary cells

Cell thawing solution (stored at 4°C) was pre-warmed to 37°C. Using an aseptic technique in a laminar flow hood, cryovials containing MNCs/CD34⁺ selected cells were removed from liquid nitrogen storage, quick-thawed by placing/swirling in a 37°C water bath and rapidly transferred to 20 mL universal tube. 15 mL of thawing solution were added; initially drop wise, whilst gently agitating the tube to prevent cell clumping. Cells were then centrifuged for 5 min at 1,500 rpm, the supernatant discarded carefully from the loose pellet using Pasteur pipettes followed by the addition of 200 µL DNase I for 5 minutes. This step was performed at room temperature (RT) with constant flicking of the tube to ensure even dispersal and enhance activity while preventing clumping of the cells. The DNase I-soaked pellet was then washed twice in pre-warmed HBSS and resuspended in either 7 mL or 30 mL of HBSS depending on cell numbers. If cell clumps were still visible a second round of 200 µL DNase I were added and the process repeated. Resistant clumps were removed after resuspension by passing through a cell strainer as excessive DNase I results in poor cell viability.

2.2.2.3 Purification of the MNC fraction

Fresh blood samples were diluted 1:4 in HBSS. After gentle mixing, 35 mL were carefully layered drop-wise on to the surface of 15 mL Lymphoprep (stored at 4°C and brought to RT) in a 50 mL falcon tube for separation by centrifugation over a

density gradient for 30 min at 1,800 rounds per minute (rpm) (centrifuge with brake off). Following centrifugation, the opaque interface containing the MNCs was carefully harvested with a Pasteur pipette, transferred into a clean falcon tube and washed twice in sterile HBSS by further centrifugation for 5 mins at 1,800 rpm (centrifuge with brake on). Cells were then counted using a haemocytometer after 1:20 dilution in 3% acetic acid to lyse red cells. The MNCs were cryopreserved according to Section 2.2.2.1 until required.

Following thawing of MNCs, clumps of dead cells (an inevitable consequence of the freeze-thaw process) were removed by Lymphoprep density gradient centrifugation. Depending on cell numbers at the time of freezing, thawed MNCs were suspended in either 7 mL or 30 mL of HBSS at RT and layered over 6 mL of Lymphoprep solution (contained in a 15 mL Falcon tube) and 15 mL (in a 50 mL Falcon tube), respectively, followed by centrifugation as described above. Lower volume samples in 15 mL Falcon tubes were centrifuged at a reduced speed of 1,700 rpm for 20 mins.

2.2.2.4 Trypan Blue exclusion assay and cell counting

Cell viability was performed on all clinical samples including fresh material and cryopreserved cells. Trypan blue is a stain that enters dead cells through damaged membranes. Using bright field microscopy, these cells appear large, blue and flat. In contrast, viable cells which exclude the stain, are seen as bright and refractile. Cells were washed and suspended in PBS according to the size of the pellet. To 20 μ L of cells an equal volume of 4% trypan blue was added and mixed well. 10 μ L were then loaded between a coverslip and standard haemocytometer Neubauer chamber and viewed under the 10x objective (low magnification) of a light microscope: total cells numbers were calculated using the formula: Total number of viable cells = (Number of viable cells in four corner squares /4) $\times 10^4 \times 2 \times$ final volume of cell suspension

$$\% \text{ viable cells} = [1.00 - (\text{Number of blue cells} \div \text{Number of total cells})] \times 100$$

2.2.2.5 Purification of CD34⁺ cells

CD34⁺ cells were labeled using CliniMac technology with minor modifications to the manufacturer's instructions. Cells were resuspended in 5 mL (modified to 2 mL for low cell number samples) of MACS and 200 μ L CliniMACS Anti-human CD34 Microbeads per 10^9 MNC were added (reduced to 100 μ L for lower cell numbers). Samples were incubated for 30 minutes at room temperature with gentle agitation on a

platform rocker to prevent antibody sedimentation and reduce cell clumping. The labeled cells were then washed and resuspended in 5 mL of MACS buffer and passed through a pre-washed, high-gradient magnetic separation LS column mounted on a MidiMACS separator. Bound CD34⁺ cells were retained in the column while unlabelled, CD34⁻ cells were flushed through and discarded. Following this, the column was washed 4 times with MACS buffer, removed from the separation stand and the bound cells eluted with 5 mL MACS buffer. Cells were counted (see Section 2.2.2.4) after which an aliquot of 2×10^5 cells were removed for assessment of purity by flow cytometry where the CD34⁺ fraction was consistently above 96%.

2.2.2.6 Exclusion of dead cells

To avoid altered methylation and gene expression patterns associated with programmed cell death, dead and apoptotic cells were eliminated from cultured U937, c6 cells and primary CML MNCs (diagnostic and remission samples) by labeling with MACS Dead Cell Removal MicroBeads and separation over an LS Column in the magnetic field of a MidiMACS Separator according to the manufacturers instructions. Briefly, under sterile conditions 1x Binding Buffer was prepared by dilution of 20x Binding Buffer Stock Solution. Cells were centrifuged, pelleted and resuspended in 100 μ L of Dead Cell Removal MicroBeads per approximately 10^7 total cells and incubated for 15 min at room temperature. The cell suspension was then applied onto a pre washed MS (up to 2×10^8 total cells) or LS (up to 2×10^9 total cells) positive selection column in a MACS separator. Following washing with 1x Binding Buffer the negative cells passed through and collected as the live fraction, which was confirmed using a Trypan Blue exclusion assay (see Section 2.2.2.4).

2.2.3 Flow Cytometry

Flow Cytometry or fluorescence-activated cell sorting (FACS) is a quantitative technique that permits the visualisation and sorting of cells according to their fluorescence intensity after antibody labeling. To characterise and enumerate cells within the HSC and HPC compartment including common myeloid progenitor cells (CMP), megakaryocyte erythroid progenitor cells (MEP), and granulocyte macrophage progenitor cells (GMP) subpopulations, multiparametric immunophenotyping using 6-colour panel assays were performed.

2.2.3.1 Fluorescent labeled antibodies

Antibody specificity/Fluorochrome	Supplier	Catalogue number
Anti-Human Lineage Cocktail 2 (lin 2) - CD3, CD14, CD19, CD20, CD56-FITC	BD Biosciences, Oxford, UK	643397
Anti-Human CD2-FITC	eBioscience, San Diego, CA	11-0029-41
Anti-Human CD4-FITC	eBioscience, San Diego, CA	11-0048-41
Anti-Human CD7-FITC	eBioscience, San Diego, CA	11-0078-41
Anti-Human CD8-FITC	BD Biosciences, Oxford, UK	347313
Anti-Human CD10-FITC	eBioscience, San Diego, CA	11-0108-42
Anti-Human CD11b-FITC	eBioscience, San Diego, CA	11-0118-41
Anti-Human CD235a (Glycophorin A)-FITC	eBioscience, San Diego, CA	11-9987-80
Anti-Human CD123-PE	BD Biosciences, Oxford, UK	555644
Anti-Human CD45RA-APC	eBioscience, San Diego, CA	17-0458-42
Anti-Human CD34-PerCP Cy5.5	BD Biosciences, Oxford, UK	347203
Anti-Human CD38-PE Cy7	BD Biosciences, Oxford, UK	335790
SYTOX Blue Dead Cell Stain	Life Technologies, Carlsbad, CA	S34857

Table 2.1 Fluorescent-labeled antibodies to characterise and enumerate haematopoietic stem and progenitor cells

2.2.3.2 Immunophenotype of haematopoietic stem (HSC) and progenitor cell (HPC) subpopulations

HSCs and HPC subpopulations were identified according to well-established immunophenotypic surface markers as follows:

Cell Type	Marker Definition
Haematopoietic stem cells (HSC)	Lin ⁻ CD34 ⁺ CD38 ⁻
Common myeloid progenitors (CMP)	Lin ⁻ CD34 ⁺ CD38 ⁺ CD123 ^{MED} CD45RA ⁻
Granulocyte macrophage progenitors (GMP)	Lin ⁻ CD34 ⁺ CD38 ⁺ CD123 ^{MED} CD45RA ⁺
Megakaryocyte-erythrocyte progenitors (MEP)	Lin ⁻ CD34 ⁺ CD38 ⁺ CD123 ⁻ CD45RA ⁻

Table 2.2 Immunophenotypic surface markers of haematopoietic stem and progenitor cells

2.2.3.3 Analysis of CD34⁺ hematopoietic stem (HSC)/progenitor cell (HPC) compartment in CML and healthy control samples

The antibody reagent cocktail was prepared in 5 mL round-bottom tubes, immediately prior to use. One hundred microliters of purified CD34⁺ cells were added. Samples were incubated for 30 minutes in the dark at 4°C and washed in 2 mL of staining buffer containing PBS and 0.1% HSA. The supernatant was aspirated, the pellets resuspended in 0.5 mL of staining buffer then placed on ice and analysed immediately.

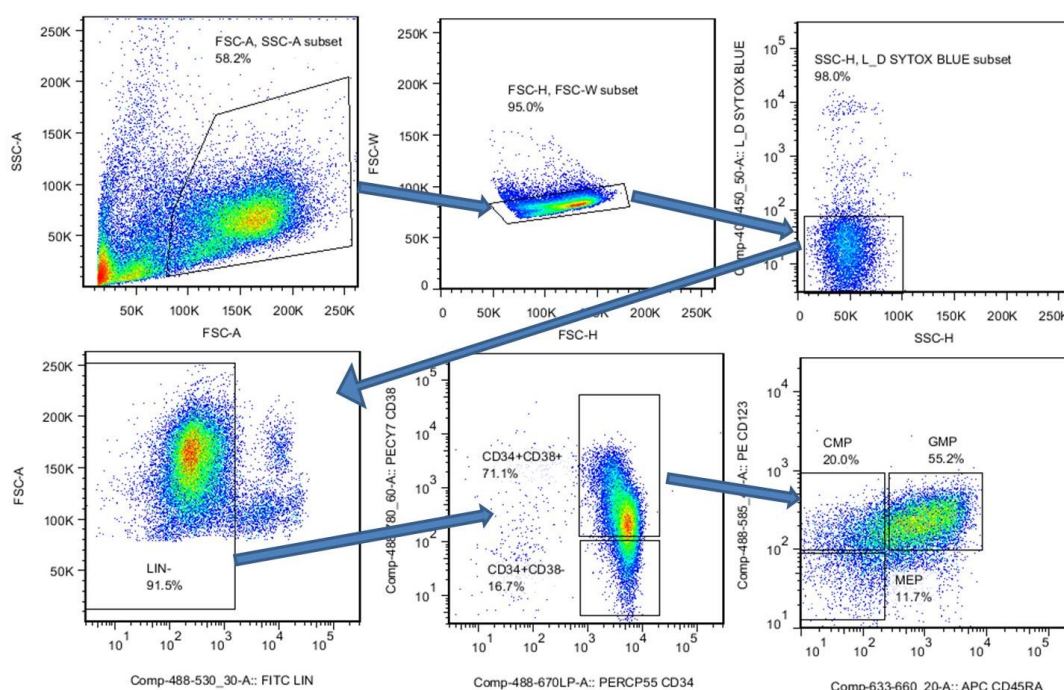
Flow cytometry data was acquired on a 3-laser BD FACSCanto II analyser. The instrument was set up using BD Cytometer Setup and Tracking (CS&T) beads. BD FACSDiva software v6.1.3 was used for data acquisition and all data analysis was performed with FlowJo v7.5 software. Data acquisition for at least 20,000 cells per test sample ensured a sufficient number of positive stained cells.

Compensation for spectral overlap was determined using single stain controls with antibody-capture CompBead particles and the compensation setup tool in BD FACSDiva software.

Isotype-matched controls (IgG1κ and IgG2α) were used at identical concentrations and staining conditions as the target primary antibodies to confirm antibody specificity and rule out non-specific Fc receptor binding. In addition unstained cells and Fluorescence Minus One (FMO) controls defined the background fluorescence and channel boundaries between positive/dull and non-staining cells.

The gating strategy is summarised in Figure 2.1. Gates on forward scatter (FSC) and side scatter, followed by FSC-width (FSC-W) and FSC-height (FSC-H), were used to exclude dead cells and debris. Next, gates were set on live/dead cell discriminator SYTOX Blue so only viable cells were selected. Lin⁻ cells were the true CD34⁺ cells. Viable CD34⁺ cells were further gated into CD34⁺CD38⁻ and CD34⁺CD38⁺ populations. The CD34⁺CD38⁻ events were HSCs and the CD34⁺CD38⁺ events were subdivided on the basis of CD123 and CD45RA expression into CMP, GMP and MEP populations (see Section 2.2.3.2).

Figure 2.1 Gating strategy for haematopoietic stem and progenitor cells



2.2.4 Cell culture

This *in vitro* cell culture work was supervised by Dr Niklas Feldhahn, PhD, Department of Haematology, Imperial College London. Dr Bryant Boulianne, PhD, was responsible for the majority of the cell culture work and Dr George Nteliopoulos, PhD, performed the Western blotting.

To elucidate the influence of BCR-ABL1 expression/activity on aberrant DNA methylation, short-term culture experiments were performed using the conditionally p210BCR-ABL-expressing (tetracycline-inducible promoter) human monocytic cell line U937p210BCR-ABL/c6 (a gift from Dr Wolfgang Seifarth, Medical Faculty Mannheim, Heidelberg University, Germany) [275].

The selection antibiotics hygromycin B and geneticin were added to cell culture medium. Cells were grown in a humidified 5% CO₂ atmosphere at 37°C at an initial concentration of 5×10^5 cells/mL medium and exponentially growing cells were used for all experiments. Culture medium was replaced every 72 hours after cell numbers were determined using a haemocytometer and viability monitored by Trypan blue exclusion assay (see Section 2.2.2.4). Triplicate cultures were set up for all groups. Untransfected U937 monocytic cells (American Type Culture Collection (ATCC), Manassas, VA, USA) were cultured under the same conditions omitting hygromycin B and geneticin. The expression of BCR-ABL1 was induced by adding 1 µg/mL doxycycline to the culture medium of c6 cells and also to U937 control cells with the aim of treating both experimental groups equally and excluding any possible BCR-ABL1 independent effects of doxycycline on DNA methylation. To test the effect of imatinib treatment, 5 µM were added to some groups.

2.2.5 Protein quantification analysis

2.2.5.1 Preparation of cell lysates

To generate protein lysates 5×10^6 cells were removed from the culture medium, washed twice in PBS, and the pellet resuspended in residual PBS. Next, the cells were lysed in 200 μ L ice-cold lysis buffer per 10^7 cells. After a 10-minute incubation on ice the cells were mechanically lysed using a needle and syringe, added to an equal volume of loading buffer and boiled for 5 minutes. Lysates were stored at -80°C for western blot analysis.

2.2.5.2 Western blotting

Samples and pre-stained protein standards/molecular weight markers (Bio-Rad Laboratories) were separated in parallel on SDS-PAGE gels in 1x running buffer at 100 volts before transferring to 0.45 μ m polyvinylidenedifluoride (PVDF) membrane (Immobilon) using a semi-dry transfer system (Bio-Rad Laboratories) for 50 minutes at 18 volts. Membranes were blocked for 1 hour at room temperature in 5% dry milk (Marvel milk powder; Nestle, UK) made up with 1X TBS-T (pH 7.6). Incubation with primary antibodies was performed overnight in 1% milk/TBS-T at 4°C . Primary antibodies used were: Anti-c-Abl (Ab-3) to detect the protein expression of BCR-ABL1 (MW: 210 kDa) and c-ABL (MW: 145 kDa), Anti-Phosphotyrosine (clone 4G10) to detect phosphorylated active Bcr-Abl1 and Anti-Phospho-CrkL (Tyr207) a downstream effector of the Bcr-Abl1 oncoprotein in CML. Anti-Lamin B1 was used as a loading control. After four washes in 1X TBS-T, membranes were incubated for one hour at room temperature with the relevant horseradish peroxidase (HRP)-conjugated secondary antibodies [rabbit anti-mouse or swine anti-rabbit (Dako)] diluted 1:2000 in 1% milk/TBS-T. Membranes were washed four times in 1X TBS-T and protein detection was performed with the enhanced chemiluminescence detection kit according to the manufacturer's instructions on Kodak Biomax MR film.

2.2.6 Nucleic acid extraction

2.2.6.1 DNA extraction from primary CD34⁺ cells and cell lines

Genomic DNA was extracted from CD34⁺ primary cells and cell lines (c6 and U937) using the QIAamp DNA Mini Blood Kit, protocol version 'June 2012'. Briefly up to 5×10^6 CD34⁺ cells in 200 μ L PBS were added to a 1.5 mL microcentrifuge tube containing 20 μ L protease. Following the addition of 4 μ L RNase A stock solution (100 mg/mL), 200 μ L buffer AL (lysis) were added and the sample pulse vortexed for 20 sec. Lysed cells were then incubated at 56°C for 10 min on a heating block. 200 μ L of molecular grade ethanol were added to the sample, which was pulse vortexed, and the mixture applied to a QIAamp Mini spin column and centrifuged at 6000g (8000 rpm) for 1 minute. The filtrate was discarded and the membrane-bound DNA washed with ethanol-containing buffers AW1 and AW2 prior to elution into 50 or 100 μ L Buffer AE (depending upon whether starting number of CD34⁺ cells were greater or less than 1×10^6).

2.2.6.2 Total RNA extraction

Total RNA was extracted using either the RNeasy Plus Micro (protocol version July 2007) or Mini (protocol version, September 2010) Kits based on whether starting CD34⁺ cell numbers were less or more than 5×10^5 respectively.

The cell pellet was resuspended by vortexing in the appropriate volume of RLT plus buffer prepared using β -mercaptoethanol. The lysate was homogenised by centrifugation (13,000 rpm for 2 min), through a QIAshredder spin column then passed through a gDNA Eliminator spin column (13,000 rpm for 30 sec). Following this, 1 volume of 70% molecular grade ethanol was added to the sample and passed through an RNeasy (MinElute in the micro kit) spin column (13,000 rpm for 15 sec) and the flow through discarded. The columns were washed in Buffer RW1 followed by Buffer RPE (13,000 rpm, 15s). Columns processed using the RNeasy Plus Mini kit were washed a further time in Buffer RW1 while 80% ethanol was substituted in the RNeasy Plus Micro protocol (13,000 rpm, 2 min). To ensure adequate drying and avoid ethanol/buffer carry over the column was centrifuged a final time for 1 or 5 mins (mini and micro kits respectively) at full speed. Finally, the spin column was transferred to a collection tube where 14 or 30 μ L (micro and mini respectively) of

RNAse-free water were added before centrifuging for 1 minute to elute the RNA. Following quantification (see **Section 2.2.6.3**), purified RNA was stored in aliquots at -80°C until required.

2.2.6.3 Quantification, and Determination of Quality of nucleic acid

RNA purity (A260/280) and concentration was measured using NanoDrop 2000 spectrophotometer (Thermo scientific). The integrity of the RNA sample, or the proportion of full-length mRNA, can have a significant impact on any subsequent amplification reactions. Reverse transcribing partially degraded mRNAs will generate short cDNA molecules that lack the sequence upstream of the break in the RNA molecule. The relative integrity of mRNA was evaluated using 2200 TapeStation System according to the manufacturer's instructions. By measuring the ratio of the 28S and 18S ribosomal RNAs (rRNA) each experimental sample is assigned an RNA integrity number (RIN) between 1 and 10. RIN values of 8 or above are considered optimal for most gene expression studies. Genomic DNA integrity was checked by running samples on agarose gels prepared using 2% agarose in 0.5X TBE and 5µL ethidium bromide and visualised under UV light. To ensure accuracy and precision, genomic DNA quantitation was assessed using both the NanoDrop 2000 spectrophotometer and the Qubit 2.0 Fluorometer and Qubit Assay Kits (Invitrogen) according to the manufacturers' instructions.

2.2.7 DNA methylation studies

2.2.7.1 *Use of HumanMethylation450 BeadChip array to measure genome wide DNA methylation changes*

The Illumina HumanMethylation450 BeadChip (HM450) is an array-based platform available since 2011 for use in human methylome-wide studies. The DNA methylation status of more than 480,000 single CpG dinucleotides (average of 17 probes per gene) distributed across the entire genome can be scrutinized in an accurate, reproducible and cost-effective manner.

The HM450 is an extension of the Illumina HumanMethylation27 BeadChip (HM27) [276, 277]. In contrast to the HM27, which is designed around a single assay (Infinium I), the HM450 was developed to combine two assays (Infinium I and II) resulting in a far greater representation of CpGs. For both assays, the underpinning principle is the evaluation of cytosine methylation through quantitative ‘genotyping’ of the C/T polymorphism generated following bisulfite conversion.

The Infinium I assay incorporates two beads per probe, one designed to bind to an unmethylated locus, the other to bind to a methylated locus. The signals from both loci are generated in the same color channel. In contrast, Infinium II bead type uses one bead per locus and employs a single base extension step after hybridisation to differentiate converted (unmethylated) from unconverted (methylated) DNA, thereby maximising possible coverage. Here, the methylated and unmethylated signals are generated in the green and the red channels, respectively. Furthermore, Infinium II chemistry utilises degenerate oligonucleotides (a group of oligonucleotides designed to bind to the same genomic location while allowing for specific single base differences), designed against all possible combinations of methylation status across the CpG sites that fall within a single probe sequence (up to 3 CpGs). In comparison, Infinium I assumes identical methylation status in all CpG sites present within a single given probe.

Coverage of HM450 is extensive and includes approximately 99% of RefSeq genes, including promoters, 5'-UTR, first exons, gene bodies and 3'-UTR regions as well as 96% of all CpG islands (in addition to multiple targets around these regions to

incorporate CpG shores and flanking regions), CpG sites outside of CpG islands, micro RNA promoter regions, DNase hypersensitive sites and disease/cancer-associated regions previously identified through genome-wide association studies (GWAS).

2.2.7.2 Bisulfite conversion of genomic DNA

The bisulfite conversion reaction enables single-base discrimination of methylation status at CpG dinucleotides. During this reaction, DNA is treated with sodium bisulfite resulting in preferential deamination of unmethylated cytosines to uracil, via sulphonation, hydrolytic deamination, and desulphonation. Methylated cytosines remain unchanged during the treatment and therefore a single base difference can be detected using a variety of sequencing methods. Genomic DNA was treated with sodium bisulfite using the Zymo EZ-96 DNA Methylation Kit (Deep-Well format) according to the manufacturer's instructions, utilising the alternative incubation conditions recommended specifically for the Illumina Infinium Methylation Assay.

5 µL of M-Dilution Buffer were mixed with each DNA sample (500 ng total DNA in 45 µL Buffer AE) in a 96-well conversion plate and incubated at 37°C for 15 minutes. Following the addition of 100 µL CT Conversion Reagent to each well, the reaction was incubated in a thermocycler overnight [(95°C for 30 sec, 50°C for 60 min) x 16 cycles] then placed on ice for 15 minutes. A Zymo-Spin I-96 Binding Plate was prepared by placing onto a collection plate and pipetting 400 µL of M-Binding Buffer to each sample well. The samples were transferred from the conversion plate and loaded into the wells of the binding plate containing the M-Binding Buffer, sealed carefully with adhesive film and mixed by inverting several times before centrifuging at 3,000 rpm for 30 seconds. The flow-through was discarded and 500 µL of M-Wash Buffer were added followed by centrifugation for 30 seconds. 200 µL of M-Desulphonation Buffer were then added to each well. The plate was left to stand at room temperature for 20 minutes before centrifuging for 30 seconds. The wells were washed twice using 500 µL of M-Wash Buffer, placed on an elution plate and the converted DNA eluted in 15 µL of M-Elution Buffer. The recovered bisulfite treated DNA was quantified by NanoDrop 2000 spectrophotometer (Thermo scientific) using the RNA setting (The absorption coefficient of single stranded bisulfite converted DNA at 260 nm resembles that of RNA) and stored at -20°C in the short term until required. For experiments where smaller sample numbers were being processed, the

EZ DNA Methylation Kit, a spin column-based kit was used with minor protocol modifications including smaller volumes of M-Wash Buffer as well as a reduced final elution volume of 10 μ L.

2.2.7.3 *Illumina Infinium Methylation Assay*

Hybridisation of bisulphite-treated DNA to the BeadChips was performed at the Genome Centre, at the Barts and London School of Medicine and Dentistry, Guy's Genomics Facility, King's College London, and UCL Genomics at the Institute of Child Health using standard Illumina protocols, which involved the following steps;

- Denaturation and neutralisation of bisulphite-converted DNA
- Isothermal amplification by overnight incubation
- Enzymatic fragmentation of bisulphite-converted DNA using end-point fragmentation to avoid over-treatment
- DNA precipitation using 2-propanol and PM1 solution followed by centrifugation
- Re-suspension of DNA in RA1 hybridisation buffer and loading on to BeadChips (twelve samples per chip)
- Overnight hybridisation
- Washing to remove unhybridised and non-specifically hybridized DNA sample from the BeadChips
- Single base extension using hybridized DNA as a template through the addition of biotin (ddCTP and ddGTP) and 2,4-dinitrophenol (DNP) (ddATP and ddTTP) labeled nucleotides
- Immunohistochemical staining using a combination of antibodies that differentiate DNP from biotin
- Imaging BeadChip on the Illumina iScan, a dual laser (wavelengths at 532 and 658 nm) fluorescent scanner with a 0.53-micron spatial resolution and dual-channel collection capacity.

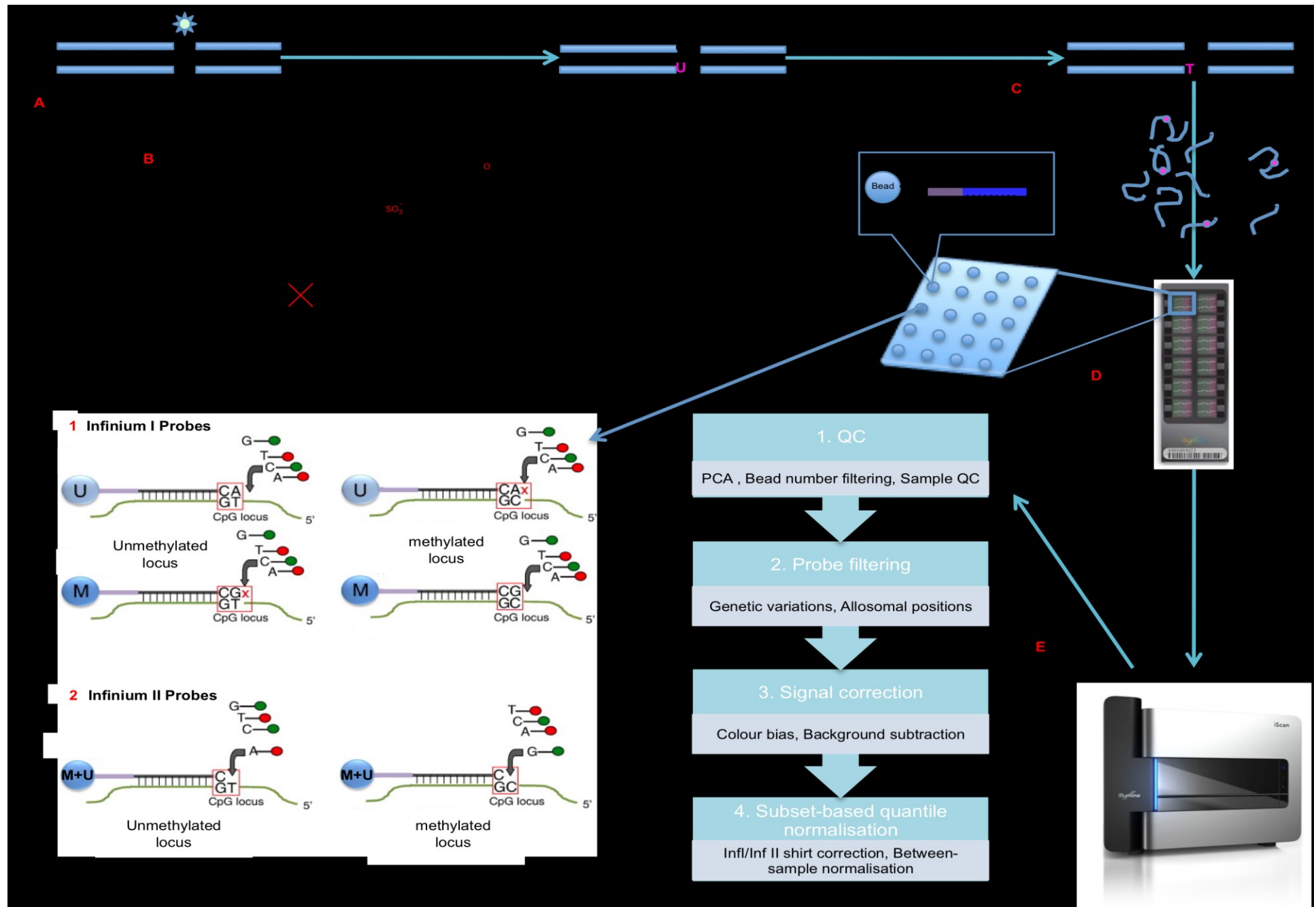


Figure 2.2 Overview of the Illumina Infinium methylation assay

(A) Genomic DNA (500 ng) is treated with sodium bisulfite (HSO_3^-) using the Zymo EZ DNA Methylation Kit **(B)** which results in the conversion of unmethylated cytosine (C) to uracil (U) through sulphonation, deamination and desulphonation (lower strand). Methylated cytosines are completely resistant to sulphonation and therefore do not convert to uracil (upper strand). This enables downstream sequencing-based experiments to differentiate between methylated and unmethylated cytosines. Following whole genome amplification the analysis of DNA methylation is reduced to single nucleotide polymorphism (SNP) analysis for thymidine (T) and cytosine (C); where C is found, the original cytosine was methylated and where T is present, non-methylated **(C)**. The HM450 array allows the interrogation of 485,836 single CpG dinucleotides in the human genome. The methylation assay was performed on 4 μL bisulfite-converted genomic DNA (50 ng/ μL) with a throughput of 12 patient samples per slide **(D)**. The array is composed of beads with an average 800,000 oligonucleotides clustered on each bead which self-assemble and are immobilised into picoliter sized pits. Each oligo consists of a 50-base (b) gene-specific probe linked to a 23-b address code, used to identify the bead location on the chip to create the .dmap file. The HM450 contains two different chemical assays carrying information about the methylation state using Type I probes, developed first for the Infinium 27K BeadChip **(1)** and Type II probes **(2)**. Type I probes require two beads to differentiate between methylation status of any given loci. Base extension is the same for methylated or unmethylated alleles since the fluorescent signal does not carry any information on the methylation status. Differentiation of unconverted and converted cytosines (T) is by the failure or success of single base extension beyond the probe sequence. Type II probes (72% of all probes represented on HM450) are attached to a single bead with a degenerate base. Methylation information of each individual CpG site is obtained through dual channel single-nucleotide primer extension with labeled dideoxynucleotides on the methylation variable position of a CpG. For both bead types, the incorporation of a fluorophore-labeled base allows a colour-based detection of sequence by laser excitement using the Illumina iScan. The preprocessing pipeline comprises four main steps **(E)** beginning with raw extracted data and leading to ready-to-analyse methylation information. *(Diagram in lower left corner adapted from Bibikova.[276])*

2.2.8 Whole-Genome Gene Expression profiling using Illumina Direct Hybridization Assay system (Direct Hybridization Assay)

This work was performed by myself at Guy's Genomics Facility, King's College London under the supervision of Mr Eric Nasser.

2.2.8.1 Illumina TotalPrep-96 RNA Amplification Procedure

During this initial step, extracted total RNA is converted to double-stranded cDNA, followed by an amplification step (*in vitro* transcription (or IVT)) to generate biotin labeled cRNA.

To synthesize the first strand of cDNA by reverse transcription, up to 500 ng of RNA in a final volume of 11 μ L in nuclease-free water were placed into the wells of a PCR plate. A reverse transcription master mix was prepared for $n + 10\%$ samples as follows;

2 μ L 10X First Strand Buffer
4 μ L dNTP Mix
1 μ L T7 Oligo(dT) Primer
1 μ L RNase Inhibitor
1 μ L ArrayScript Reverse Transcriptase

9- μ L reverse transcription master mix were added to each sample. The plate was covered and incubated for 2 hours at 42°C in a thermal cycler then held at 4°C/placed on ice. A second strand master mix was prepared on ice for $n + 10\%$ samples as follows;

63 μ L Nuclease-free Water
10 μ L 10X Second Strand Buffer
4 μ L dNTP Mix
2 μ L DNA Polymerase
1 μ L RNase H

80 μ L of the second strand mix were added to each sample in the PCR Plate and incubated at 16°C for 2 hours. Following thorough resuspension, 180 μ L of room temperature-equilibrated cDNApure were added to each well before transferring the

samples from the PCR plate to a U-bottom plate. To ensure adequate binding of cDNA to the cDNAPure magnetic beads, the plate was placed on a shaker and gently agitated for 2 minutes. The plate was then placed on a magnetic stand for 7 minutes during which time the mixture became transparent and the beads formed pellets against the magnets. The supernatant was carefully aspirated without disturbing the magnetic beads and discarded. The plate was removed from the stand and washed twice with 150 μ L cDNA Wash Buffer. To evaporate the residual ethanol from the wells, the plate was vigorously shaken for a further 2 minutes following the second wash and capture.

To elute the cDNA from the magnetic beads 20 μ L of preheated (55°C) nuclease-free water were added to each sample. The plate was vigorously shaken for 4 minutes until the magnetic beads were fully dispersed then move to the magnetic stand for magnetic bead capture until the mixture became transparent. As the final step in cDNA purification, 17.5 μ L of the supernatant, containing the eluted cDNA was placed into the wells of a new PCR plate.

For in vitro cRNA synthesis the IVT master mix was prepared for $n + 10\%$ samples as follows;

2.5 μ L Biotin-NTP Mix

2.5 μ L T7 10X Reaction Buffer

2.5 μ L T7 Enzyme Mix

7.5 μ L of IVT master mix were added to each sample, mixed, covered with aluminum sealing foil and incubated at 37°C for 14 hours. The following day, 70 μ L cRNA binding mix and 95 μ L 100% ethanol were added to each sample before transferring to a clean U-bottom plate with gentle shaking for 3 minutes to facilitate the binding of cRNA in the samples to the binding beads. In the same way described above, the plate was placed on the magnetic stand for 7 minutes to enable capture of the bead-bound cRNA while discarding the supernatant. After two wash/recapture steps using 100 μ L of cRNA wash solution the purified cRNA were eluted from the RNA Binding Beads by adding 50 μ L preheated (55°C) cRNA Elution Buffer to each sample. cRNA concentration was determined by measuring sample absorbance at 260 nm using a NanoDrop Spectrophotometer and size distribution evaluated using 2200 TapeStation System (expected range 250-5500 nucleotides).

2.2.8.2 Direct Hybridization Assay

During this part of the procedure, the amplified pool of biotin-labeled cRNA corresponding to the polyadenylated mRNA fraction is normalized and applied to Illumina HumanHT-12 v4.0 Expression BeadChip. The RNA-loaded BeadChips (12 samples per chip) are placed in Hybridisation Chambers and incubated in a Hybridization Oven for 14–20 hours at 58°C.

The cRNA sample tubes were preheated to 65°C for 5 minutes, mixed by pulse vortexing, centrifuged down briefly and left to cool to room temperature. Next, 750 ng of each cRNA sample were added to a hybridization tube and made up to 5 µL using RNase-free water. To each sample 10 µL of HYB were then added. Following this the Hybridisation Chambers were carefully assembled, 200 µL of HCB dispensed into the humidifying buffer reservoirs. To prepare the BeadChips for hybridisation each BeadChip was placed in a Hybridisation Chamber Insert in the correct orientation and the 15-µL samples loaded on to the center of each inlet port. Four Hybridisation Chamber Inserts (containing the sample-laden BeadChips) were placed into each Hybridisation Chamber, and incubated for 18 hours at 58°C on a rocker in the Hybridisation Oven. The following day, the BeadChips were removed from the oven and chambers then washed which involved the following steps;

- Seal removal while BeadChips remains submerged in diluted Wash E1BC buffer
- High temp wash at 55°C using diluted High-Temp Wash buffer in the Hybex water bath, 10 minutes
- First room-temp wash in diluted Wash E1BC buffer, 5 minutes
- Ethanol wash, 10 minutes
- Second room-temp wash by plunging the rack in and out of diluted Wash E1BC buffer 10 times followed by further 2 minute submersion
- Blocking using Block E1 buffer, 10 minutes.

Following the above washing and blocking steps, Cy3-Streptavidin was added in order to bind the analytical probes on the BeadChip and allow differential detection of signals during scanning. Briefly each BeadChip was transferred to the Block E1 buffer/Cy3-Streptavidin-containing wash tray, covered and placed on a rocker mixer (medium speed) for 10 minutes. This was followed by a wash in diluted Wash E1BC

Buffer, drying by centrifugation at 1,400 rpm, at room temperature for 4 minutes and scanning on the iScan System within one hour of the final room temperature wash.

The iScan Reader uses a laser to excite the fluor of the single-base extension product on the beads of the BeadChip sections. Light emissions from these fluors are then recorded in high-resolution images of the BeadChip sections. Data from these images were initially downloaded to Illumina's GenomeStudio Gene Expression Module.

2.2.9 Mutation analysis by targeted sequencing using Ion Torrent technology

Dr George Nteliopoulos, PhD carried out the library preparation in the Molecular Unit of the Imperial Molecular Pathology Laboratory, Hammersmith Hospital, London under the supervision of Dr Gareth Gerrard, PhD and Miss Alexandra Foong. In addition Miss Alexandra Foong was responsible for writing and updating the standard operating procedure associated with the workflow.

The Ion Personal Genome Machine (PGM) sequencer uses a sequencing-by-synthesis method and emulsion PCR similar to other pyrosequencing-based platforms but rather than measuring light released from chemiluminescent reagents, it measures the direct release of H^+ ions during base incorporation by semiconductor sequencing technology.

2.2.9.1 PCR amplify genomic DNA targets

The first step in the workflow was library construction. The process involved taking the genomic DNA from CML patient samples, fragmenting it to a uniform size and then adding sequencing adapters.

A master mix was prepared for $n + 1$ reactions;

4 μ L 5x Ion AmpliSeq HiFi Mix

10 μ L 2x Ion AmpliSeq Primer Pool (custom)

Next, 10 ng gDNA made up to 6 μ L in nuclease-free dH₂O were added to single well of a 96-well PCR plate into which 16 μ L of the PCR master mix were added. The plate was sealed, vortexed and briefly spun before placing in a G-Storm (GS4) thermal cycler

- (1) 99°C for 2 minutes (enzyme activation)
- (2) 22 cycles of: 99°C for 15 sec (denature) and 60°C for 4 minutes (anneal/extend)
- (3) 10°C ∞

2.2.9.2 Partially digest primer sequences

Following the PCR step, the plate was spun down briefly; the film seal removed and 2 µL FuPa Reagent were added to each amplified sample before resealing and placing back in thermal cycler to run the following short program;

- (1) 50°C for 10 minutes
- (2) 55°C for 10 minutes
- (3) 60°C for 20 minutes
- (4) 10°C Hold (for maximum 60 minutes)

2.2.9.3 Ligate barcode adapters to the amplicons and purify

For each adapter barcode used, a 1:4 dilution was prepared (stored at -20°C) of Ion P1 Adapter and Ion Xpress Barcode X per 4 reactions as indicated;

- 2 µL Ion P1 Adapter
- 2 µL Ion Xpress Barcode X
- 4 µL Nuclease-free water

To set up and run the ligation reaction, 4 µL Switch Solution together with 2 µL of diluted barcode adapter mix and 2 µL of DNA Ligase were added to each digested sample. The plate was covered with adhesive film, vortexed thoroughly, and spun briefly before loading in to the thermal cycler to run the following incubation:

- (1) 22°C for 30 minutes
- (2) 72°C for 10 minutes
- (3) 10°C Hold (for maximum 60 minutes)

To purify the unamplified library 45 µL of Agencourt AMPure XP Reagent (room temperature for 30 mins and vortexed to disperse the beads) were added to each library, pipetted and incubated for 5 minutes at room temperature after covering with protective film and spinning down. The plate was then placed on a DynaMag-96 Side Magnet for 3 minutes, until the solution cleared. The supernatant was removed

without disturbing the pellet and 150 μ L of freshly prepared 70% ethanol added while and the plate was moved from side to side in the two magnet positions to wash the beads. The supernatant was discarded without disturbing the pellet and the process repeated for a second wash. The plate was left in the magnet to allow the beads to air-dry at room temperature for 5 minutes avoiding over drying. Finally, the plate was removed from the magnet, 50 μ L of Low TE were added to each pellet to disperse the beads, the plate was sealed, vortexed thoroughly and spun down.

2.2.9.4 Quantify the unamplified library

Accurate quantification of amplifiable library molecules was important to avoid over- or underestimation of DNA-bearing beads and ensure equal representation of each library. This was achieved using Kapa Sybr Fast Universal qPCR Kit For Ion Torrent platform. Quantification was achieved by inference from a standard curve generated using six supplied DNA Standards according to the manufacturers instructions.

The qPCR mix was prepared as follows:

6 μ L KAPA SYBR FAST qPCR Master Mix containing Primer Premix
2 μ L ddH₂O
2 μ L Diluted library DNA or DNA standard (1-6)
(10 μ L Total)

Using the run conditions:

- (1) 95°C for 3 minutes (initial activation/denaturation)
- (2) 35 cycles of: 95°C for 3 seconds (denaturation) and 60°C for 25 seconds (annealing/extension).

Based on the calculated qPCR template dilution factor, the sample libraries were diluted to a concentration of 20 pM with low TE, combined as 20 μ L single pool in a LoBind tube (Eppendorf) and made up to a final volume of 25 μ L.

2.2.9.5 Clonal amplification by emulsion PCR

This part of the workflow was fully automated by way of the Ion OneTouch 2 System (including Ion OneTouch 2 Instrument and the Ion OneTouch ES Instrument) using the Ion OneTouch 200 Template Kit v2 DL according to the manufacturer's instructions consisting of the following steps

- Clonal amplification of generated barcoded library templates onto Ion PI Ion Sphere Particles (ISPs) within oil microdroplets using emulsion PCR.
- ‘Breaking’ the emulsion by organic extraction and centrifugation
- Template-positive ISP enrichment/recovery in a glycerol gradient with the Ion OneTouch ES and enrichment bead kit.
- Quality control check with Ion Sphere Quality Control Kit using the Qubit 2.0 Fluorometer.

2.2.9.6 Ion Torrent Personal Genome Machine (PGM) sequencing

The polyclonal percentage and quality of enriched, template-positive ISPs were determined using the Ion Sphere Quality Control Kit (Life Technologies) and the Qubit 2.0 Fluorometer. Samples with less than 30% polyclonality and enriched, template-positive ISPs greater than >80% were deposited on an Ion 318 Chip v2 and sequenced by the Ion Personal Genome Machine (PGM) System using Ion PGM Sequencing 200 kit v2 (Life Technologies). A cut-off with a quality score of Q17 (2% errors, corresponding to 1 base error per 50 bases) was used as a measure of sequencing success.

2.2.9.7 Sanger sequencing validation

Mutations identified were validated using Sanger sequencing on a 3500 Genetic Analyser (Applied Biosystems) (performed by Miss Alexandria Foong).

2.2.10 Statistical tests

Basic statistical analysis including calculation of means, medians, standard deviations and confidence intervals was performed using Microsoft Excel 2010 and GraphPad Prism 5.0.

2.2.10.1 Power calculation

Power calculations were determined using the `pwr.t2n.test` function in the R, `pwr` package. The effect size d was estimated using a power of 95% and a genome-wide significance level $1e-8$, resulting in an effect size $d = 2.12$. To calculate the corresponding beta difference to this effect size the standard deviation of each probe

for CML-CP and healthy control sample were determined. The mean SD was then used to derive an expected beta value difference of 0.082. In conclusion, the study was appropriately powered to detect beta value differences of 0.2, which formed the cut-off for CML-CP DMPs.

NCML-CP= 44

Ncontrol= 20

2.2.10.2 Chi-Squared test

This non-parametric, goodness-of-fit test was used to test whether there were non-random associations between categorical variables by comparing the observed and expected frequencies in each. The Fisher's exact test was used when the smallest expected value was less than 5.

2.2.10.3 Analysis of variance (ANOVA)

ANOVA was used to analyse the differences between several group (three or more) means (groups or variables) for statistical significance. $P < 0.05$ was considered significant.

2.2.10.4 Multiple comparison correction

A major consideration for multiple statistical testing relates to the fraction of false positives results. This is especially relevant with respect to microarray data, which generate quantities for tens of thousands of genes. For example, if 10^4 statistical tests are generated and for all of them the null hypothesis is actually true, it is expected that 500 tests will be significant if $p < 0.05$ by chance alone. While there is no universally accepted approach for dealing with the problem of multiple comparisons, for the purposes of this work two statistical methods were used to address the issue.

The Bonferroni correction was used to control the family-wise error rate on small numbers of multiple comparisons. Here, the critical value (alpha) for an individual test was found by dividing the family-wise error rate (usually 0.05) by the number of tests. Therefore when 100 statistical tests were performed, the critical value for an individual test was $0.05/100=0.0005$. However, with large numbers of multiple comparisons e.g expression levels of 2×10^4 genes, only genes with huge differences in expression will have $p < 0.0000025$ ($0.05/20000$) leading to a high false negative rate and important differences being missed.

The Benjamini–Hochberg procedure [278] was the approach used to control the proportion of false positives or the false discovery rate (FDR) for microarray data in this study. The Benjamini-Hochberg critical value was calculated by ranking unadjusted p-values within the dataset in order, from smallest to largest; the smallest p-value having a rank (i) of $i=1$, dividing this by m, the total number of tests, and multiplying by Q, the false discovery rate, $(i/m)Q$. For the purposes of this work the FDR was 0.01 (1%) while statistical significance remained $p < 0.05$.

2.2.11 Bioinformatic analysis

Dr R. Lowe (Blizard Institute, Barts and The London School of Medicine and Dentistry, London, UK) performed **all** analyses unless otherwise stated.

2.2.11.1 Analysis of 450k DNA methylation data

Data were extracted from Illumina's BeadScan software as idat files and imported into R (2.14.1) [279] for use with the Bioconductor package (<http://www.bioconductor.org>) minfi [280]. All samples underwent quality control (QC) checks to validate various steps including bisulfite conversion, staining, extension and hybridisation. Sex chromosomes and those autosomal probes mapping with 90% accuracy to a sex chromosome were removed from the main analysis.

Samples were normalized by quantile normalization of Type I Methylated (Red), Type I Methylated (Green), Type I Unmethylated (Red), Type I Unmethylated (Green), Type II Methylated and Type II Unmethylated separately. Transformed beta values (M-values) were calculated using the log₂ ratio of the intensities of the methylated channel versus unmethylated channel at the interrogated CpG site. Since the logit transformation of the beta value, the M value reduces the severe heteroscedasticity associated within subpopulations and permits the assumption during testing that variances within groups are equal [281]. Differentially methylated positions (DMP) were calculated using an F-test with a genome-wide corrected p-value < 0.01 . Correlations were calculated using a Pearson's product moment correlation co-efficient. Permutation tests were used to calculate fold enrichments and associated p-values (10,000 permutations were used in all cases). The eBayes and ImFit functions in limma were used to incorporate patient age, chip number and

position as covariates in the analysis for correction to find probes that were differentially methylated between the various categorical groups with a FDR 1% (See section 2.2.10.4) [282].

The probes were mapped to the February 2009 human reference sequence (GRCh37/hg19) genome [283] using the BLAST-like alignment tool (BLAT) [284]. In order to functionally annotate the differentially methylated genes (corresponding to the genomic location of DMPs), the Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.7 tools were accessed [285]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [286] and Gene Ontology (GO) enrichment analysis [287] were specifically utilised for pathway analysis of these genes (considering pathways with Benjamini–Hochberg corrected $p < 0.01$, $> 10X$ enrichment).

Meta-analyses were performed using Marmal-aid [288]. Developed in-house, Marmal-aid (<http://marmal-aid.org>) is a standardised database that incorporates all publicly available HM450 data into a single repository to enable customizable methylation comparisons of defined genomic regions between various reference tissues and generated data [289].

2.2.11.2 Analysis of HT-12 gene expression data

After image processing, raw (pixel-level data from TIFF images) BeadArray expression data were exported as tab-delimited Illumina idat files from GenomeStudio for processing in the R software environment. Arrays were analysed using the lumi package from the open-source Bioconductor project. Data were transformed using the variance stabilising transform and normalised using Robust Spline Normalisation (RSN). Differentially expressed genes between the various categorical groups were identified using a moderated t-test in *limma* [282] with a FDR 1%. Following differential expression analyses, probes were annotated according to their manufacturer-assigned and RefSeq IDs. Those probes assigned a bad or no match quality score were discarded. Downstream analysis of the gene expression data including gene ontology and pathway enrichment analyses were performed using GO and KEGG tools (see Section 2.2.11.1).

2.2.11.3 Analysis of Ion Torrent targeted gene sequencing data

Raw data from sequencing runs were automatically transferred to the Torrent Server, Torrent Suite where Ion Reporter Software was used to perform variant calling and read mapping alignment against human reference sequence (GRCh37/hg19). A number of steps were used to filter/exclude nucleotide variants identified in the screening including a) those not annotated as pathogenic variants (b) those variants called in both CML and healthy controls and (c) variants representing probable mapping ambiguities. Annotation was performed using Oncotator, A Cancer Genome Atlas (CGA) web-based application interpretation tool for annotating human genomic point mutations and indels [290].

3 Identification of a unique DNA methylation signature in CML cells

3.1 Background

Chronic myeloid leukaemia (CML) is a myeloproliferative neoplasm thought to originate from the acquisition of a BCR-ABL1 fusion gene by a single pluripotent stem cell, whose progeny then gain a major proliferative advantage over their normal counterparts.

Despite this consistent event, CML is clinically heterogeneous. This is exemplified by the unpredictable leukocyte doubling times in untreated patients and the striking variability in response to treatment with tyrosine kinase inhibitors (TKI). Although the majority of patients who commence imatinib respond well, approximately 30% of those newly diagnosed with CML do not achieve optimal responses and acquire clinical resistance while others demonstrate primary drug failure or experience intolerable drug-related toxicities [46].

While a relatively small proportion of CML-CP patients develop secondary resistance through BCR-ABL1 kinase domain mutations, the remaining cases have no clear aetiology [51]. Numerous studies have explored the possibility that genetic aberrations in addition to BCR-ABL1 such as mutations, copy number variation, or other genomic rearrangements are responsible but with inconsistent results.

Though attention has focused mainly on the genetic events underlying carcinogenesis, it is increasingly apparent that epigenetic alterations also influence the natural history of malignant disease. Disruption to global DNA methylation is detected early in leukaemia suggesting a critical role for this process in initiating, maintaining and driving the malignant phenotype [291, 292]. Tumour suppressor genes are frequently silenced by CGI-associated promoter hypermethylation, while hypomethylation leads to activation of oncogenes and transposable elements located within intronic or intergenic regions [116].

Figuerola and colleagues generated DNA methylation profiles for 344 patients with acute myeloid leukaemia (AML) and identified a 15-gene methylation classifier that could separate the disease into 16 subgroups and predict for overall survival [293]. While in chronic lymphocytic leukaemia (CLL), extensive characterisation of the methylomes from 139 patients confirmed distinct patterns corresponding to the somatic mutated (M-CLL) and non-mutated (U-CLL) immunoglobulin heavy chain variable region (IGHV) gene status, reflecting the differentiation stage from which these subtypes were derived [294]. When applied to a multivariate regression model, DNA methylation signature together with CD38 expression and lactate dehydrogenase levels were the only factors with independent clinical prognostic value.

Aberrant epigenetic reprogramming might contribute to CML pathogenesis or underlie the variations in responses and outcome to a given TKI therapy. This chapter focuses on the identification of a CML methylation signature using a genome-wide approach in a cohort of clinically well-characterised patients and explores the possibility of whether such profiling can be used to accurately predict an individual's response to treatment.

DNA methylation is an attractive early diagnostic marker for predicting clinical response and offers several practical advantages over genetic mutation approaches. Firstly, aberrant DNA methylation occurs at a higher frequency in cancer when compared with genetic alterations and has been correlated with gene expression changes for a number of TS genes supporting its use as an early cancer detection-screening tool. Where mutational analysis often necessitates whole gene studies (point mutations can be detected at various locations along the locus), DNA methylation assessments tend to be more focused and involve smaller regions e.g. CGI hypermethylation of the promoter.

When stored appropriately, DNA from clinical samples is stable and fairly resilient to degradation. As technology and platforms have improved, it has become possible to use smaller starting amounts of material while increasing throughput at a relatively low cost. Furthermore, levels of DNA methylation can be compared accurately to absolute reference points i.e. a CpG site can be 100% methylated or entirely unmethylated which greatly simplifies the design of any assay. Lastly and most

importantly for patients, acquired epigenetic changes are potentially reversible which provides the opportunity for therapeutic intervention.

3.2 Hypothesis

Aberrant DNA methylation contributes to CML pathogenesis and underlies the variation in response and outcomes on imatinib therapy.

3.3 Objectives

1. To generate and validate epigenome-wide DNA methylation profiles of CML CD34⁺ cells using the Illumina HM450 array, with no *a priori* hypothesis as to where this variation, if any would be occurring.
2. To identify and validate the DNA methylation differences between CML-CP haematopoietic progenitor cells (HPC) and their normal counterparts.
3. To define differences in the DNA methylome of HPCs from patients who respond to TKI therapy with those who do not.
4. To perform relevant pathway analyses of any identified methylation perturbations in CML.
5. To compare the methylation variation between CML and other blood and solid cancer types.

3.4 Results

3.4.1 Patients

The institutional research ethics committee approved the study and all subjects provided written informed consent in accordance with the Declaration of Helsinki. Copies of the signed consent forms were filed in the patient notes and stem cell laboratory. For patients enrolled on the STI571 Prospective International Randomised Trial (SPIRIT 2) Study, further ethical approval was sought from the National Cancer Research Institute (NCRI) CML subgroup for access to the SPIRIT 2 biobank.

Cryopreserved mononuclear cells (MNCs) from 46 untreated patients with newly diagnosed CML in the chronic phase (CML-CP) who started treatment with imatinib at the Hammersmith Hospital between April 1999 and August 2009 were available for this study (Table 3.1). Patients were classified as responders (n=29) if they subsequently achieved complete cytogenetic response (CCyR) within 12 months of commencing imatinib therapy at a starting dose of 400 mg/day. Non-responders (n=17) never attained major cytogenetic response (MCyR) (n=12) or lost a previously attained CCyR despite continuing imatinib (n=5).

All responders underwent granulocyte-colony-stimulating factor (G-CSF) mobilisation of Ph^- peripheral blood stem cells after a median of 23 months on imatinib therapy (range, 13-55 months) as an additional strategy for reducing the number of contaminating Ph^+ stem cells with the intention of improving autologous peripheral blood stem cell transplant (PBSCT) outcome if required sometime in the future [295]. In the event, all patients maintained a durable response to TKI therapy; no cells were used for this purpose and made available for this study.

Healthy controls consisted of cryopreserved G-CSF mobilized peripheral blood MNCs collected (in excess requirement) from 20 healthy donors between December 2000 and May 2011 destined for allogeneic stem cell transplant procedures (Table 3.2).

The validation group consisted of a further 48 diagnostic CML-CP cryopreserved samples collected between 2009 and 2011 from TKI-naïve patients enrolled on the

STI571 Prospective International Randomised Trial (SPIRIT 2) Study (EudraCT: 2007-006185-15), a Phase III, multicentre, open-label, prospective randomized trial comparing imatinib (Glivec) 400 mg/day versus dasatinib (Sprycel) 100mg/day (Table 3.3). 30 patients (15 imatinib, 15 dasatinib) achieved CCyR within 12 months from the commencement of therapy and were classified as responders whilst 18 (9 dasatinib and 9 imatinib) non-responders failed to achieve this 12-month milestone. The difference in criteria for non-response between the test and validation set relates to evolving guideline definitions and shorter follow-up for the SPIRIT 2 cohort where patients failing to respond optimally to first line therapy were withdrawn from study.

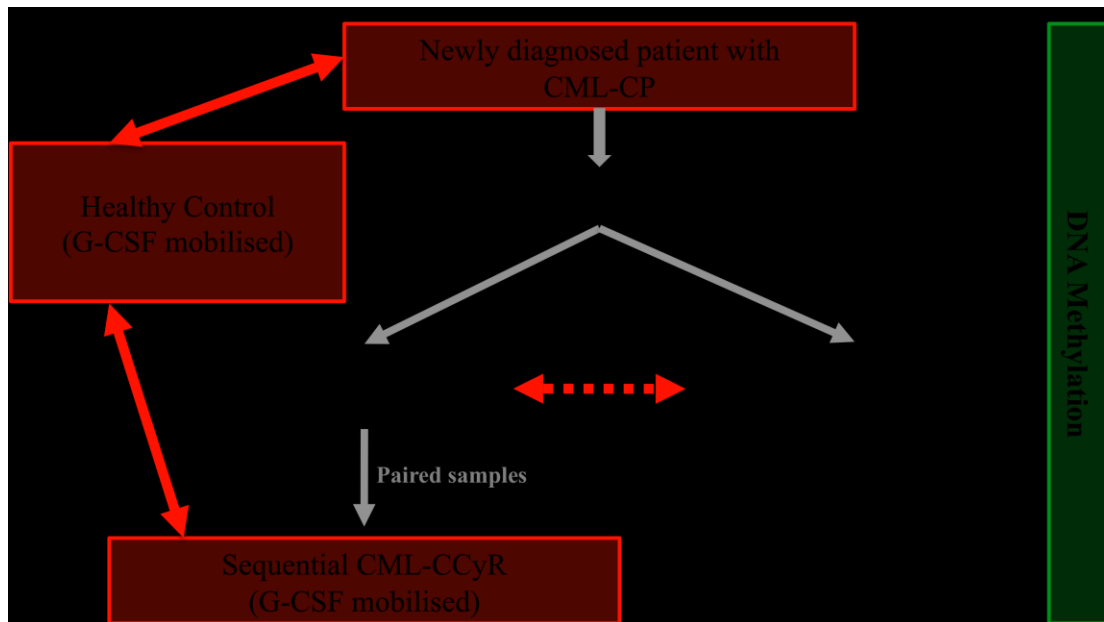


Figure 3.1 Overall strategy for CML patient study

Diagnostic samples from patients (**n=46**) were collected prior to tyrosine kinase inhibitor treatment. Subsequently these patients either failed (**n=17**) or responded (**n=29**) to first line therapy (**n=29**). Paired samples were collected from those individual patients who achieved at least a complete cytogenetic response (CCyR) (**n=29**). Overall this allowed genome-wide DNA methylation comparisons to be made between CML patients and healthy controls as well as healthy controls and patients at the time of CCyR (red solid arrows). Any differences in methylation patterns within the diagnostic samples might explain subsequent response to TKI therapy (red broken line).

Characteristic	Responders (n=29)	Non-responders (n=17)
Age, years		
At diagnosis		
Median	46	47.5
Range	24-65	20-67
Sex, (no, %)		
Male	14 (48.3)	12 (70.6)
Sokal risk group, (no, %)		
Low	12 (41.4)	2 (11.8)
Intermediate	13 (44.8)	5 (29.4)
High	3 (10.3)	10 (58.8)
Missing	1 (3.4)	NA
BCR-ABL1 transcript type, (no, %)		
e13a2	11 (37.9)	6 (35.3)
e14a2	15 (51.7)	9 (52.9)
e13a2 and e14a2	3 (10.3)	2 (11.8)
BCR-ABL1/ABL1 ratio, %		
At diagnosis		
Median	71.3	72.5
Range	13.2-120.3	30-126.4
Time to major molecular response (MMR), months		-
Median	14	-
Range	3-42	-
Time from diagnosis to imatinib therapy, months		
Median	3.5	4.2
Range	0-7.2	0-12.4
Therapy prior to imatinib, (no, %)		
Hydroxycarbamide	18 (62.1)	10 (58.8)
Interferon	1 (3.4)	3 (17.6)

Table 3.1 Biological and Clinical Characteristics of Hammersmith Hospital Patients (Discovery Cohort)

Patient Identifier	Age at mobilisation	Sex	CD34% purity	HM450	HT-12
A1	23	M	98.0	X	X
A3	31	M	97.5	X	-
A4	44	F	99.0	X	X
A5	33	M	97.0	X	X
A7	20	M	97.0	X	X
A9	43	M	98.5	X	X
A11	26	M	98.0	X	X
A13	44	M	99.1	X	X
A14	16	M	97.3	X	X
A15	28	M	96.2	X	X
A16	45	M	99.0	X	X
A18	52	F	98.5	X	-
A29	24	M	99.0	X	X
A30	56	M	98.4	-	X
A31	58	F	96.8	X	-
A33	32	M	99.0	X	X
A34	32	F	98.5	-	X
A36	46	F	97.0	X	X
A37	47	F	98.5	X	-
A38	17	F	98.0	X	X
A39	60	M	97.0	X	X
A40	57	F	98.5	X	-
A41	53	M	98.0	-	X
A42	53	M	97.5	-	X
A43	47	M	98.5	-	X
A44	51	M	97.0	-	X
A45	52	F	99.0	-	X

Table 3.2 Healthy Control Characteristics

HM450 (Human Methylation 450 BeadChip), HT-12 (HumanHT-12 Expression BeadChip discussed in Chapter 5).

Characteristic	Responders (n=30)	Non-responders (n=18)
Age, years		
At diagnosis		
Median	53	52.5
Range	21-74	20-80
Sex, (no, %)		
Male	19 (63.3)	12 (66.7)
Sokal risk group, (no, %)		
Low	13 (43.3)	3 (16.7)
Intermediate	7 (23.3)	8 (44.4)
High	10 (33.3)	7 (38.9)
BCR-ABL1 transcript type, %		
e13a2	10 (33.3)	5 (27.8)
e14a2	17 (56.7)	11 (61.1)
e13a2 and e14a2	3 (10.0)	2 (11.1)
BCR-ABL1/ABL1 ratio, %		
At diagnosis		
Median	68.9	72.0
Range	21.6 -115.0	28.3-120.7
Time from diagnosis to TKI therapy, months		
Median	0.36	0.42
Range	0-2.5	0-1.9
Therapy prior to TKI, (no, %)		
Hydroxycarbamide	14 (46.7)	10 (55.6)
Interferon	0	0
First line TKI therapy (no, %)		
Imatinib	15 (50.0)	9 (50.0)
Dasatinib	15 (50.0)	9 (50.0)

Table 3.3 Biological and Clinical Characteristics of SPIRIT 2 Patients (Validation Cohort)

3.4.2 HM450 quality control and preprocessing (normalisation)

To ensure all HM450 data generated were robust and reproducible, a series of quality control (QC) and preprocessing, normalisation steps were performed using minfi Bioconductor R package. Raw methylation data representing two different color channels were extracted as .idat files using Illumina BeadStudio software prior to any background subtraction or normalisation.

3.4.2.1 Quality Control

Approximately 800 probes are incorporated into the array design for the purpose of QC checks. Initial analysis of the data involved assessing sample staining, extension, hybridisation, target-removal and bisulfite conversion efficiency as well as checking negative control probes and those designed for between-array-normalisation.

Principal component analysis (PCA) was performed to detect possible batch effects since the samples were processed on multiple arrays at three genome centres at different time points during the project. To avoid position bias, samples were randomly allocated to array positions using a number generator. The Hammersmith hospital (discovery set) cohort (n=46) (Table 3.1) along with all healthy control samples (n=20) (Table 3.2) were processed in parallel on six arrays at the Barts Genome Centre. The SPIRIT 2 validation cohort samples (n=48) (Table 3.3) were hybridized at the Guy's Genomics Facility, King's College London six months later, while DNA from cell line experiments (Chapter 4) and adult bone marrow samples (n=15) were run at the UCL Genomics Facility, Institute of Child Health one year after this. The reasons for using different laboratories were mainly practical and based on the availability of technical expertise, anticipated processing times and associated costs.

Negative control probes consist of randomly permuted bisulfite-converted sequences that do not contain CpG dinucleotides and for this reason should not hybridise to DNA template. The mean intensity of these probes therefore produces an estimation of the system background.

A detection p-value was calculated for all probes based on the empirical cumulative distribution of these background level probes. Any containing a detection p-value <

0.01 were removed, essentially removing 3209 probes which displayed intensities close to that of background.

Accurate DNA methylation analysis is dependent on the prerequisite that bisulfite conversion of genomic DNA (gDNA) is complete following treatment. Unless all unmethylated cytosines are deaminated to uracil they will be incorrectly interpreted as methylated cytosines following downstream experiments. At the same time, bisulfite treatment dramatically alters the chemical and physical properties of DNA from a large double-stranded molecule to unstable single-stranded fragments rendering it prone to degradation. A number of steps were carried out to guarantee the quality of input gDNA including measuring samples using UV spectrophotometry and fluorometry (See Chapter 2, Materials and Methods). Secondly, the conversion kit used was manufacturer recommended with 'alternative' optimised incubation conditions specific for HM450. While the high thermocycler temperatures avoided ssDNA re-annealing to dsDNA during the conversion reaction, exposure times were kept to a minimum to limit the risk of sample degradation. Following treatment with bisulfite, samples were stored and transported at -20°C and always hybridised within 72 hours of conversion. Despite these measures, one sample CML-CPNR#17 was removed from further analyses at this stage due to a poor bisulfite conversion rate both quantitatively (91.1%) and relative to all other samples within the same batch (Figure 3.2). In addition a second sample, CML-CPNR#1 was removed due to concerns that an extended delay in starting TKI therapy could impact on disease biology and subsequent clinical outcome. This left 29 responders and 15 non-responders available for analyses.

3.4.2.2 Handling divergence of Type I and II probes using subset quantile normalization

As described in Chapter 2, Materials and Methods, the HM450 has two types of probes incorporated into the array design namely Type I and II. Type I probes measure methylated and unmethylated CpGs in the same colour channel and therefore require representation by two beads at the same location while Type II probes, measure methylation status in separate colour channels by means of just a single bead.

When plotted as a function of density (using a kernel density estimation with a Gaussian smoothing function and bandwidth = 0.05), the distribution of beta values for both probe types are bimodal showing two clear peaks corresponding to low/unmethylated and highly/fully methylated probes but with a shift in the distribution between probes types (Figure 3.3).

The divergence in distributions is probably driven by both biological and technical factors. However, the precise contribution of each of these factors on the observed differences is impossible to ascertain since type I and II probes never query the same CpGs. In terms of location, Type I probes are situated in CGIs containing CpGs that are less likely to be partially methylated rather fully methylated or unmethylated. Conversely, Type II probes have a wider genomic distribution where the GC content and methylation state will demonstrate greater variability. With respect to technical considerations, Type II probes demonstrate a lower dynamic range and a higher standard deviation in signal intensity due to the fact that the background signal is the product of two rather than a single colour (as is the case for type I). Finally, the probe design of Type I assume co-methylation of adjacent CpGs at query sites. In contrast, the Infinium II design uses a ‘degenerate’ R-base allowing adjacent CpGs to be individually assessed.

To account for these technical differences in probe type distribution, a subset quantile normalisation approach was used to normalise samples [296]. The intensity values of the red and green channels of type I and type II probes were handled separately equating to 6 separate groups i.e. Green Unmethylated Type I probes, Green methylated Type I probes, Red Unmethylated Type I probes, Red Methylated Type I probes, Green Methylated Type II probes and Red Unmethylated Type II probes. This

was in preference to other rescaling techniques that attempt to repair and merge the two probe types using peak-based correction methods.

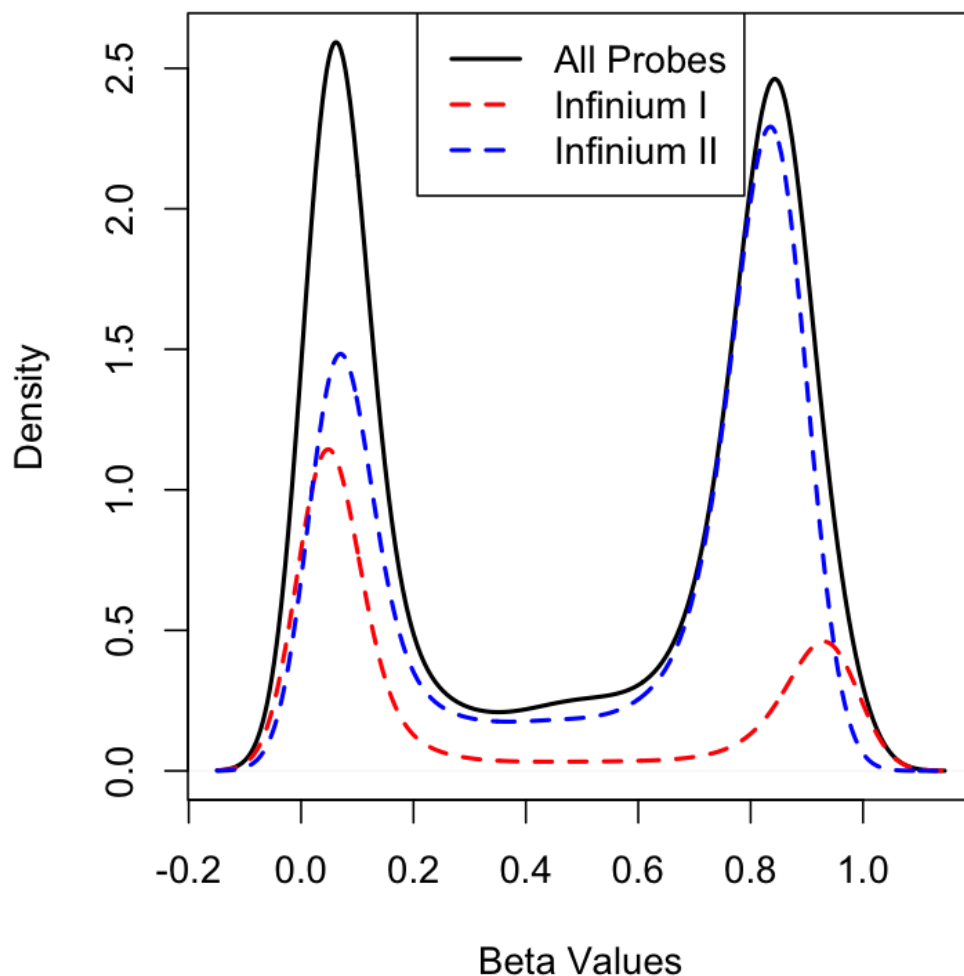


Figure 3.3 Density plots of the beta-values for the two Infinium assay types considered (red: Infinium I; blue: Infinium II; black: all probes) from Hammersmith cohort and health control samples.

Note that peak-based correction of Infinium I and II probes was avoided to prevent artificially reducing the power of detecting Type I probes or inflating the power in Type II probe detection. While this may lead to preferential selection of Type I probes (Type II probes are less effective detecting extreme beta values i.e. 0 and 1), this bias is minimal, restricted to small regions of the genome and is not anticipated to affect downstream analysis. Type I and II probes were analysed completely separately until the point of functional pathway analysis. (y-axis = CpG density; x-axis = beta value).

3.4.2.3 Data filtering

This step involved the removal of any probes that could interfere with methylation level variation as a result of genetic differences between samples. Following gender checks to support correct sample labeling (Figure 3.4), with the exception of cell line samples, 11648 probes mapping to X and Y chromosomes were removed from further analysis in all samples to ensure DMPs were not the result of different numbers of X and Y chromosomes in mixed sex experimental groups.

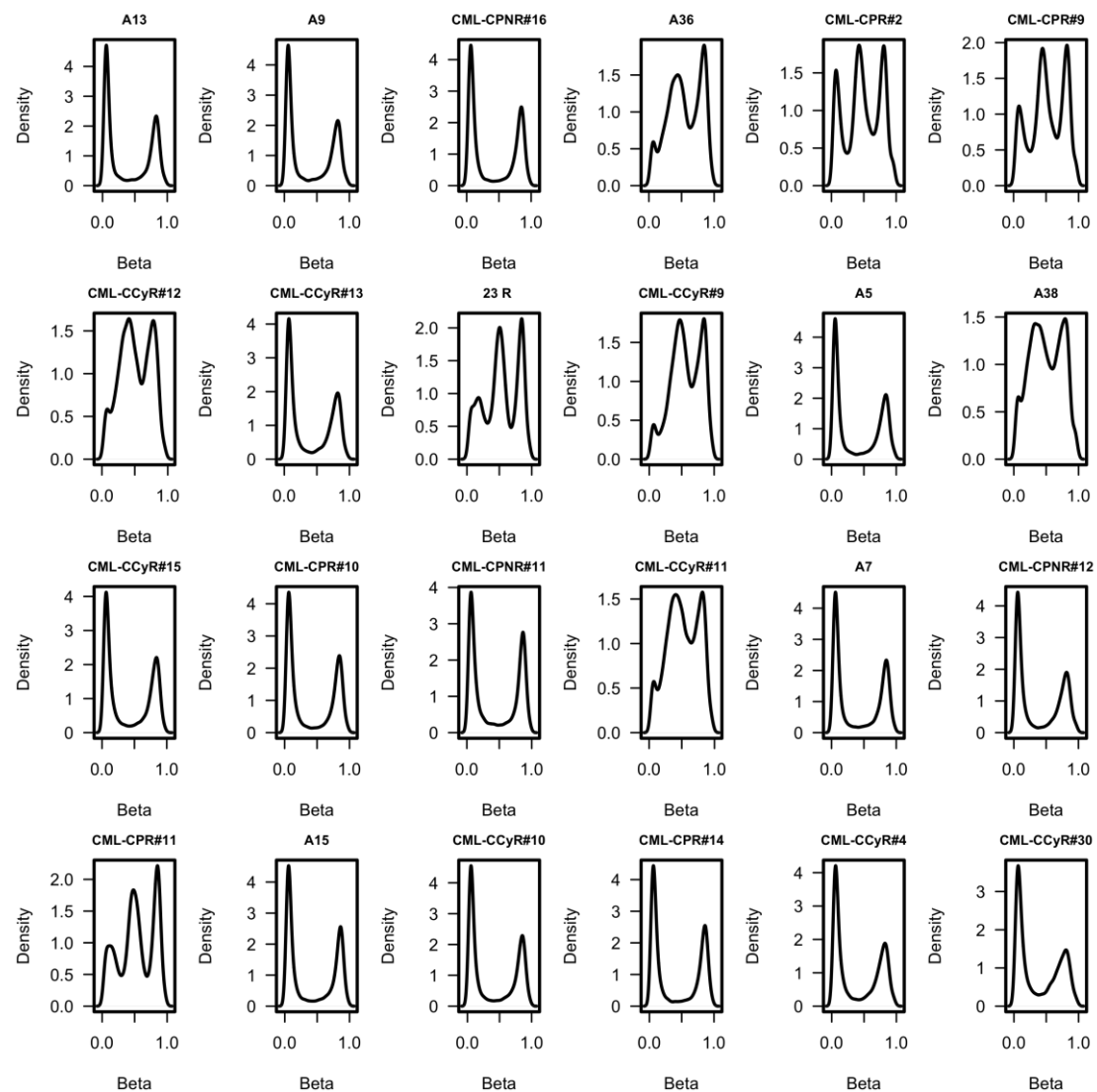


Figure 3.4 X-chromosome beta-value density plot in male and female study samples.

Example plots from the Hammersmith cohort and healthy control samples (same processing batch) were used to confirm correct labeling of arrays and rule out cross-contamination. Probes located on the X chromosome in males show the typical bimodal peaks with β -values < 0.2 (hypomethylated) and > 0.7 (hypermethylated) (e.g. samples A13 and A9, top row). In contrast, three peaks can be seen in female samples due to a high number of hemi-methylated probes ($0.4 < \beta < 0.6$) as a result of X inactivation with one allele fully methylated and the other completely unmethylated (e.g. 2R and 9R, top row). (y -axis = CpG density; x -axis = beta value).

A total of 40,592 probes mapping to more than one location were filtered out. Cross hybridisation is a recognised feature of bead arrays and certain probes are more susceptible to the process than others as a result of regions of sequence homology.

As previously mentioned described, *p*-values are generated for each probe on the HM450 that essentially describes the variance in intensity generated from the beads per probe. The higher the intensity variation the greater the signal noise so only probes with a *p*-value of < 0.01 were taken forward and included in the analysis.

Furthermore, single nucleotide polymorphisms (SNP) can influence methylation differences at CpG sites by affecting binding and subsequent probe intensities when compared to non-SNP associated CpGs [297]. According to Illumina annotation, 59,892 HM450 probes overlap with SNPs, some of which are known to affect methylation [298].

In this work, the effect of underlying genetic polymorphisms was assessed by identifying and cross-referencing those probes contained or extending into SNPs. Of 4,070 probes separating CML patients from healthy controls (see later in the chapter), 472 were in close proximity to annotated SNPs. When the same numbers of probes were repeatedly chosen at random from the array, a mean of 517.35 SNP-associated probes were selected. These observations suggest that CpGs separating the samples were less enriched for SNP-associated probes than expected (*p*-value, 0.015) and selection was therefore more likely to be driven by methylation differences rather than genetic ones. This lead to the conclusion that SNP-associated probes could be included in subsequent data analysis to maximise the power of the array in detecting methylation differences.

3.4.3 Identifying differentially methylated positions (DMPs)

Following QC and probe filtering 430,065 probes were available for analysis in the Hammersmith cohort and healthy control samples. Once normalised, the intensities of these probes were converted into beta-values. The beta-value for each probe is a methylation score ranging between 0 (unmethylated, U) to 1 (fully methylated, M) on a continuous scale and is calculated from the proportion of fluorescent signal intensity from the red and green channels of the M and U alleles using the standard formula

(1):

$$\beta = M / (M + U + 100) \quad (1)$$

The value 100 is included to compensate for the possibility of dividing by zero. As probe intensities are much larger than 100, the beta values are unaffected by this small addition.

All subsequent analyses including plots were performed using custom scripts in R. A differentially methylated position (DMP) was called when CpG methylation changes were > 0.2 between groups with a genome wide correct p value < 0.01 and the values were consistent within the groups ($SD < 0.1$).

3.4.4 Cluster analysis

The first step was to observe overall methylation pattern differences in CD34⁺ cells of CML patients and healthy controls. To achieve this, probes from the Hammersmith cohort including healthy controls were used as input for unsupervised hierarchical clustering using Wards's method in R [299]. This process constructs a binary tree by sequentially merging similar samples or probes based on a measure of similarity. The resulting clusters were driven by disease state rather than arrays or individual genetics (i.e. the diagnostic and remission samples from the same individual did not preferentially cluster), allowing the profiles to be further investigated.

Three distinct groups were identified: (i) CML samples collected at the time of diagnosis in the chronic phase of disease (CML-CP), (ii) samples collected at the time of complete cytogenetic remission in the same patients (CML-CCyR), and (iii) healthy controls (Figure 3.5).

These three groups formed two major clusters, one derived from CML samples including responders (CML-CP-R) and non-responders (CML-CP-NR) and the other from the controls and CML-CCyR samples. In the latter cluster CML-CCyR samples were distinct but epigenetically closer to normal CD34⁺ cells than CML-CP. The exception being one healthy control sample consistently clustering with the CML-CCyR samples. Interestingly this outlier sample belonged to the allogeneic sibling stem cell donor for a patient with CML raising the possibility of a genetic event that preceded or predisposed to the acquisition of the BCR-ABL1 fusion gene. This association could not be investigated any further in the context of this study.

All CML samples, regardless of disease stage exhibited a greater degree of inter-individual methylation differences than the unaffected healthy controls. This is consistent with the findings of other studies reporting considerable inter-individual methylation differences in cancer compared to the non-disease state [189, 300, 301].

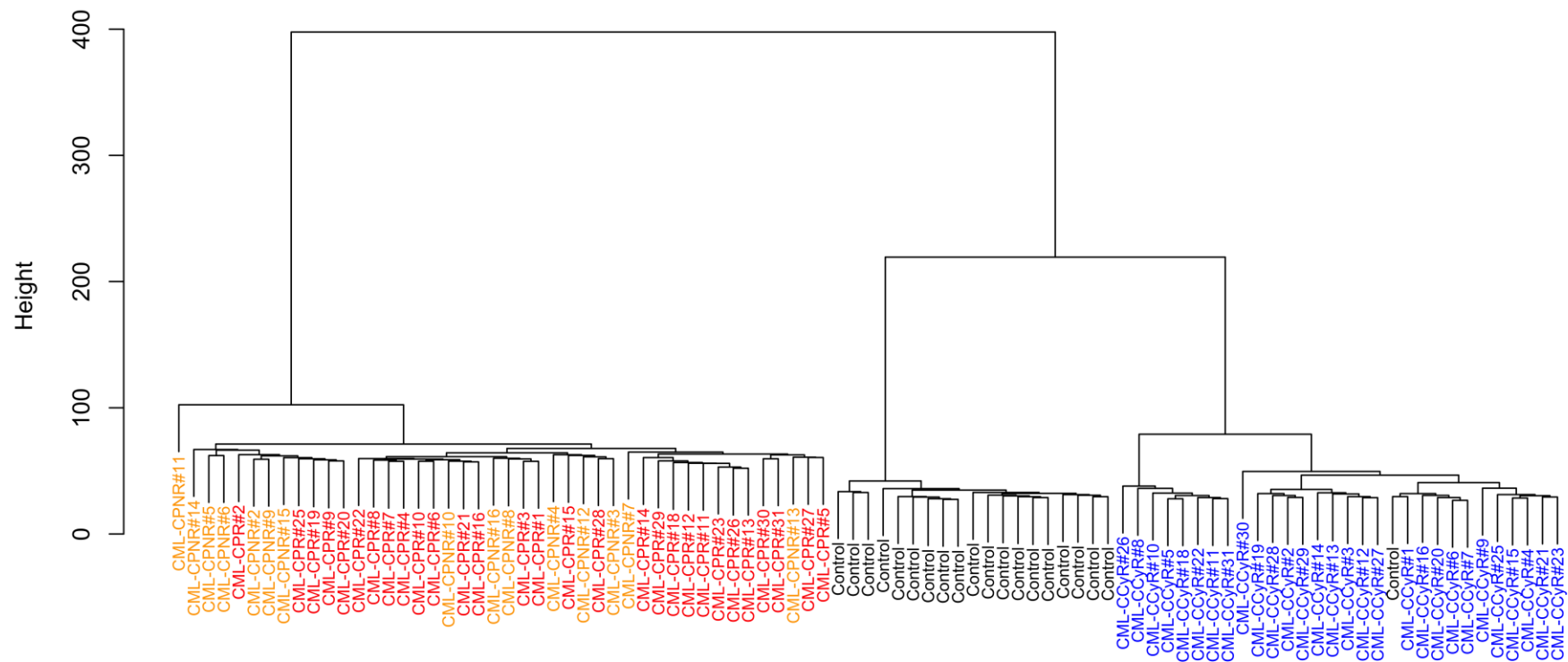


Figure 3.5 Dendrogram showing unsupervised hierarchical clustering of CD34⁺ cell methylation profiles in CML samples from the Hammersmith cohort and healthy controls following HM450 preprocessing.

Constructed using the Euclidean distance and complete linkage in a divisive ‘top-down’ approach. The vertical distance between the joining points of various branches are proportional to the degree of similarity between objects. The ‘leaves’ in the lower part of the dendrogram form 3 distinct groups. CML-CP patients [including responders (CML-CPR, red), non-responders (CML-CPNR, orange)], CML-CCyR (blue) and healthy controls (black).

3.4.5 The CML signature is significantly hypomethylated relative to healthy controls

3.4.5.1 Generating and defining a methylation signature in CML

Recently, a number of diseases including leukemia have been further defined on the basis of genome wide DNA methylation alterations [302]. To further investigate the DNA methylation signature of CML-CP the probe differences causing the clustering of CML-CP (CML-CP-R and CML-CP-NR combined) patient samples and healthy controls were called.

Using the eBayes and lmFit functions in limma, 47,152 probes separated these two groups (genome wide corrected p-value <0.01). To filter further and in order to ensure robust cancer related methylation differences, only those probes with average methylation differences greater than 20% were considered, reducing the number to 4,070 (Figure 3.6). Of these, 1,562 (38%) were hypermethylated in the CML-CP with respect to the healthy controls while 2,508 (62%) were hypomethylated, consistent with the widespread observation that cancer-associated methylation differences represent overall global loss of DNA methylation with foci of aberrant hypermethylation (Figure 3.7).

Genomic feature analysis revealed that CML-CP DMPs were significantly enriched in gene bodies (Permutation test; Fold Change = 1.32, p-value < 1e-4) and intergenic regions (Permutation Test: Fold Change = 1.19, p-value < 1e-4), with relative underrepresentation or depletion above background at promoter transcription start sites (TSS) (TSS200: Permutation Test; Fold Change = -2.03, p-value < 1e-4 and TSS1500: Permutation Test; Fold Change = -1.62, p-value < 1e-4) and associations with CpG islands (Permutation Test; Fold Change -0.8, p-value < 1e-4) (Figure 3.8).

Examining methylation state in a probe-wise manner is a necessary step in the data analysis pipeline. However, locus-by-locus assessment of CpGs in a particular health or disease state without consideration for the methylome as a whole is difficult to interpret and could simply represent heterogeneity in the underlying cell type or disease [303]. One approach to assessing the biological validity of a particular DMP ('the index') is to look to consecutive probes within a few thousand base pairs for

evidence of co-methylation as part of further reaching differentially methylated regions (DMRs) [304].

Where possible, the extent of co-methylation was investigated in the 4,070 probes that separated CML samples from healthy controls. Due to the lack of resolution of HM450, 3,635 (89%) of these were found to have neighbouring probes within a 2 kbp territory which could be analysed. A high proportion of these (84%) exhibited concordant directionality (Fold Enrichment; 1.34, p-value < 1e-4) providing high confidence for the validity of the original DMPs (Figure 3.9).

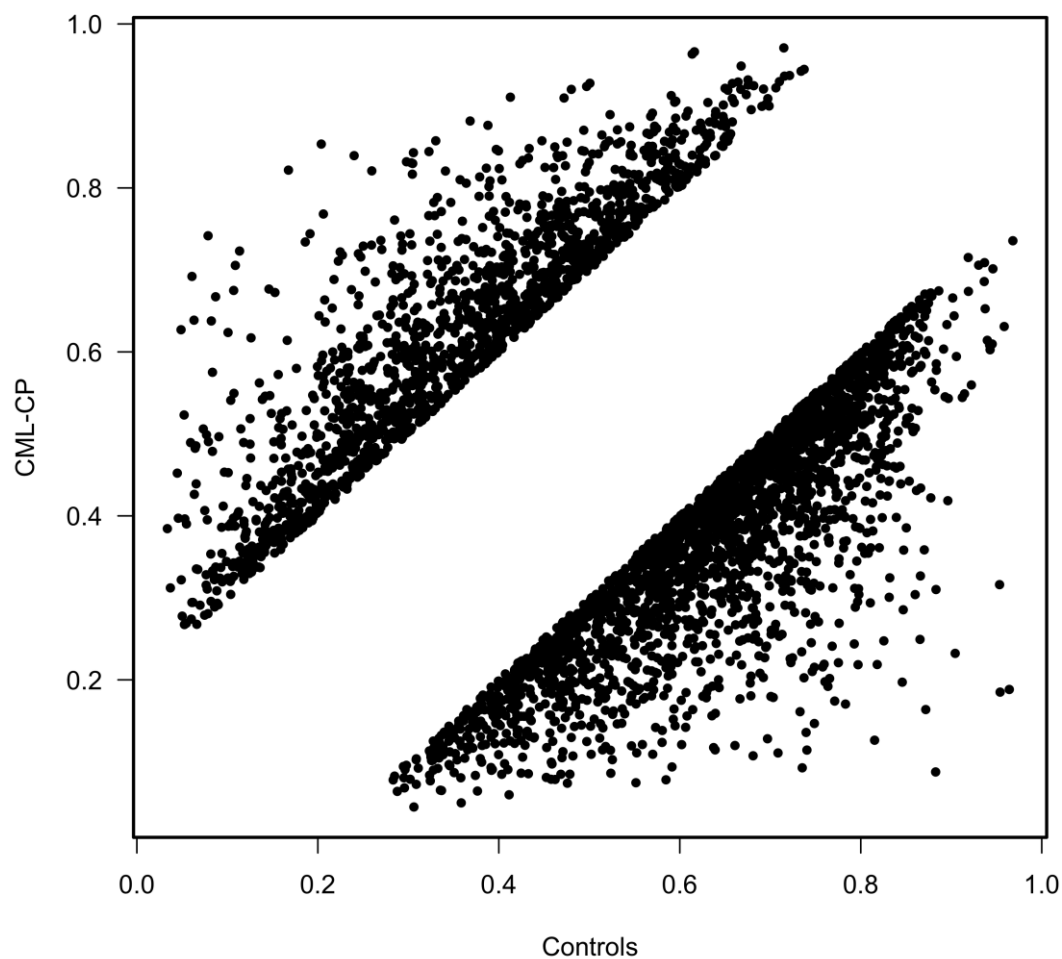


Figure 3.6 Scatter plot displaying CML-CP differentially methylated probes (DMPs) represented as beta values (measured on a 0 - 1 scale).

Limma linear regression model (genome-wide adjusted p-values of <0.01) with additional filter of beta value > 0.2 was performed on the full dataset following pre-processing. 4,070 DMPs separate CML-CP samples from healthy controls.

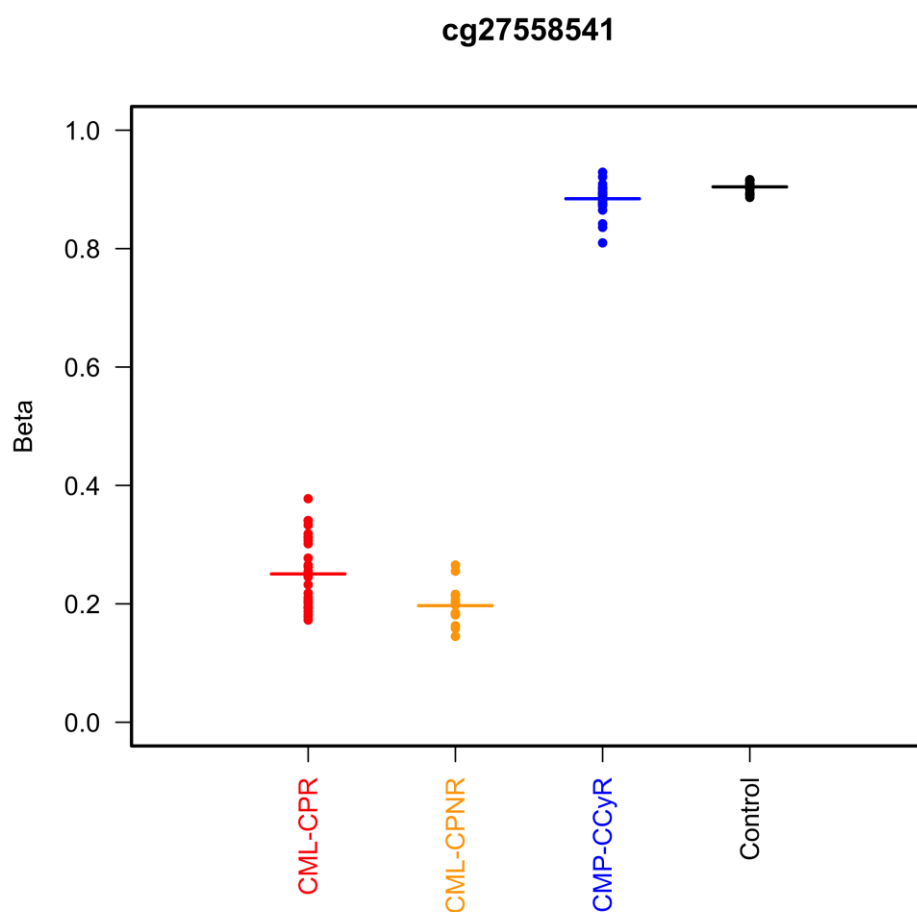


Figure 3.7 Box plot of an exemplar CML-CP differentially methylated probe (DMP)

Plotted are the beta values for all Hammersmith patients and control samples for probe ID cg27558541. The central line represents the group's median beta value. Marked hypomethylation is evident at the time of diagnosis in all CML-CP samples with respect to healthy controls. In this example, methylation restoration towards normal occurs at the time of disease remission in patients.

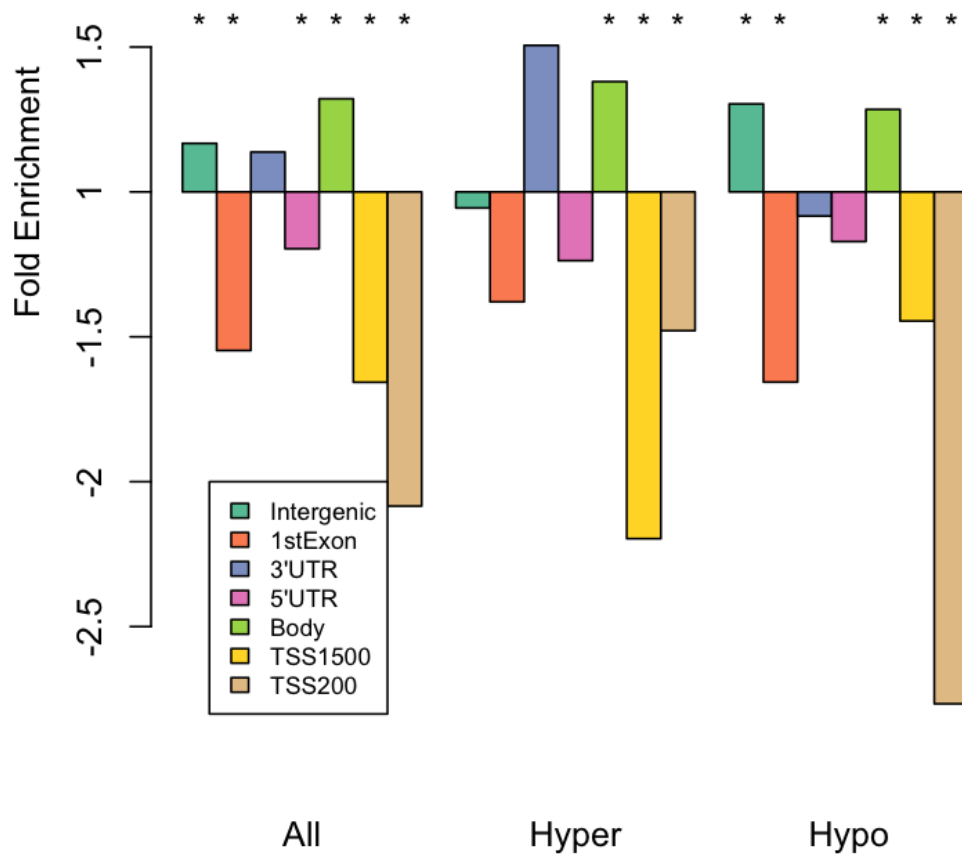


Figure 3.8 The relative genomic distribution of CML-CP DMPs (compared to healthy controls).

Genomic distribution data is shown for all DMPs (left), and separately for hyper- and hypomethylated DMPs (middle and right respectively). The y-axis shows fold change enrichment or depletion. DMPs were categorised according to their relative location to the closest gene [TSS200 (within 200 base pairs of a transcription start site), TSS1500 (within 1500 base pairs of a transcription start site, UTR (untranslated region). 5' side (leader sequence), 3' side, (trailer sequence)]. Asterix indicate significant bars.

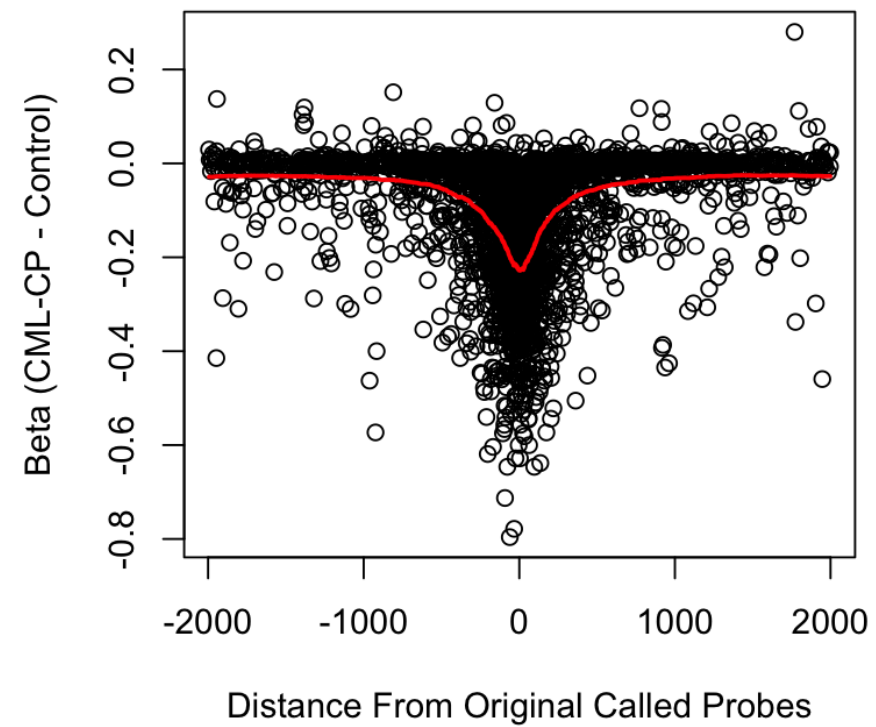
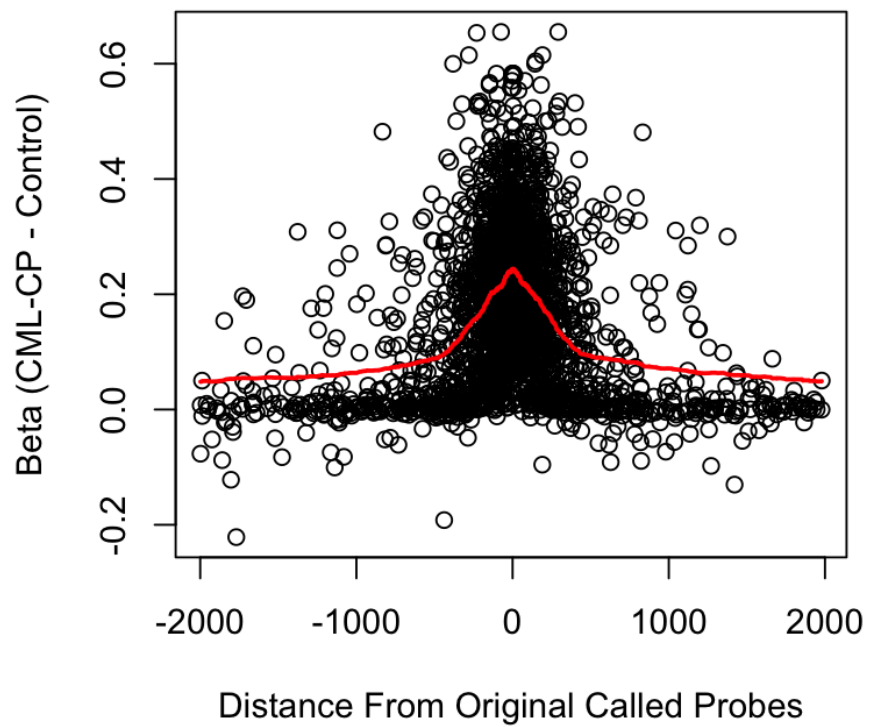


Figure 3.9 Identification of aberrant co-methylation in CpGs flanking CML DMPs

CML DMPs were separated into those that were hypermethylated (left) and hypomethylated (right) relative to healthy controls. The x -axis represents distance away from the called probe measured in base pairs (bp) and the y -axis the beta value difference between CML and healthy. The red line is the smoothed mean of the points. There is clear evidence of aberrant co-methylation, the strength and correlation of which deteriorates over distances $>1,000$ bp as expected.

3.4.5.2 Biological validation of the CML methylation signature in an independent patient cohort

Depending on the nature and goals of a particular study, validation of significant hits can be achieved through technical replication using an independent experimental approach, biologically using unrelated samples/data sets or both.

When HM450 was first made available, many groups used pyrosequencing and other targeted PCR-based methods to validate identified regions of interest from the array with robust accuracy. Since this time and with the increasing availability of next generation sequencing platforms, the HM450 has also demonstrated reproducibility in cross-validation comparisons against these methods [277].

This translational study focused on biological replication to validate the DMPs identified in the Hammersmith patients with the expectation that if the same CML-CP signature were robust in an independent cohort, the biological findings would provide important insights into DNA methylation reprogramming in CML. To this extent, the SPIRIT 2 study samples were used.

Following quality control and preprocessing using the same steps described above, all samples from SPIRIT 2 showed satisfactory bisulfite conversion (Figure 3.10). After removal of all sex chromosome-linked probes ($n=11,648$), those demonstrating cross-reactivity ($n=40,590$) and probes with detection p -value < 0.01 cut-off ($n=1193$), 432,081 probes were available for analysis (genome-wide adjusted p -values of < 0.01).

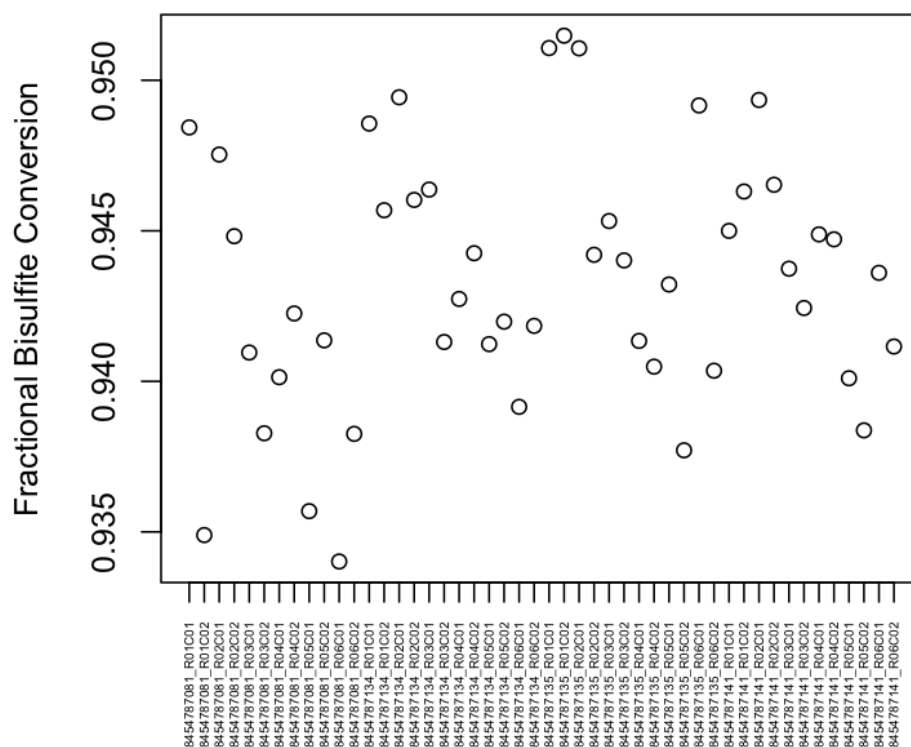


Figure 3.10 Bisulfite conversion efficiency of SPIRIT 2 validation cohort

The variability of conversion efficiency is greater in the SPIRIT 2 samples than the Hammersmith cohort but overall samples show $> 90\%$ conversion rate and were available for use in subsequent analyses (x -axis represents individual samples; y -axis bisulfite conversion efficiency).

Hierarchical clustering demonstrated that all CML samples (Hammersmith and SPIRIT 2) grouped together (Figure 3.11). However, with the exception of one sample an obvious batch effect was present, unrelated to the disease pathology and somehow associated with the HM450 BeadChips themselves or the experimental processing conditions.

The SPIRIT 2 samples were received from multiple hospital sites across the UK but processed (i.e. selection of MNCs and freezing) in the same clinical laboratory as the Hammersmith discovery cohort. However, in contrast to the Hammersmith samples, it was never the intention to collect and store SPIRIT 2 cells for therapeutic use. This may have impacted on the prioritisation and processing times resulting in ex-vivo cellular changes.

In addition many of the SPIRIT 2 samples were received via the postal system increasing the collection to processing time. This was particularly problematic for samples sent at the end of the week, which were not received until the start of the next. In comparison Hammersmith samples were collected in large volumes with same-day processing. In keeping with these concerns, it was noted at the time of thawing that cell viability was considerably lower in the SPIRIT 2 samples.

Finally, the two cohorts were processed at different genome facilities six months apart. The Hammersmith samples were hybridised using automated robotic systems at the Barts Genome Centre, while Kings Genomics used a manual protocol for the SPIRIT 2 samples. Any one or a combination of these factors might have contributed to the differences observed and explain the close but separate clustering of the samples seen in Figure 3.11.

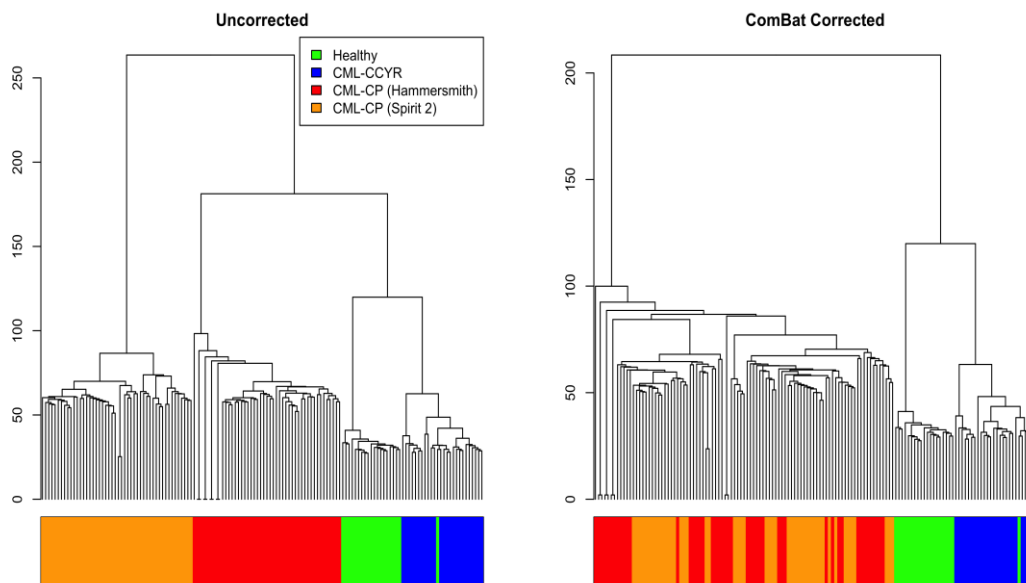


Figure 3.11 ComBat uncorrected and corrected dendrograms showing hierarchical clustering of all CML-CP samples from Hammersmith Hospital Patients (Discovery Cohort) and SPIRIT 2 Patients (Validation Cohort).

Surrogate variables are covariates constructed directly from high-dimensional data that can be used to adjust for subsequent unknown sources of noise artefact. The sva package in R includes functions for eliminating batch effects from high-throughput studies including methylome data sets. Batch effects between the Hammersmith and Spirit 2 cohorts were removed by the ComBat module using default parameters for the phenotypes: healthy, CML-CCYR and CML-CP.

To further investigate the extent of the batch effect between the two cohorts and estimate the degree of individual beta-value correlation, all beta-values from the two groups (normalised separately) were plotted against each other in an ‘across batch’ (between-group) analysis (Figure 3.12). The majority of beta-values were either heavily methylated (0.8-1.0) or unmethylated (0-0.20) with a satisfactory across batch correlation, as evidenced by a high coefficient of determination, R squared (R^2) = 0.97. The notable presence of extreme outlying probes between and within CML sample batches was not evident in the healthy control samples (Figure 3.13) and likely due to inter-individual cancer associated variation.

Finally, an across batch plot using only the 4,070 CML DMPs demonstrated a high degree of correlation ($R^2 = 0.98$, p-value < 2.1×10^{-16}) to allow meaningful analysis (Figure 3.14).

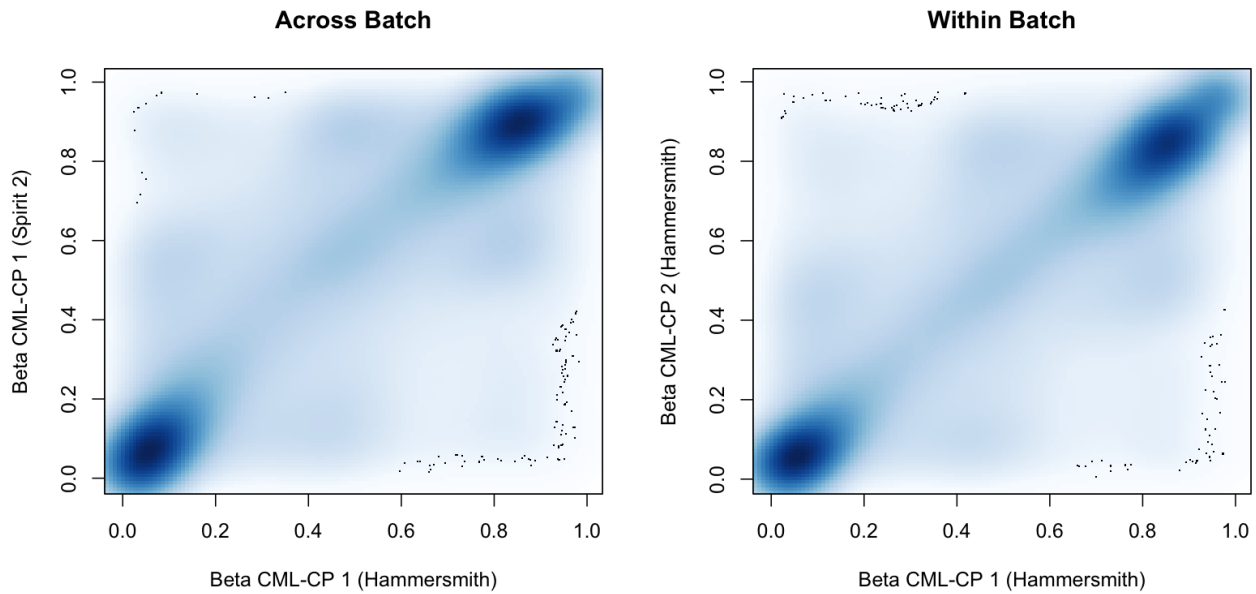


Figure 3.12 Across batch and within batch comparison of separately normalised beta-values.

Within any array-based experiment, variations between batches (‘batch effects’) are frequently observed, despite standardised experimental protocols. The beta-value correlation between the two batches of CML patients; Hammersmith and SPIRIT 2 are shown (left). For comparison, correlation between 2 individual patients from the Hammersmith (discovery) cohort, within batch is also displayed (right). The correlation coefficient, $R^2 = 0.97$. Further normalisation was avoided given the potential risks of data over-correction. Extreme outlying probes likely indicate DNA methylation deregulation occurring in leukaemia.

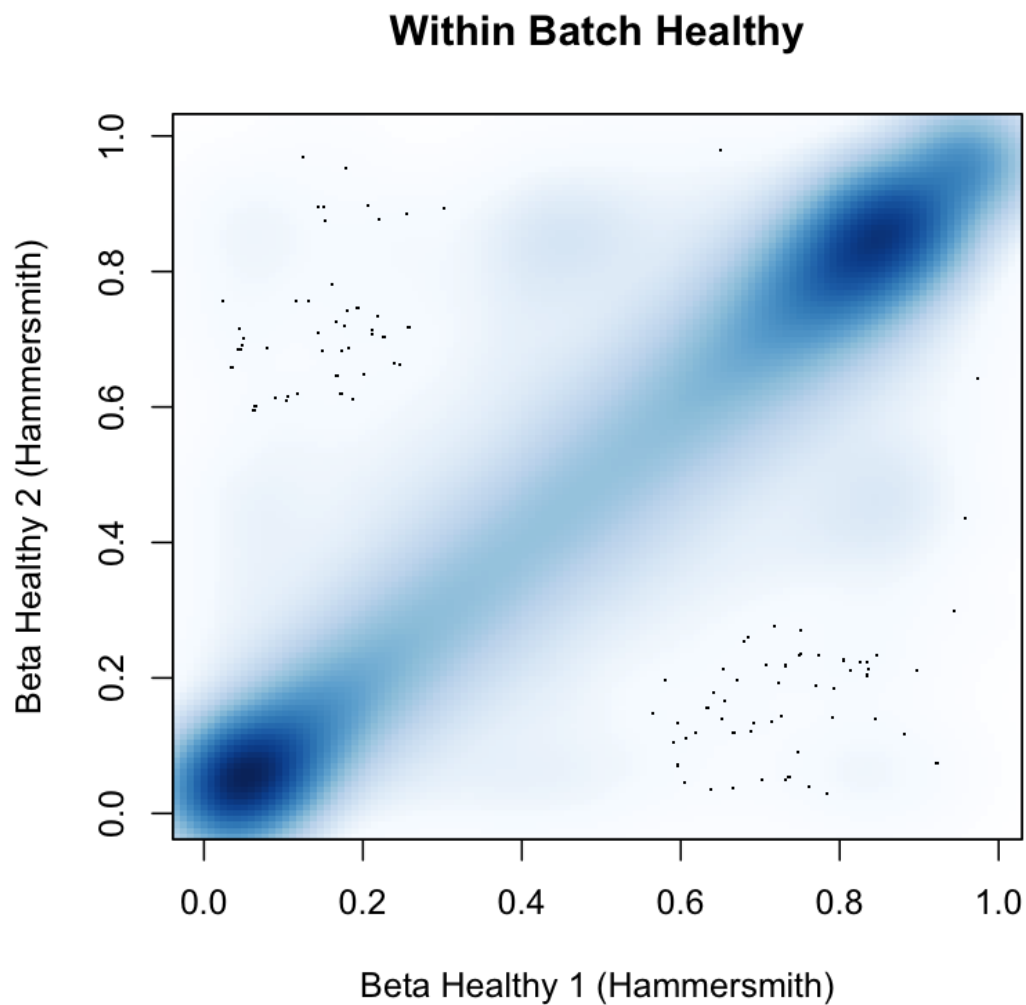


Figure 3.13 Scatter plot of beta-value correlation between healthy controls

Within batch beta-value correlation coefficient $R^2 = 0.99$. As expected in healthy controls there is less inter-individual variation with fewer extreme outliers compared to CML samples.

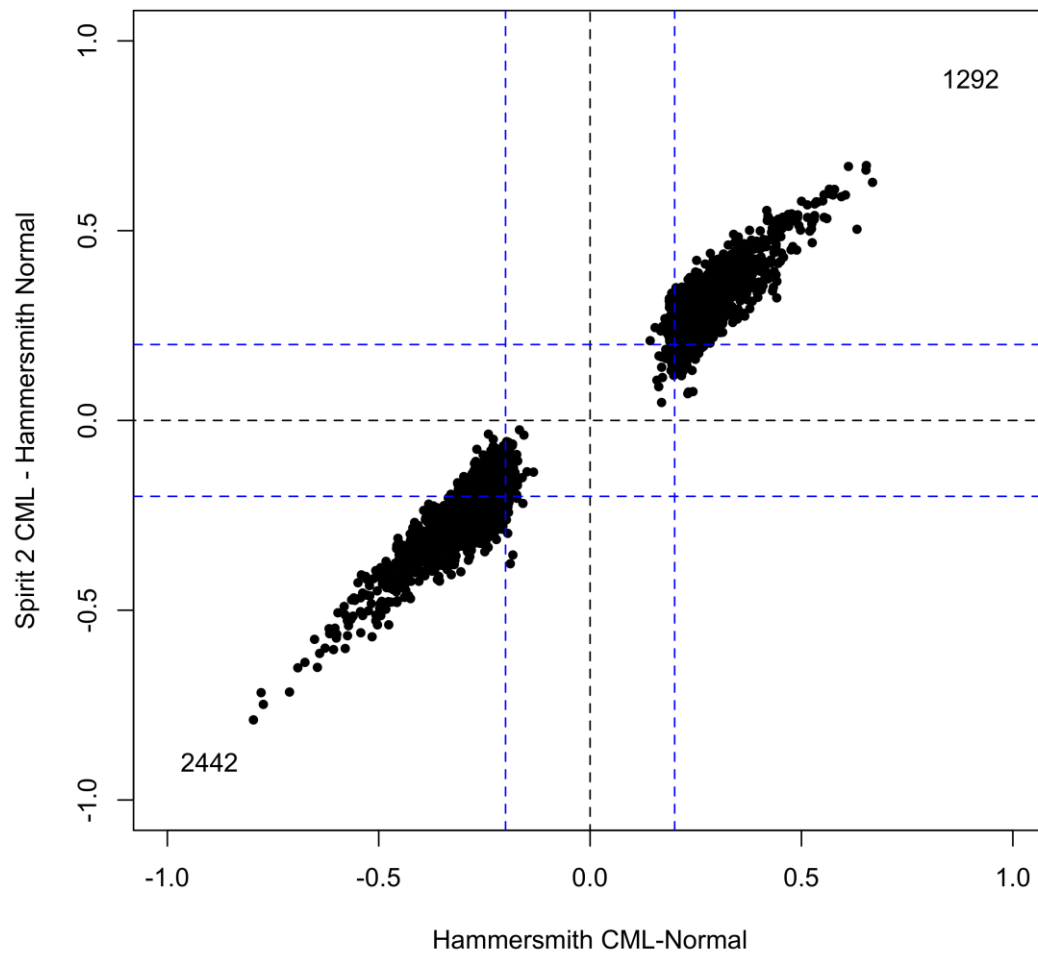


Figure 3.14 Across-batch validation of the 4,070 Hammersmith CML-CP DMPs

Plotted using ComBat uncorrected data to avoid ComBat over fitting by phenotype. Since differences are in agreement despite an obvious batch effect, this implies these differences are larger than the batch effect itself. No calling using the Spirit 2 differences were made as an undefined number of these will represent batch effects between Spirit 2 CML and Hammersmith CML-Normal. Overall 3,734 DMPs validate across the two cohorts.

[y-axis = DMP beta-value difference between CML discovery (Hammersmith) and validation (SPIRIT 2) cohorts; x-axis = significant DMPs separating CML patients from healthy controls].

3.4.5.3 Gene Ontology analysis of differentially methylated CpGs in CML-CP

To gain additional insights in to the genes corresponding with the CML-CP DMPs, those that could be mapped were functionally annotated using Gene Ontology (GO) in DAVID (Database for Annotation, Visualization and Integrated Discovery).

Using the complete HM450 probe set as background and the Benjamini-Hochberg method to correct for multiple hypothesis testing ($p < 0.01$), pathway analysis was performed on all DMPs and repeated for those that were hyper- or hypomethylated relative to healthy controls. All lists are available as supplementary files. The generated lists used for discussion consider all probes regardless on the basis that an altered pathway in CML could contain a series of genes that may be hyper- or hypomethylated in comparison to healthy controls yet still be part of the same pathological process.

GO analysis revealed 273 significant terms containing 10 or more genes showing enrichment for biological categories associated with general features of CML including cell adhesion (GO:0007155, p-value: $3.93\text{e-}21$), induction (GO:0006917, p-value: $3.63\text{e-}12$) and regulation of apoptosis (GO:0042981, p-value: $6.69\text{e-}18$) and programmed cell death (GO:0043067, p-value: $1.76\text{e-}17$) as well as leukocyte (GO:0045321, p-value: $4.01\text{e-}09$) and myeloid leukocyte activation (GO:0002274, p-value: 0.0065). Notably, there was significant overrepresentation of probes associated with positive (GO:0033674, p-value: $4.86\text{e-}06$) regulation (GO:0043549, p-value: $6.28\text{e-}06$) and activation (GO:0032147, p-value: $9.87\text{e-}04$) of protein tyrosine kinase activity.

Following the annotation of the genes by biological processes, the KEGG pathway database was used to explore the molecular interactions and networks involved. To reduce the overall probability of a Type-1 error (i.e. false positive discoveries) an FDR, Benjamini and Hochberg correction was applied using the stringent cutoff, $p < 0.01$. Only pathways containing at least 10 annotated genes were included to ensure meaningful analysis. This resulted in a total of 49 significant pathways including Pathways in cancer (hsa05200, fold enrichment 15.5X; p-value: $7.65\text{e-}43$), Jak-STAT (hsa04630, fold enrichment 19.4X; p-value: $5.92\text{e-}25$), Wnt (hsa04310, fold enrichment 15.3X; p-value: $5.68\text{e-}19$), TGF-beta (hsa04350, fold enrichment 14.1X; p-value: $9.15\text{e-}10$), p53 (hsa04115, fold enrichment 15.3X; p-value: $1.36\text{e-}08$)

signaling pathways as well as Chronic myeloid leukemia (hsa05220, fold enrichment 14X; p-value: 2.95e-08). Amongst the 11 genes represented in the Chronic myeloid leukemia pathway, ABL1 was included. ABL1 has been previously been shown to be aberrantly methylated in CML [234, 236].

These data reveal that specific DNA methylation pathway perturbations could play a role in dysregulating genes involved in CML leukaemogenesis. Whilst highly relevant and very encouraging, the results should be interpreted with a degree of caution for a number of reasons. Firstly the lack of significance in pathways not represented cannot be regarded as proof of insignificance. In the same way, a GO term that is highly significant can be generated and represented by a single gene, which is perhaps less revealing and informative than a number of genes in the same category attracting lower statistical significance. As a technical consideration, genes are not represented equally on HM450 in terms of probe coverage, which can complicate testing. Indeed, higher numbers of probes associated with a gene can artificially lower the significance of it, resulting in a larger *p*-value and biological bias.

Although GO and KEGG databases are useful tools for interpreting high throughput methylation data, it is important to consider that when compared with mRNA expression array data the approach is less direct and therefore more challenging to interpret. CpGs have different outcomes on gene expression depending on genomic location so additional analysis is required to further define the relevance and draw the correct conclusions.

3.4.6 Integrative analysis of CML-CP DMPs across different cancer types using Marmal-aid identifies commonalities between tumour entities

A number of studies have revealed that although different cancers display distinct methylation patterns, similarities or ‘pan-cancer signatures’ exist between tumour types in the wider context [305, 306]. Comparing the CML methylome to other cancers might provide some novel insights into the molecular mechanisms and epigenetic features of leukaemogenesis.

Having defined a series of CML-CP DMPs, further comparisons were made with external HM450 data using Marmal-aid [289]. Marmal-aid is a publicly available HM450 database combining information from a number of sources including The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) into a single location using a common annotation format and R package interface (available samples, n= 14,586, updated January 2015).

Sub-signatures were derived on two-dimensional cluster analyses from a variety of tumour types including solid and haematological malignancies as well as equivalent healthy tissues.

The data were first stratified into hyper- and hypo-DMPs i.e. CpGs that were more or less methylated in CML-CP relative to healthy controls. These groups were further sub-divided based on an agglomerative hierarchical cluster analysis to yield a total of six sub-clusters (Figure 3.15).

Cluster 5 contained a subset of 701 CpGs that were depleted for CpG islands (CGIs) and promoters but modestly enriched for probes located in gene-bodies. The average methylation beta-values were lower in the CML-CP compared with all other tissues and blood samples including different leukaemia types. If the DMPs defining CML are a composite of pan-cancer associated and BCR-ABL1-driven changes, this sub-cluster appeared to represent CML-specific DMPs.

The proposed biological functions of the genes associated with these CpGs were investigated using KEGG (considering pathways with Benjamini-Hochberg corrected, $P < 0.01$, >10X enrichment) resulting in 39 significant pathways. Of these, 24 were

unique i.e. pathways not significant in other clusters and included chronic myeloid leukemia (hsa05220, fold enrichment; 15.8X; $p= 0.0078$), Jak-STAT (hsa04630, fold enrichment 22X; $p= 1.98e-08$) and TGF-beta (hsa04350, fold enrichment 16.8X; $p= 0.0013$) signaling. This is supportive of the proposition that the probes in cluster 5 reflect CML-unique changes and in addition included pathways that have frequently been described in contributing to the disease pathogenesis.

Next, cluster 1 contained a subset of 608 CpGs that displayed higher methylation levels across tissues and the corresponding solid tumor types than healthy blood samples. These probes also demonstrated hypermethylation in CD34⁺ CML-CP cells and untreated AML peripheral blood mononuclear cells (MNCs) but not acute lymphoblastic leukemia (ALL) MNCs, giving rise to the possibility that this cluster represents a myeloid leukemia-specific sub-signature.

Gene functions of the 23 pathways in this cluster organised using KEGG annotation included chemokine signaling (hsa04062, fold enrichment 19.6X; $p= 3.13e-05$), cytokine-cytokine receptor interaction (hsa04060, fold enrichment 13X, $p= 9.90e-04$) and leukocyte transendothelial migration (hsa04670, fold enrichment 22.8X; $p= 7.46e-04$) in which the chemokine receptor CXCR4 and its ligand CXCL12 feature repeatedly.

The CpGs contained within cluster 2 were relatively hypomethylated in healthy tissue and normal blood cells particularly with respect to leukemia, but also cancers from different tissues. This overall gain of CpG methylation suggests that cluster 2 might characterise changes common to multiple tumour types and represent a pan-cancer methylation signature across samples.

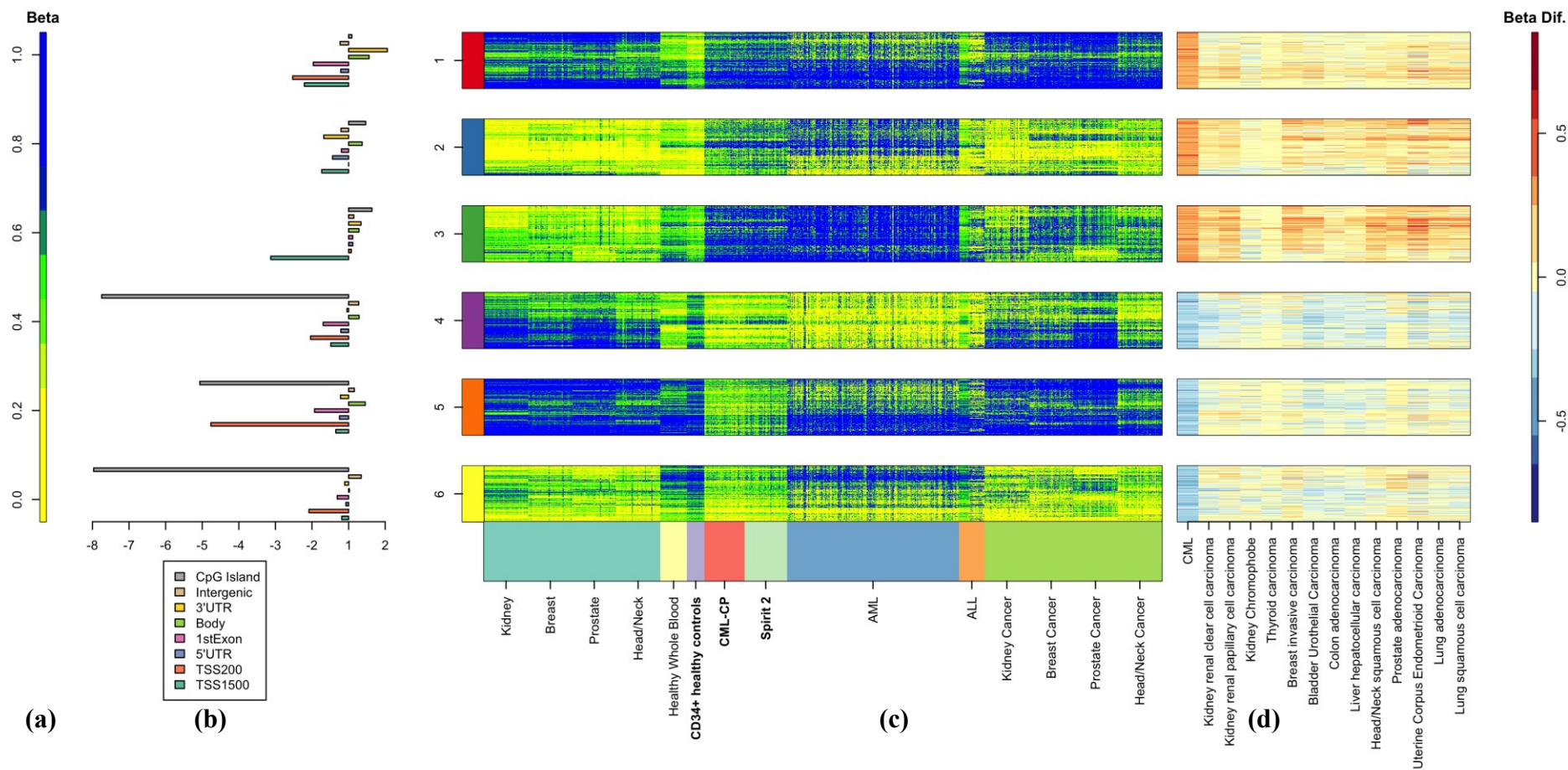


Figure 3.15 Marmal-aid heat map of the 4,070 CML-CP DMPs (vertical axis) separated into 6 clusters (horizontal axis) based on unsupervised hierarchical clustering.

This plot highlights the methylation similarities and differences between CML-CP DMPs in the context of other healthy tissues and cancers including blood and leukemia. From left to right: **a)** the absolute beta-values using color scale 1 (blue) = fully methylated, 0.5 (green) = hemimethylated, 0 (yellow) = unmethylated. **b)** The relative genomic distributions of DMP clusters with fold enrichment or depletion **c)** probes clustered based on their methylation values across all samples and divided into six distinct clusters. Vertical points represent individual probe beta values of Hammersmith CML-CP samples (pink) and SPIRIT 2 samples (light green) in relation to CD34⁺ healthy controls (lilac), whole blood (yellow) n= 35, healthy tissues (turquoise) n= 215, as well as leukemias [AML (blue) n= 196, ALL (orange) n= 31] and solid cancers (green) n= 211 **d)** the directionality of methylation changes in CML-CP compared to other solid cancers. In clusters 1-3, CML-DMPs show aberrant gains in methylation across all solid malignancies while conversely in clusters 4-6 there is relative loss.

3.4.7 A DNA Methylation classifier does not predict response to TKI therapy in pre-treatment CML samples

There exists notable inter-individual heterogeneity in responses to imatinib therapy for CML patients. With the lack of available molecular tests to determine outcome, one main objective of this work was to establish whether a DNA methylation signature at diagnosis could accurately predict for imatinib response. Identifying high-risk patients early in their care pathway would allow management to be tailored towards more intensive strategies from the outset such as the use of second and third generation TKIs or allogeneic stem cell transplantation, in suitable cases.

3.4.7.1 Predicting TKI response using Random Forest modeling

Using Random Forest, an ensemble classifier based on a decision tree algorithm, a model was grown to separate responders from non-responders by DNA methylation profile, age, sex and Sokal score.

Supervised learning is a form of machine learning where objects in a given collection are classified using a set of attributes or features. The classes are known in advance (i.e. responders and non-responders) and the goal is to build a model that can assign patients to the class based on features (i.e. DMPs and important clinical parameters) then accurately predict the class membership of new individuals using these.

To this extent, the Hammersmith cohort ‘training set’ was used to inform and stabilise the classifier while as before, the SPIRIT 2 samples were the ‘test or validation set’ to assess the degree of error.

With a permutation test it was possible to select 756 probes (759 features including age, sex and Sokal score) that separated the two groups ($p < 0.01$) (Figure 3.16). Using these, a ‘leave-one-out’ cross-validation procedure was repeated 1000 times for each sample and used as a minimal validation set to estimate the prediction accuracy. Of the 17 non-responders, 2 patients (NR13 and NR16) were incorrectly identified by the model as having responder methylation phenotypes 100% of the time while a further 3 samples (NR2, NR11 and NR7) were incorrectly predicted as responders 10%, 13% and 27% of the time respectively (Table 3.4).

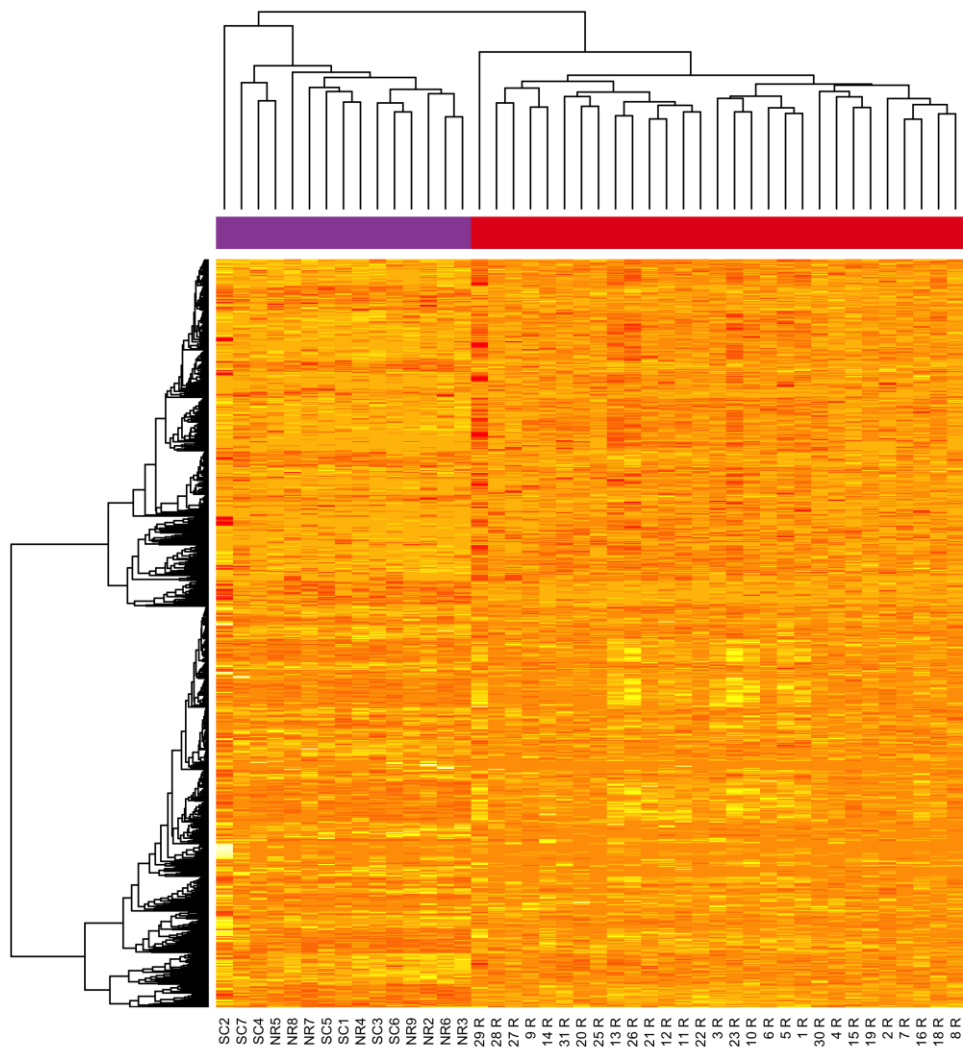


Figure 3.16 Heat map representation (scale by row). 756 probes (759 features including age, sex and Sokal score) probe DNA methylation classifier Responder vs. Non-responder using Random Forest machine learning

Individual patients are represented in columns (CML-CP-NR purple, CML-CP-R red) and methylation in rows. Darker shade represents increasing degree of methylation (i.e. yellow < orange < red). Sample 16R was excluded from the classifier since no Sokal score was available.

	Predicted Non-Responder	Predicted Responder
Non-Responder	12.502	2.498
Responder	0	28

Table 3.4 Confusion matrix (contingency table) for leave-one-out cross-validation procedure (repeated 1000 times per sample) on training set (Hammersmith Hospital).

Note sample 16R was excluded from the classifier, as no Sokal score was available.

It was also possible to show that amongst the 756 probes, co-methylation was increased compared to the background set and 69% of those located in CpG islands were hypermethylated in non-responders.

However, although the model looked encouraging, when applied to the validation set, the classifier could not reliably separate patients based on their TKI response (Table 3.5).

	Predicted Non-Responder	Predicted Responder
Non-Responder	17	1
Responder	29	1

Table 3.5 Contingency table showing the 759-feature classifier performance for prediction of response to TKI therapy, in the SPIRIT 2 validation set.

There are a number of reasons why a signature of response was not found including insufficient sample sizes to underrepresented HM450 probe coverage of the relevant genomic regions predictive of response. It is also conceivable that the DNA methylation changes at diagnosis are so homogeneous among patients that a consistent response sub-signature is not present. After all, imatinib is a molecule designed to target a specific genetic, rather than epigenetic aberration.

One further consideration relates to the administration of hydroxycarbamide, a deoxyribonucleotide inhibitor frequently used as a cytoreductive agent prior to TKI therapy in patients presenting with leukocytosis. Since there is some suggestion that the use of this drug can induce DNA methylation alterations [307] every effort was made to collect diagnostic samples before any doses were given. Although a proportion of patients from both cohorts received hydroxycarbamide, in the Hammersmith group all doses were taken following clinical sample collection. Conversely, nine patients from the SPIRIT 2 cohort (4 CML-CP-R and 5 CML-CP-NR) appear to have received hydroxycarbamide for less than 5 days preceding sample collection, giving rise to the possibility that this may have induced methylation changes. Although there was no evidence to support this concern on hierarchical clustering analyses (i.e. these patients did not cluster together separately from others), this highlights one of the challenges associated with clinical studies involving multicenter participation.

3.4.7.2 Relabeling patients according to ELN updated recommendations

In 2012, Marin *et al* at the Hammersmith hospital were able to demonstrate that patients with more than 10% detectable BCR-ABL1 transcripts at 3 months after starting imatinib therapy had significantly lower event-free and overall survival probabilities and were destined to fare poorly [49]. The need for early clinical intervention in this patient category was subsequently confirmed by other groups and incorporated into the 2013 updated European LeukemiaNet (ELN) CML management recommendations [48].

At the time this project started, attainment of CCyR one year from starting TKI treatment was considered an optimal therapeutic response. Yet, in light of these new study findings the ELN have since reclassified this level of residual disease as ‘sub-

optimal' and patients in this category attract a warning label for close clinical observation. Inevitably, these revisions impact on the labeling of drug response in this study.

In a final attempt to define a response signature, all CML-CP patients were re-categorised according to 2013 ELN recommendations. Any patient failing to demonstrate BCR-ABL1 transcripts $< 10\%$ three months after starting imatinib were classified as non-responders. Patients lacking 3-month transcript values were omitted if the 6-month levels were also unavailable. Similarly, cases involving clinical uncertainties e.g. excessive delays in commencing therapy secondary to healthcare funding/drug access issues were removed due to concerns that these could impact on disease evolution and confound data interpretation. After all, a patient experiencing treatment delays and failing to achieve optimal transcript levels at a specified time point may have demonstrated an improved response had therapy commenced sooner.

The clinical case notes of the Hammersmith cohort were reexamined and discussed in the department of Haematology to ensure a consensus was reached as to whether borderline cases were classified as responders, non-responders or should be excluded by the new response criteria. Correspondingly, the trial database from SPIRIT 2 was used as a suitable alternative to the medical notes, which were unavailable due to the multi-site recruitment nature of the study. As described previously, the patients in this cohort were randomised to receive either imatinib or dasatinib as first-line therapy in contrast to Hammersmith patients, all of who were prescribed imatinib. As a second-generation TKI dasatinib has demonstrated faster and deeper responses in the upfront setting and can benefit patients with CML-CP who are first-line imatinib-resistant or intolerant [308-310].

Patients randomised to receive dasatinib were originally included in this study to increase access to TKI non-responder samples and find out whether a potential response signature was agent- or drug class-specific. With no obvious response signature, all dasatinib responders were removed from the final analyses since it was conceivable a patient failing to respond to imatinib could respond to dasatinib thereby complicating an imatinib (agent-specific) classifier. Dasatinib non-responders were included however, on the basis that in clinical practice a person who fails first-line dasatinib treatment would be commenced on an alternative second or third generation

TKI rather than imatinib i.e. it is likely that the same patient would have failed upfront imatinib had it been given as an alternative.

For the Hammersmith cohort, in addition to 16R (no Sokal available), NR1 (treatment delay) and NR17 (incomplete bisulfite conversion), a further 5 patients (3 CML-CP-R and 2 CML-CP-NR) were excluded using the strict criteria outlined above leaving 38 patients available for analysis. Twenty-one of the original 29 CML-CP-R patients retained the same label, while 4 were re-categorised as CML-CP-NR having failed to achieve BCR-ABL1 transcript ratios $< 10\%$ at 3 months on the IS. Thirteen of 17 (15 excluding NR1 and NR17 previously removed) CML-CP-NR patients remained unchanged. Overall, 21 CML-CP-R and 17 CML-CP-NR made up the newly labeled discovery set.

For the SPIRIT 2 cohort, all 15 dasatinib CML-CP-R patients were omitted in addition to 1 patient previously labeled as a dasatinib non-responder. This patient did achieve BCR-ABL1 transcript ratios $< 10\%$ at 3 months before failing therapy. Thirty-two remaining patients were available for analysis consisting of 16 imatinib responders (15 CML-CP-R patients receiving imatinib along with one patient formerly labeled as CML-CP-NR but once again had achieved transcript ratios $< 10\%$ at 3 months prior to failing imatinib therapy) and 16 TKI non-responders (imatinib $n=8$, dasatinib $n=8$).

The revised Hammersmith CML-CP-R and CML-CP-NR subjects along with age, Sokal score and sex, were used to train and stabilise a new random forest classifier. A 'leave-one-out' cross-validation procedure repeated 1000 times for each sample resulted in 578 probes (581 features including Sokal score, age and sex) ($p<0.01$) separating the groups (Table 3.6), which were then applied to the Spirit 2 validation set. Despite the above measures including re-categorization as well as incorporation of important clinical parameters, no reliable separation of patients was possible based on response to TKIs (Table 3.7).

	Predicted Non-Responder	Predicted Responder
Non-Responder	16.716	0.284
Responder	0.965	20.035

Table 3.6 Confusion matrix (contingency table) for leave-one-out cross-validation procedure (repeated 1000 times per sample) on training set (Hammersmith Hospital) using revised ELN criteria for response

581-feature classifier based on 578 probes, Sokal score, age and sex

	Predicted Non-Responder	Predicted Responder
Non-Responder	1	15
Responder (imatinib)	5	11

Table 3.7 Contingency table showing the 581-feature classifier performance for prediction of response to TKI therapy, in the SPIRIT 2 (validation) cohort using revised ELN criteria for response

(Fisher's test $p=0.1719$)

3.4.8 Despite restoration towards normal, residual DNA methylation differences exist at the time of remission in G-CSF mobilised samples

The work so far has identified a core set of probes that distinguished CML-CP from healthy controls. Because DNA methylation is a reversible epigenetic modification, the next important step was to elucidate the epigenome-wide differences between paired CML-CP and CML-CCyR samples as well as CML-CCyR and healthy controls seen on hierarchical clustering (Figure 3.5).

3.4.8.1 Flow cytometric analysis of CD34⁺ haematopoietic stem and progenitor cells

The CD34⁺ compartment consists of a number of different cell types including primitive haematopoietic stem cells (HSC) as well as more committed progenitors (HPC). At the time of CML-CCyR, haematopoiesis may not be the same as healthy bone marrow. Fewer multipotent HSCs (stem cell exhaustion) combined with the ongoing effects of imatinib could lead to differences in the proportions as well as functioning of the progenitor cells. If large enough, these might be reflected as alterations in the global gene expression and subsequently DNA methylation patterns of cells.

While the functional capabilities of CD34⁺ cells in CML are outside the scope of this work, as a first step and to ensure that the DNA methylation changes associated with the clusters, particularly the small differences between CML-CCyR and healthy controls, were not solely attributable to variations in the cellular composition of the CD34⁺ compartment i.e. the proportions of haematopoietic stem and progenitor cells, comparisons were made between groups [CML-CP-NR/R (n=6), CML-CCyR (n=6) and healthy controls (n=7)].

The primitive stem (Lin⁻CD34⁺CD38⁻) and progenitor (lin⁻CD34⁺CD38⁺) haematopoietic cell fractions including common myeloid progenitors (CMPs), granulocyte macrophage progenitors (GMPs) and megakaryocyte erythroid progenitors (MEPs) were determined by multiparameter flow cytometric quantification (Figure 3.17 to Figure 3.20). Data acquisition was performed on BD FACSCanto II and analysed with FlowJo software using methodologies and gating strategies described previously in Chapter 2 Materials and Methods.

Overall the trend was that CML-CP patients had higher proportions of primitive HSCs as well as GMPs but lower numbers of CMPs and MEPs with respect to CML-CCyR and healthy control samples (Figure 3.21). This group also displayed marked inter-individual variability within subsets particularly CMP (range 16-61%) and GMP (range 3-55%) populations which is not surprising given that normal haematopoietic homeostasis is profoundly disrupted during leukaemogenesis. CML-CCyR patients have similar profiles to non-leukaemic controls which mean that importantly, the close but distinct clustering is unlikely to be cell composition-based. No described differences were greater than 2-fold between any groups or statistically significant using the non-parametric Mann-Whitney U (Wilcoxon rank) test ($p < 0.05$). Altogether, these observations suggest that differences observed in methylation patterns between experimental groups are not the result of cell type composition within the CD34⁺ compartment.

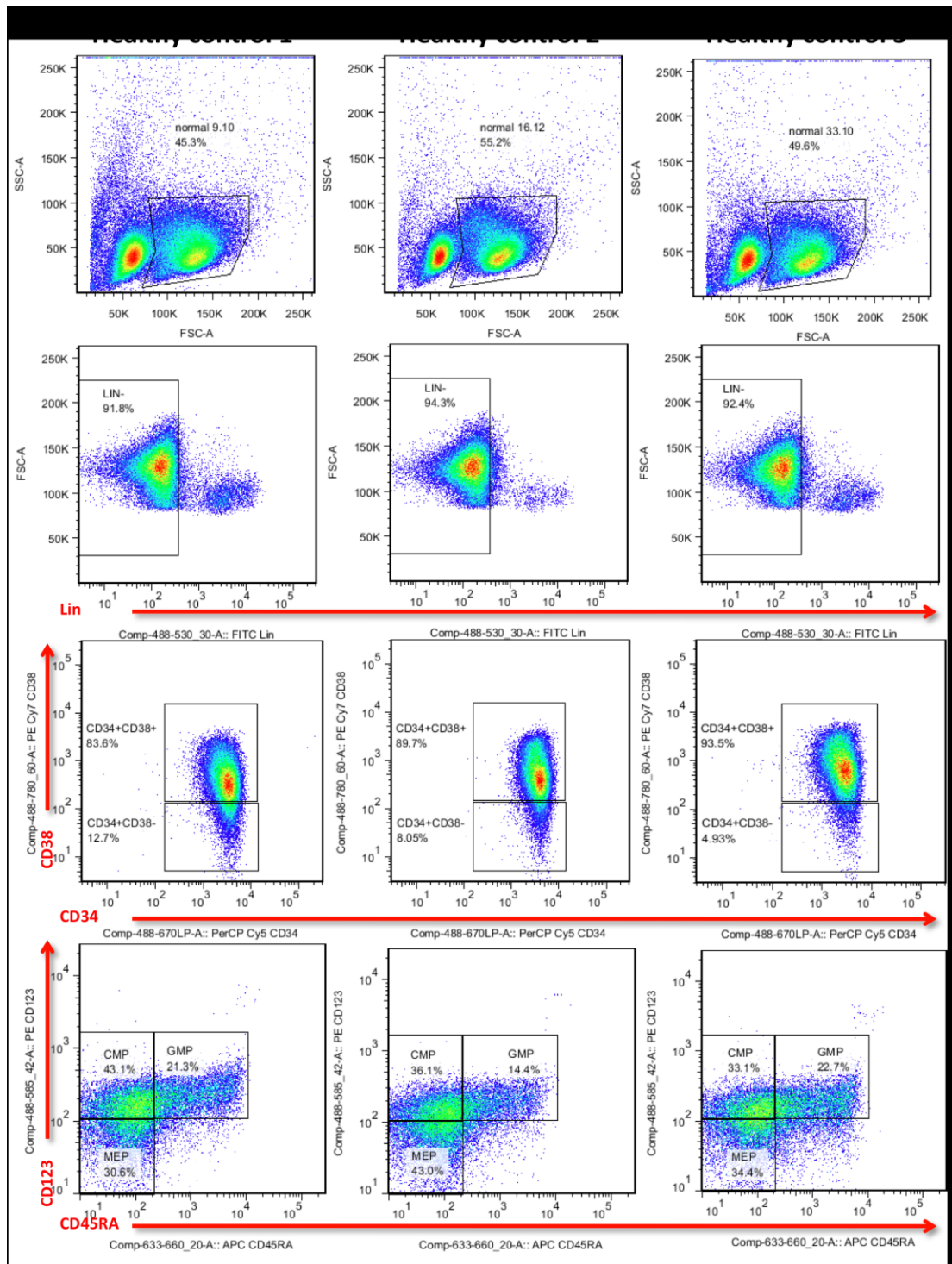


Figure 3.17 FACS plots of CD34⁺ haematopoietic stem (HSC) and progenitor cell (HPC) populations from 3 healthy subjects following G-CSF mobilisation

Top row shows light scatter plots (not shown graphically in Figure 3.18 to Figure 3.20) below. Polygonal regions define percentage count of cell subsets.

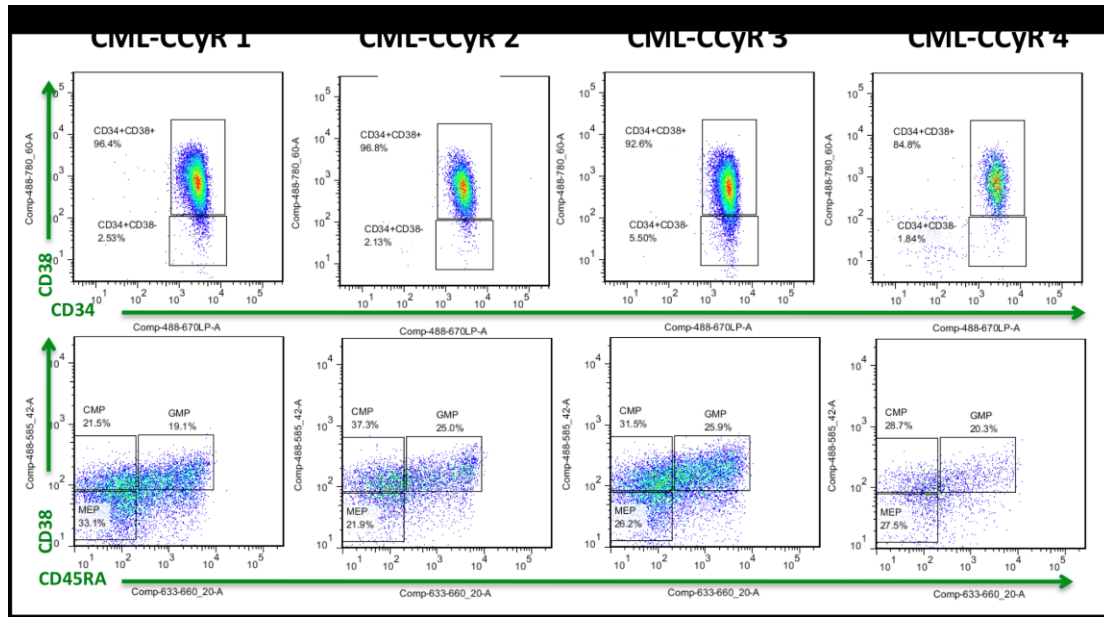


Figure 3.18 FACS plots of CD34⁺ haematopoietic stem (HSC) and progenitor cell (HPC) populations from 4 CML-CCyR patients on imatinib therapy following G-CSF mobilization.

CD34⁺ enriched cells were stained with anti-Lin-FITC, CD34-PerCPCy5.5, CD38-PECy7, CD45RA-APC and CD123-PE. A polygonal region was first drawn around the Lin⁻ cells on the scatter plot. This region was used to set gates on a plot of CD34 versus CD38 fluorescence (top row). A further region was set on the CD34⁺, CD38⁺ cells and used to gate for common myeloid progenitors (CMP) (Lin⁻CD34⁺CD38⁺CD123^{MED}CD45RA⁻), granulocyte macrophage progenitors (GMP) (Lin⁻CD34⁺CD38⁺CD123^{MED}CD45RA⁺) and megakaryocyte-erythrocyte progenitors (MEP) (Lin⁻CD34⁺CD38⁺CD123⁻CD45RA⁻).

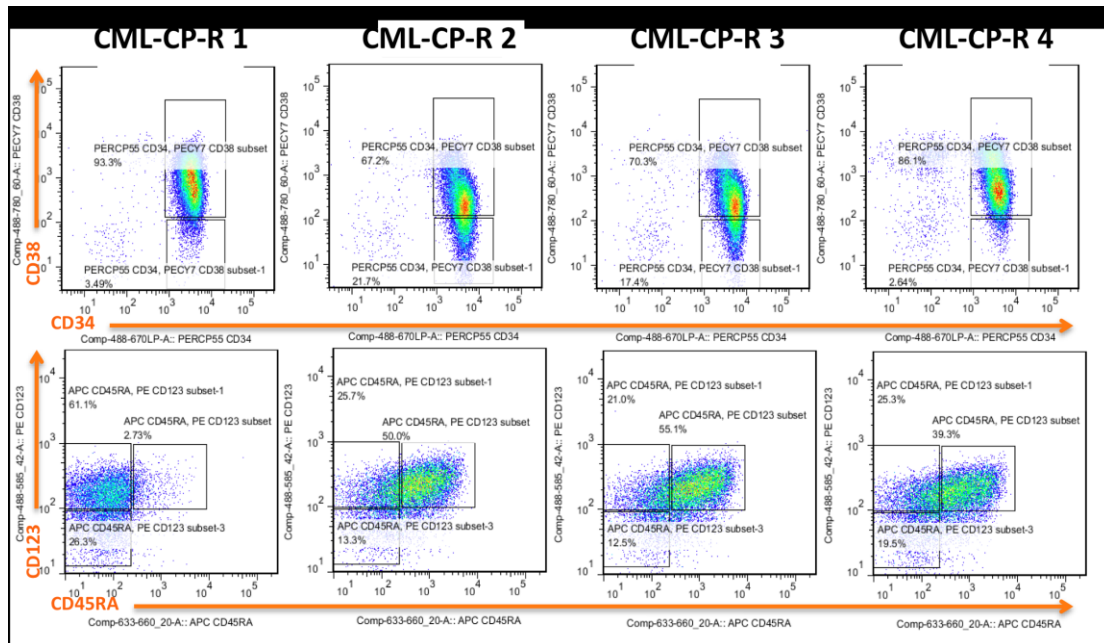


Figure 3.19 FACS plots of CD34+ haematopoietic stem (HSC) and Progenitor cell (HPC) populations from 4 untreated CML-CP patients (R, responders).

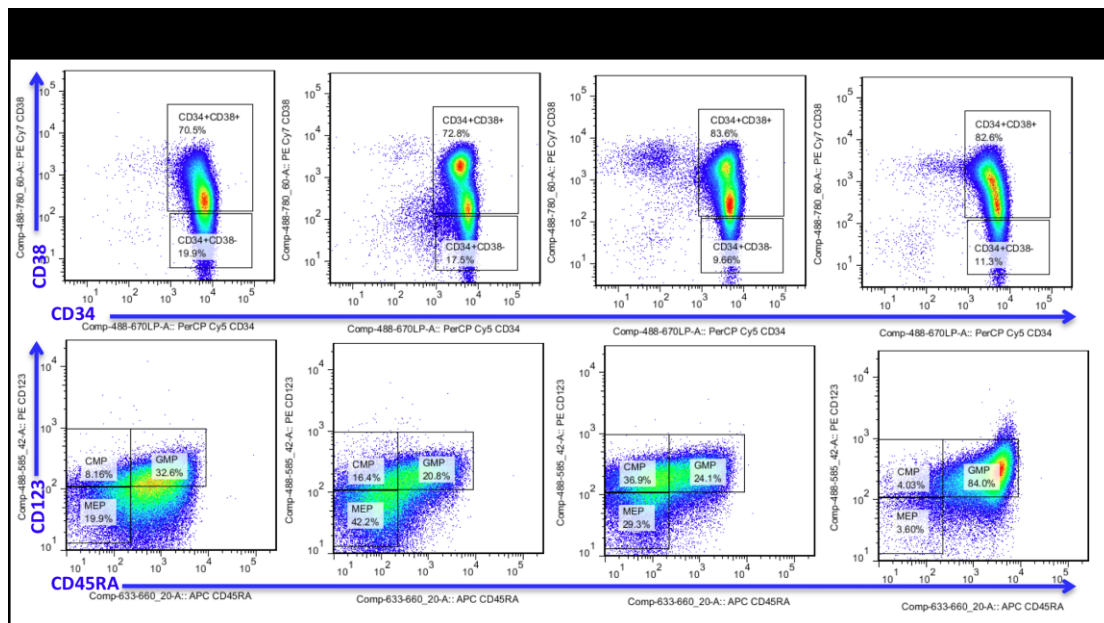
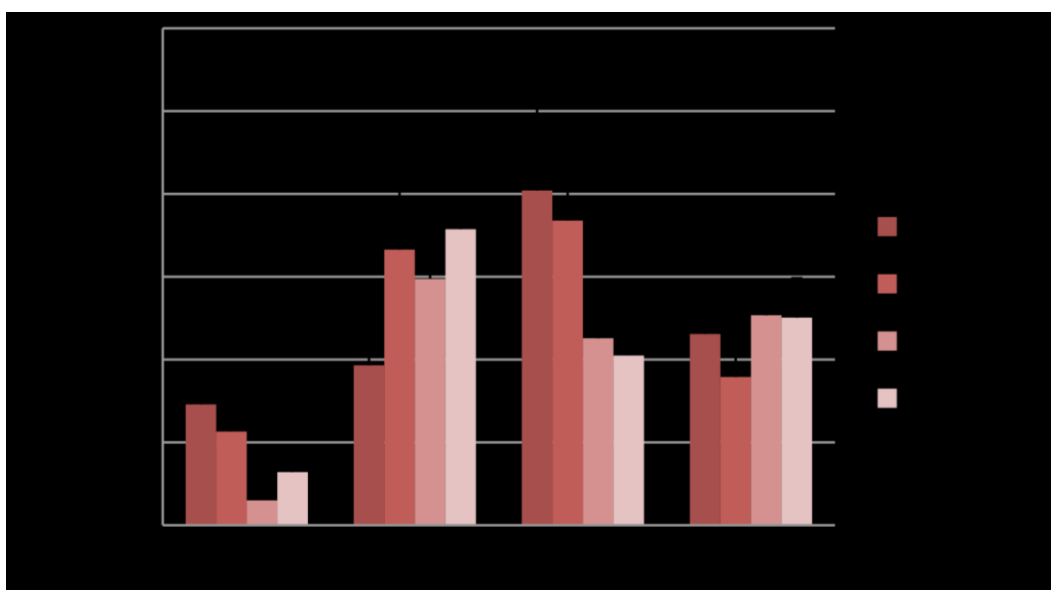


Figure 3.20 FACS plots of CD34+ haematopoietic stem (HSC) and Progenitor cell (HPC) populations from 4 untreated CML-CP patients (NR, non-responders)



	HSC			CMP			GMP			MEP		
	CCyR	Control	Control	CCyR	Control	Control	CCyR	Control	Control	CCyR	Control	Control
	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.
	CML-CP	CML-CP	CCyR	CML-CP	CML-CP	CCyR	CML-CP	CML-CP	CCyR	CML-CP	CML-CP	CCyR
<i>p-value</i>	0.0949	0.2331	0.4312	0.474	0.1783	0.0921	0.132	0.0734	0.2925	0.1797	0.5152	0.9021

Figure 3.21 The proportion of FACS sorted CD34⁺ stem and progenitor cell subsets in CML patients (CML-CP-R/NR and CML-CCyR) and healthy controls

Analysis of the sorted subsets demonstrates that while there is variation between sample types, none of these reach statistical significance using Mann–Whitney U (Wilcoxon rank) test ($p < 0.05$). Note, CML-CP (responder and non-responder) samples were combined for the purpose of analysis. The error bars denote the standard error of the mean (SEM).

3.4.8.2 Patients at the time of CCyR show global methylation patterns closer to healthy controls than paired CML-CP samples

The median time to mobilisation from commencement of imatinib was 22.7 months (range 13.3-55.6 months). Each patient had achieved at least CCyR on imatinib (confirmed by conventional cytogenetic analysis or FISH) with all but 3 individuals attaining a major (MMR) on the international scale (IS) (BCR-ABL/ABL ratio median 0.01%, range 0-0.25%).

Samples were collected following G-CSF administration with patients receiving a median of 5 doses (range 4–7 doses). A total of twenty-two individuals (76%) underwent a median of two (range 1-3) apheresis procedures, collecting a median dose of $2.3 \times 10^6/\text{kg}$ CD34⁺ cells (range $0.9\text{--}5.5 \times 10^6/\text{kg}$). Due to inadequate cell doses, 7 patients elected to undergo an additional harvest procedure on a separate occasion that necessitated a short (7-10 day) break in TKI therapy and higher doses of G-CSF prior to leucapheresis in an attempt to maximise the CD34⁺ stem cell yield.

At the time of CML-CCyR, 3,440 (85%) of the 4,070 CML-CP DMPs (filtered for >20% differences) had methylation values that had returned to within 20% of the healthy controls including 94% of hyper- and 86% of hypomethylation probes. Despite this restoration towards normal, variability at individual sites was evident (Figure 3.22) with 24,597 probes (52% of the original 47,152 called DMPs before the 20% filter was applied) still showing significant methylation differences (F-test with genome wide corrected p-value < 0.01) (Figure 3.23 and Figure 3.24).

On further investigation, the degree of restoration towards ‘normal’ in CML-CCyR patients was independent of a patient’s methylation state at diagnosis; CML-CP patients with the most disrupted global methylation patterns did not show the greatest residual differences at the time of remission.

Furthermore the probes showing the highest degree of restoration (smallest beta-value differences between CML-CCyR and healthy controls) were consistently the same probes between individuals i.e. a restoration signature amongst patients.

In an attempt to explore the variable degree of methylation restoration in CML-CCyR patients, clinical parameters were tested using a Pearson's correlation coefficient (r).

There was no statistically significant correlation between the extent of restoration and the duration of diagnosis/time treated on imatinib prior to mobilisation ($r = 0.090$), the level of the BCR-ABL1 transcripts at the time of mobilisation ($r = 0.118$) or the time to attainment of MMR regardless of whether this was achieved before or following mobilisation ($r = -0.044$).

Given that both CML-CCyR and healthy controls groups were exposed to G-CSF, the DMPs between them were separated further into those assumed to be the result of residual disease (less than 20% differences between CML-CP and CML-CCyR compared to healthy controls) and those possibly associated with imatinib therapy (more than 20% differences between CML-CCyR and the other groups) (Figure 3.25).

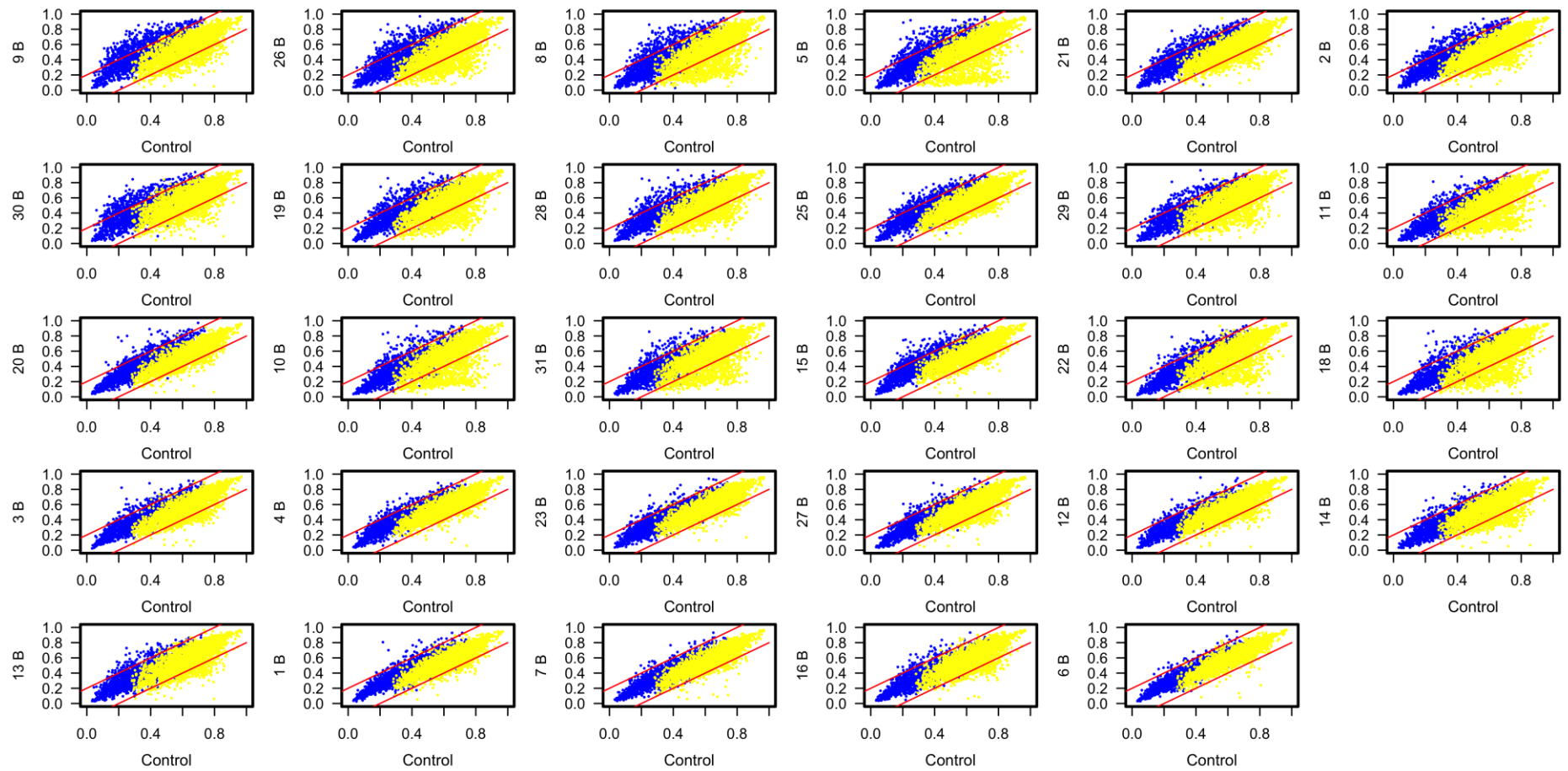


Figure 3.22 Restoration of CML DMPs at the time of CML-CCyR

Each small plot represents an individual patient at the time of remission (CML-CCyR). Blue points are those DMPs ($> 20\%$ differences) that were hypermethylated at diagnosis (CML-CP) compared with controls. Conversely, the yellow points were hypomethylated. Black points within the red parallel lines were probes showing $< 20\%$ differences at diagnosis i.e. were considered to show smaller differential methylation differences. In all patients there is a clear trend for methylation beta values returning closer to normal, the degree however, varies between individuals. Patients 5B and 26B (top, left) have global methylation patterns that correlate less closely with healthy controls compared to patients 6B and 7B (bottom right) where the majority of original DMPs have been restored. (x -axis beta value healthy control, y -axis beta value CML-CCyR).

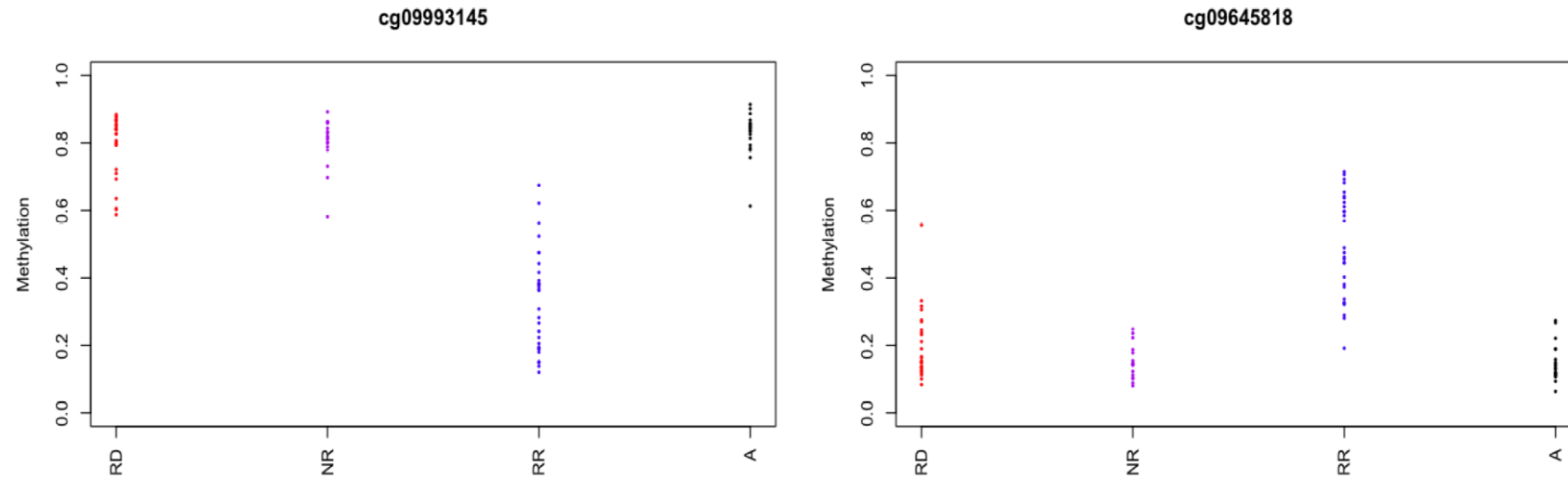


Figure 3.23 Probes representing methylation patterns unique to CML-CCyR samples

Despite having untreated disease, all CML-CP patients [red, responder (RD) and purple, non-responder (NR)] have similar methylation beta values (y -axis) as healthy control samples (G-CSF mobilised) (black, A). Strikingly, the CML-CCyR patients (G-CSF mobilised) (blue, RR) show the greatest differences. Probe cg09993145 (left) shows marked hypomethylation and cg09645818 (right) relative hypermethylation in the CML-CCyR samples compared to the other groups.

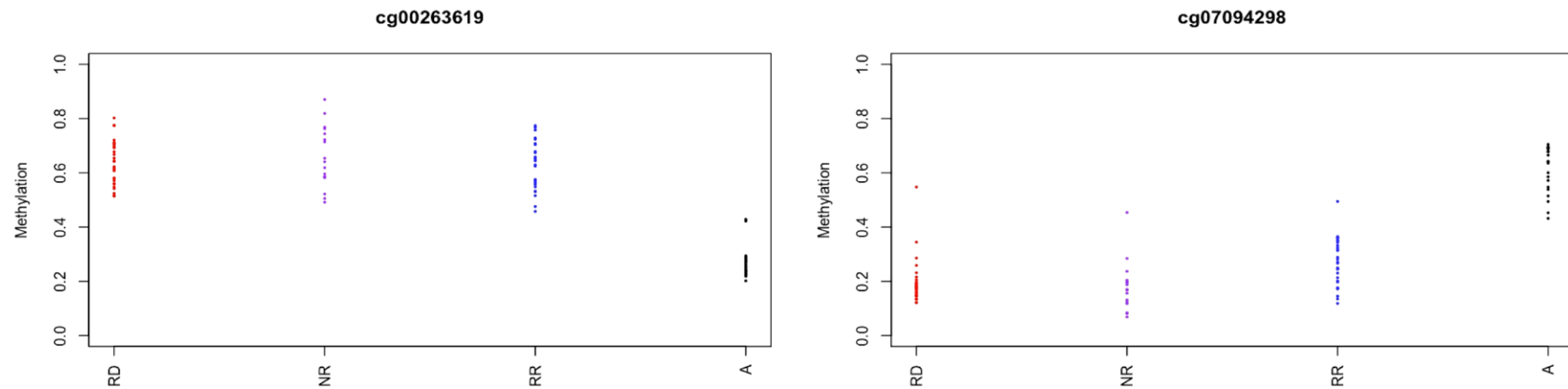


Figure 3.24 Probes representing methylation patterns common to all CML samples

Probes cg00263619 and cg07094298 represent CML-unique methylation changes. On the left, all CML samples regardless of stage or response to TKI therapy [red, responder (RD) and purple, non-responder (NR)] show aberrant hypermethylation relative to the healthy controls (black, A). Conversely, the right plot shows relative hypomethylation. These probes are not therapy-driven since CML-CCyR patients were on imatinib therapy in contrast to CML-CP patients. Both CML-CCyR and healthy controls have received G-CSF.

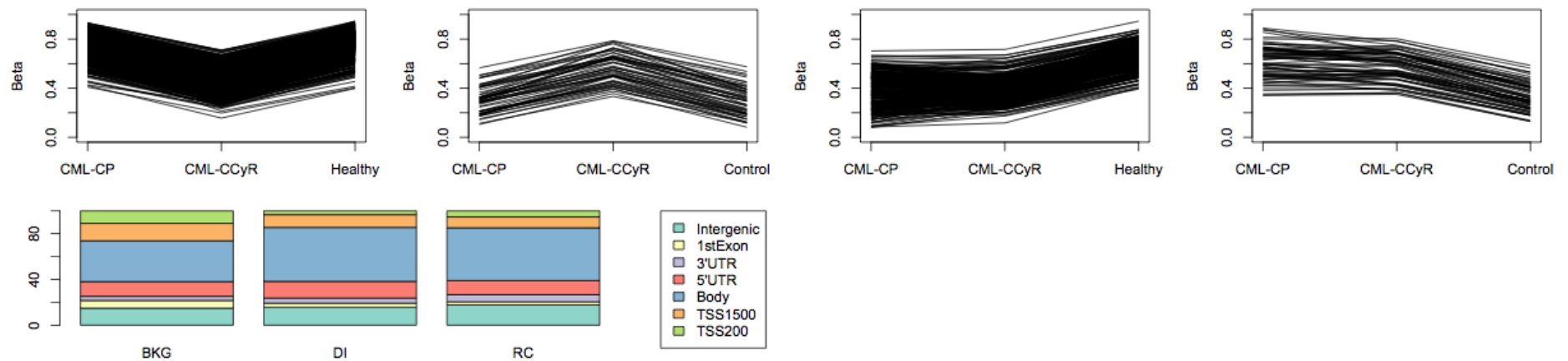


Figure 3.25 Drug induced and residual disease DMPs in CML-CCyR samples

Drug induced (DI) probes represent CpGs that are only differentially methylated in CML-CCyR samples (with > 20% filter applied). Patients in this category are the only group receiving long-term TKI therapy. Top panel, (left graphs) display possible DI methylation changes of which 1710 (97%) probes are hypomethylated and 53 are hypermethylated with enrichment for gene body methylation (Fold enrichment: 2.1, p-value < 1e-4) and intergenic regions (Fold enrichment: 1.7, p-value < 1e-4) but depletion in promoters (Fold enrichment: 0.89, p-value = 0.0035) and first exons (Fold enrichment: 0.43, p-value < 1e-4) compared to background (BKG, lower panel).

CML-CCyR DMPs with less than 20% residual differences between CML-CP and healthy controls i.e. partial restoration towards normal have been labeled 'residual CML' (RC). RC differences (top panel, right) in CML-CCyR samples consist of 302 (81%) hypomethylated and 71 (19%)

hypermethylated probes. RC probes also showed enrichment within gene bodies (Fold enrichment: 2.0, p-value $<1e-4$) and intergenic regions (Fold enrichment: 1.8, p-value $<1e-4$) with depletion in first exons (Fold enrichment: 0.52, p-value $<1e-4$).

3.4.8.3 The effects of combined exogenous G-CSF and TKI therapy drives methylation differences in CML-CCyR samples

The DMPs that defined CML-CCyR and separated this group from the others were further interrogated using Marmal-aid.

CML-CCyR samples were compared to CML-CP, and healthy control study samples along with publically available HM450 data from other blood cell types including CD19⁺, CD14⁺, CD8⁺, CD4⁺ and CD56⁺ expressing cells as well as whole blood samples and PBMCs (Figure 3.26).

Unexpectedly, 12 of 29 (41%) CML-CCyR samples clustered separately from the rest of the group and appeared to have methylation patterns for the selected probes that closely resembled CD14⁺ cells, a cell surface protein expressed on monocytes. The remaining 17 CML-CCyR samples clustered more closely with whole blood and terminally differentiated leucocytes than other CD34⁺ cells such as cord blood and healthy G-CSF mobilised controls.

The striking methylation similarities between CD14⁺ monocytes and a proportion of the CML-CCyR cells required further investigation to see whether the findings could be attributed to artefactual or biological causes.

The possibility of selecting the incorrect fractions during sample processing was ruled out on the basis that CD34⁺ purity had been assessed by single colour FACS. Furthermore, from the laboratory records, CML-CCyR samples (including those displaying CD14-like methylation patterns) had been processed in an identical manner on different days over a number of weeks alongside CML-CP samples which were unaffected i.e. the finding was not batch-related.

To eliminate the possibility of CD14 and CD34 surface co-expression representing an exogenous G-CSF-driven intermediate immunophenotype, 3 healthy controls, 3 CML-CCyR samples exhibiting CD14⁺ cell-like methylation patterns and 3 CML-CCyR samples resembling healthy whole blood were subjected to flow cytometry after staining with CD14 and CD34 antibodies as previously described. The CD34 negative fractions from the same samples were used as controls.

It was possible to demonstrate immunophenotypically that the cells were CD34 surface marker positive and did not co-express CD14 (Figure 3.27). Within the CD34⁺ population, CD14⁺ cells were present in greater numbers in all CML-CCyR samples when compared with healthy controls, regardless of which methylation clustering they showed.

Both CML-CCyR and healthy control samples were collected by G-CSF mobilisation. It is possible that treatment with haematopoietic growth factor resulted in disruption to cellular methylation patterns accounting for the differences observed. Yet, given these two groups were exposed to the drug with only half of a single group exhibiting alterations, it is highly unlikely that G-CSF alone can account for the divergence in appearance of the CML-CCyR samples.

From the clinical data it became apparent that those samples with CD14-like methylation patterns showed a tendency towards lower cell numbers and the need for repeat harvest procedures requiring higher doses of G-CSF.

While this may in part reflect the underlying mechanism for the altered patterns, higher doses of G-CSF cannot exclusively account for the findings as the aliquots of cells used in this study were all from Day 1 harvest samples (the same as healthy controls) regardless of whether the optimal clinical dose was achieved or additional harvest procedures attempted.

Another consideration is the observation that temporarily withholding imatinib therapy two weeks prior to stem cell harvest improved collection success rates in 'poor-to-mobilise' patients perhaps implicating TKIs. Indeed no G-CSF mobilised healthy controls had CD14-like patterns despite receiving the same doses of the growth factor. One inference that could be drawn is that the interaction of imatinib and G-CSF were driving the observed differences.

Re-examining Figure 3.22, it was clear that the degree of restoration towards normal in the CML-CCyR samples was entirely driven by the unexplained divergence in methylation patterns. Without exception, the CD14-like samples correlated less with healthy controls than those clustering with whole blood. Pathway analyses of CML-

CCyR DMPs were delayed on the basis that a proportion of the probes were representing drug effects/interactions rather than genuine disease-state differences.

One way of approaching the unexpected clustering discrepancy was to remove all the CD14-like samples from further analysis and only consider the remaining samples. But to further elucidate and better understand the reasons underpinning the separation it was decided that further sample collections should be carried out.

In an attempt to detect genuine CML-CCyR differences and remove the need for exposure to G-CSF, a proportion of CML patients used in this study consented to additional bone marrow aspiration procedures. A total of 8 patients, 4 that clustered with healthy whole blood and 4 with CD14⁺ cells were collected according with full ethical approval. These patients remained on TKI therapy at the time of the aspiration since discontinuation of treatment is entirely avoided unless clinically indicated outside the controlled clinical trial setting. In addition, 7 equivalent frozen, healthy control bone marrow samples, from different individuals to those used previously were obtained from The John Goldman Stem Cell Laboratory selected closely on the basis of age and ethnicity. After collection, and following ficol density separation, patient bone marrow samples were frozen down for at least one month to reproduce the storage conditions of the healthy control samples.

The samples were processed in a total of three batches in an identical manner to the previous study samples with a combination of CML-CCyR and healthy control bone marrow per batch to control for any bench-based variation e.g. different aliquots of kit reagents. Bisulfite converted DNA from CD34⁺ selected cells were hybridised to one of four HM450 chips using a random number generator to allocate positions and eliminate possible processing bias. The remaining array positions were allocated to cell line samples and duplicate CML-CCyR (n= 6) and healthy control (n= 2) bone marrow samples for internal validation across Beadchips, which were processed at the UCL Genomics facility, Institute of Child Health, London.

Following pre-processing and normalisation, unsupervised hierarchical clustering confirmed that the bone marrow samples grouped closely together (Figure 3.28). In keeping with previous observations, the greatest differences seen were between the mobilised CD34⁺ CML-CCyR samples and all other remission/healthy control groups. Although lack of remaining material meant it had not been possible to

rehybridise any of the G-CSF mobilised samples i.e. CML-CCyR or healthy controls to the additional BeadChips, the healthy G-CSF-mobilised CD34⁺ cells clustered very closely with CD34⁺ bone marrow samples despite a batch effect.

Marmal-aid was used to look at the clustering in the context of other cells types. Using sets of 10 probes, defined as being cell subtype specific i.e. analogous to a cell surface markers, beta-value data were extracted for CD34⁺ (GEO accession numbers GSM1108073, GSM1108075, GSM1139408 and GSM1139409), CD14⁺ (GSE43976), CD4⁺, CD19⁺ and CD56⁺ cells as well as eosinophils and neutrophils (All GSE35069) (Figure 3.29). One drawback is the lack of demographic data associated with the GEO samples. CD34⁺ cells were all male without a specified age while the remaining samples were collected from a combination of males (59%) and females (41%). Only the series accession GSE35069 provided an age range (24-51 years) but not individual ages of the study subjects.

In the G-CSF mobilised CML-CCyR samples, a loss of CD34⁺ methylation marks and gain in CD14⁺ cell patterns could be seen which were not present in any other groups including the paired bone marrow samples (on imatinib therapy with no recent G-CSF exposure) or healthy controls (no imatinib therapy but short-term G-CSF administration). Taken together, the methylation pattern changes seen in mobilised CML-CCyR samples are likely to be related to the combined effects of imatinib and G-CSF but the reasons for this remain unclear.

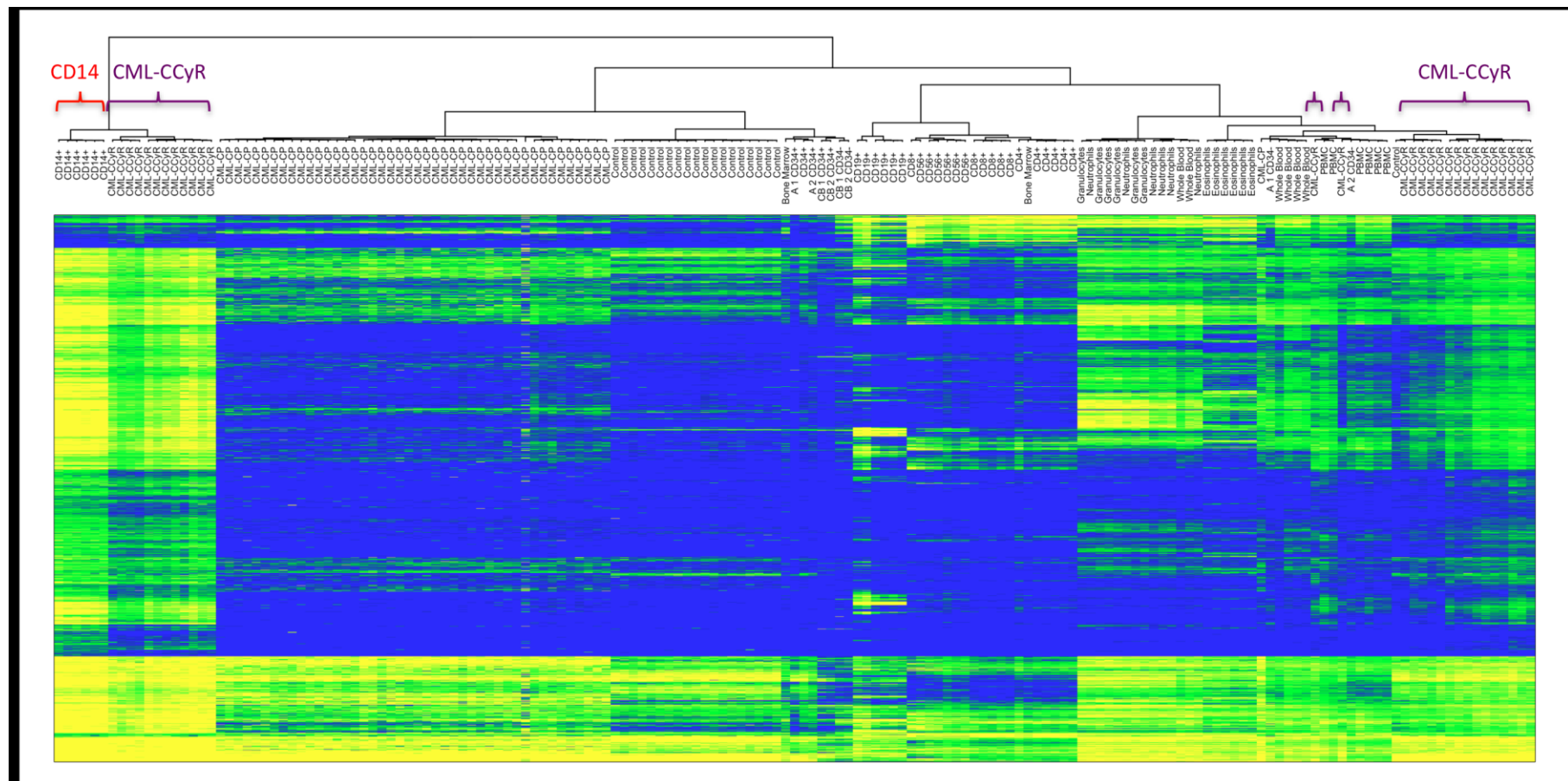


Figure 3.26 Dendrogram showing hierarchical clustering of CML-CCyR samples with other blood cell types using Marmal-aid.

CD34⁺ CML-CCyR samples cluster in two distinct groups; those with CD14⁺ cells and those with whole blood/PBMCs (highlighted). Healthy mobilised CD34⁺ cells from this study cluster with unmobilised CD34⁺ cord blood cells from publically available data through Maraml-aid .

Absolute beta-values using color scale 1 (blue) = fully methylated, 0.5 (green) = partially methylated, 0 (yellow) = unmethylated.

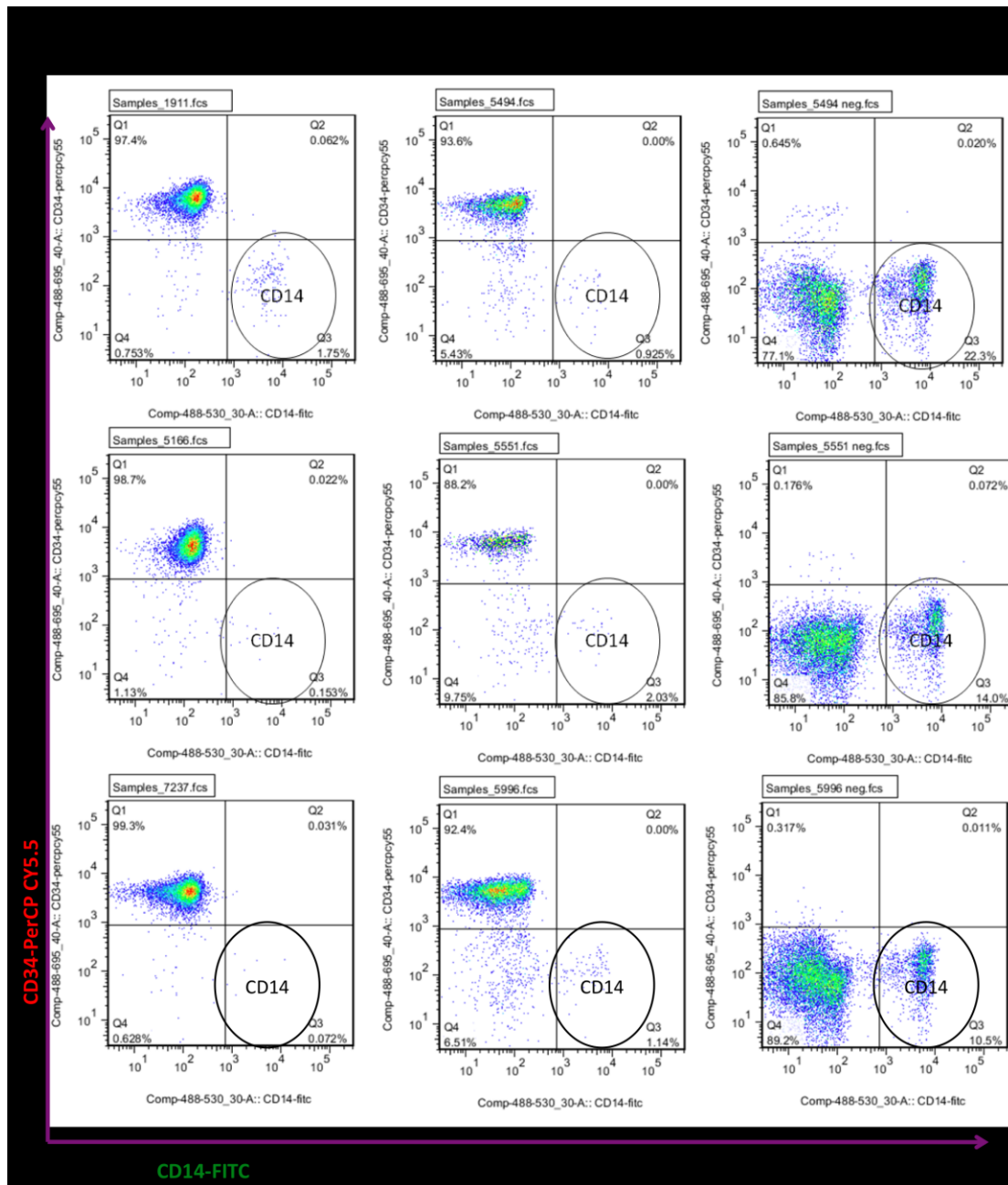


Figure 3.27 FACS plots comparing CD14 expression on ‘CD14-like’ CML-CCyR cells and healthy control samples

CD34⁺ cells do not co-express CD14 (middle column). As a control and for comparison, CD34⁻ cells from ‘CD14-like’ CML-CCyR samples were also analysed. As expected, as this stage of differentiation, a proportion of these cells are CD14⁺ (CD34⁻) monocytes (right column).

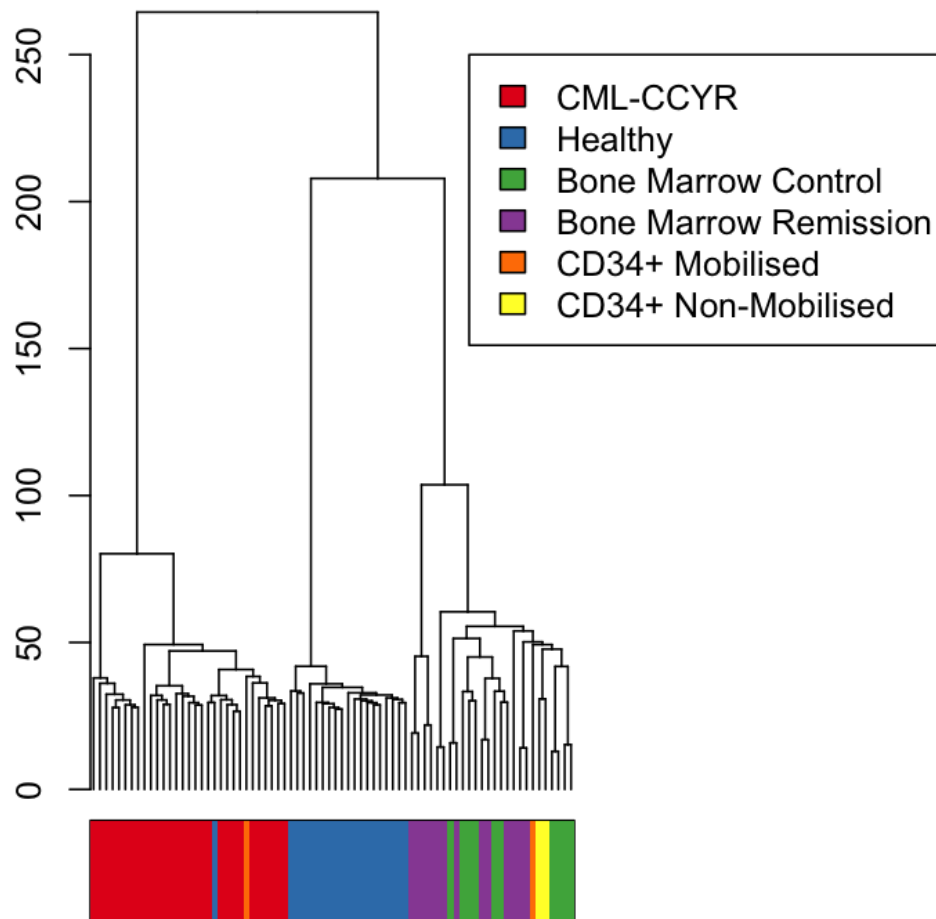


Figure 3.28 Cluster dendrogram comparing CD34⁺ cell methylation differences in G-CSF mobilised and bone marrow harvested samples

CD34⁺ CML-CCyR mobilised cells (red) cluster furthest as a group from all other samples. With the exception of one sample (discussed earlier), mobilised healthy controls from the Hammersmith cohort (blue) cluster much closer to CD34⁺ bone marrow cells [CML-CCyR (purple), healthy controls (green)]. The remaining samples, CD34⁺ mobilised (orange) and CD34⁺ non-mobilised (yellow) cells were publically available HM450 data through Maramal-aid. The non-mobilised cells are CD34⁺ cord blood cells and are indistinct from the CD34⁺bone marrow samples. The fact that healthy GCSF mobilised cells cluster closer to healthy blood and bone marrow samples than CML-CCyR despite being processed in the same batch as the

latter, suggests that clustering is in part determined by disease state rather than batch. While some degree of batch effect is likely, the precise contribution of batch versus disease state is not possible to determine.

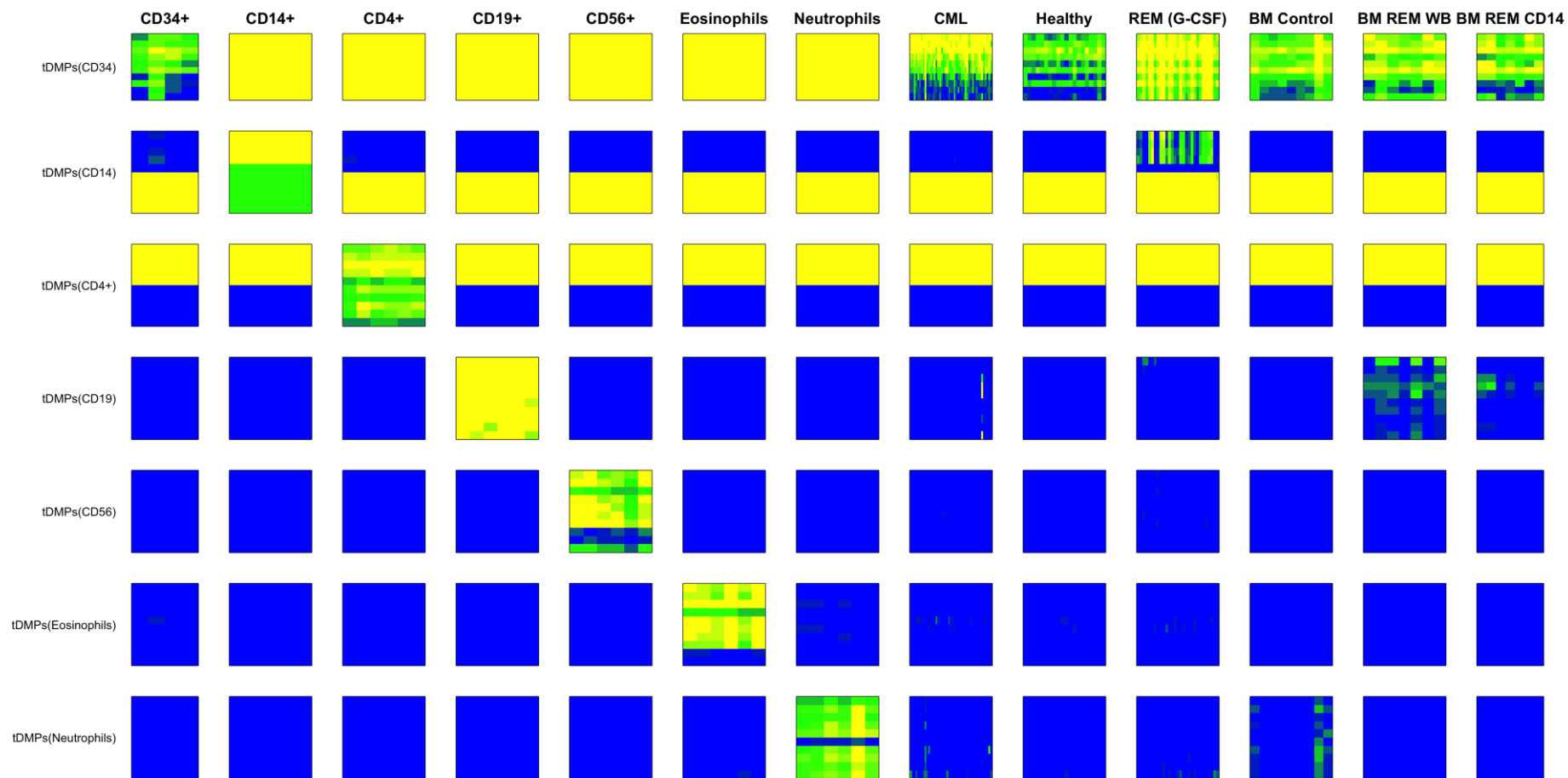


Figure 3.29 Cell-type composition methylation differences using Marmal-aid

Each column represents grouped samples extracted from Maraml-aid including 7 different blood cell subtypes (CD34, CD14, CD4, CD19, CD56, eosinophils and neutrophils) and the study samples [Hammersmith/SPIRIT 2 CML-CP (CML), G-CSF mobilised CD34⁺ healthy controls (healthy), G-CSF mobilised CD34⁺ CML-CCyR (REM G-CSF), CD34⁺ healthy control bone marrow (BM Control) and CD34⁺ bone marrow CML-CCyR samples with methylation patterns at the time of mobilisation that resembled whole blood (BM REM WB) or CD14⁺ cells (BM REM CD14)].

Each row represents 10 probes with the smallest p-values that distinguish the 7 different cell subtypes from each other. The resulting diagonal boxes from CD34⁺ to neutrophils represent these unique differences.

The first plot in the top row defines CD34⁺ cells and consequently all study samples reflect the cell-type of origin to a variable extent with the healthy mobilised blood and all bone marrow samples exhibiting the highest degree of correlation. CML-CP samples show a loss of methylation (green to yellow) at some DMPs in keeping with cancer changes but retain recognisable marks of CD34⁺ cells while mobilised CML-CCyR samples (REM G-CSF) have greatly disrupted methylation levels across all probes.

The CpGs that delineate CD14⁺ cells are shown in the second row. In contrast to all other cell types and samples, a proportion of G-CSF mobilised CML-CCyR samples (REM G-CSF) have losses in methylation consistent with gaining CD14-like methylation signatures.

3.4.8.4 CML-CCyR is a distinct epigenetic phenotype

Despite the strong correlation in methylation beta-values between G-CSF mobilised and bone marrow (BM) healthy control samples, only DMPs between CML-CCyR and healthy control bone marrow samples were called to standardise sampling procedures and starting material.

A total of 5,846 DMPs separated CML-CCyR-BM and healthy control CD34⁺ bone marrow samples (adjusted p-value < 0.01) including 2,253 with beta-value differences more than 10%, with 137 of these greater than 20% (i.e. 93 hypo- and 44 hypermethylated probes) (Table 3.8 and Table 3.9). Overall, the CML-CCyR-BM signature showed a global loss of methylation compared with healthy controls.

Of the 93 hypomethylated probes with > 20% differences in CML-CCyR-BM samples, it was possible to map 55 (59%) to RefSeq genes with 30 out of 55 (56%) located in CpG islands (GGI). Interestingly, six genes were represented by multiple probes providing evidence of biological validity given the random placement of these probes on HM450.

Loci cg00446123, associated with a CGI within the transcription start site to 200 bp upstream (TSS200) of LIME1 had the greatest significant difference between remission and control samples. In addition, two other probes cg21201401 and cg20513976 mapping to the same CGI were among the top ten remission hypomethylated loci.

LIME1 encodes the Lck-interacting transmembrane adapter 1 phosphoprotein that is preferentially expressed in haematopoietic cells [311] and involved in both T-cell receptor (TCR)-mediated T-cell activation [312] and more recently, the initial activation of B-cell receptor (BCR)-induced B-cell signaling through interaction with molecules Grb2, Lyn, and p85, the regulatory subunit of PI3K [313].

Moreover, treatment with imatinib is associated with impaired B-cell responses, specifically reduction in IgM memory B cells through mechanisms likely associated with off-target inhibition of kinases important in B-cell signaling [314]. Taken together, LIME1 TSS200 hypomethylation could reflect the fact that the effects of ongoing TKI therapy influence the CML remission signature.

Next, the methylation levels of the 5,846 probes were investigated in the context of the original CML-CP DMPs (n= 50,718 before the beta-value 20% cut off). Although 2,117 (including 34 out of 137 remission DMPs with greater than 20% differences) were significantly altered at diagnosis and remained so at remission, probe directionality between the two disease states were not always consistent at given CpGs i.e. a hypomethylated probe at diagnosis could demonstrate gains in methylation at CML-CCyR to become significantly hypermethylated compared to controls. Moreover, 3,729 (63.8%) of the remission DMPs were novel and not present at diagnosis (Figure 3.30 and Figure 3.31).

Gene ontology (GO) enrichment analysis of the 5,846 remission probes was performed after separating CpGs in to those that were hypo- or hypermethylated in relation to healthy controls. For the hypermethylated remission probes, 229 significant pathways ($p < 0.01$, Benjamini-Hochberg correction, containing at least 10 annotated genes) were over represented including many associated with programmed cell death such as induction of apoptosis by extracellular signals (GO:0008624, fold enrichment 6.8X; p -value: 6.87×10^{-9}) and positive regulation of programmed cell death (GO:0043068, fold enrichment 3.25X; p -value: 7.08×10^{-13}). There was also considerable selection for categories relating to protein tyrosine kinase signaling including the activation and regulation of phosphorylation (GO:0042325, fold enrichment 2.52X; p -value: 4.25×10^{-8}), protein kinase activity (GO:0045859, fold enrichment 3.07X; p -value: 3.57×10^{-9}) and the protein kinase cascade (GO:0007243, fold enrichment 3.04X; p -value: 3.35×10^{-9}). Amongst the other noteworthy terms, response to drug (GO:0042493, fold enrichment 2.41X; p -value: 0.0024) consisted of 26 genes including members of the solute carrier family 22 and in particular, the organic cation transporter SLC22A1 (hOCT1), a drug influx pump responsible for transportation of imatinib into target cells [60, 62].

One probe, cg02915920 located in the gene body of ABL1 appeared in 23 of the significant pathways and was consistently hypermethylated in remission samples compared to healthy controls (beta-value 0.82 vs. 0.73 respectively, genome-wide corrected q -value: 0.008).

A total of 252 biological processes were over represented for genes associated with hypomethylated remission probes. The derived categories showed considerable

overlap with those generated for the hypermethylated probes, including the GO terms discussed above, sometimes involving different genes in these same pathways.

The presence of dysregulated DNA methylation across gene regulatory networks at the time of CML-CCyR was evident but the biological relevance not entirely clear. That many of the DMPs were absent in healthy controls and patients at diagnosis supports the idea that TKI therapy is associated with dynamic epigenetic changes resulting in a distinct remission phenotype.

Table 3.8 CpGs showing >20% hypomethylation differences in CML-CCyR samples with respect to healthy control bone marrow (FDR < 0.01)

Illumina Probe ID	Chromosome	MAPINFO probe location	SNP	UCSC Gene Name	UCSC Gene position	UCSC CpG Island	BM Healthy Control	BM Remission	q-value
cg00446123	20	62367888		LIME1	TSS200	chr20:62368955-62371962	0.6041924	0.300921102	5.21E-05
cg03309308	16	68525329					0.61683703	0.41181193	5.21E-05
cg21201401	20	62367884		LIME1	TSS200	chr20:62368955-62371962	0.534838102	0.25650306	5.74E-05
cg04083553	3	128778575		GP9	TSS1500		0.753907248	0.551157182	8.58E-05
cg13475995	10	5882576				chr10:5882303-5882635	0.700619592	0.433909694	8.86E-05
cg08063160	2	238578376	rs4663780	LRRFIP1	Body		0.71269796	0.481625918	8.86E-05
cg20513976	20	62367893		LIME1	TSS200	chr20:62368955-62371962	0.543458433	0.282787897	0.000104553
cg03382805	22	23905582				chr22:23908322-23909178	0.789729688	0.564911502	0.000193383
cg27106643	18	77257352		NFATC1	Body	chr18:77255680-77256046	0.742872515	0.427874822	0.000260528
cg26580869	19	611506		HCN2	Body	chr19:611291-611579	0.795461938	0.588227305	0.000276613
cg06804344	3	128778699		GP9	TSS1500		0.542168612	0.339780035	0.00031329
cg14763104	11	2406384		CD81	Body	chr11:2406711-2406927	0.727148931	0.509934652	0.00031329
cg06687504	6	167263454		RPS6KA2	Body		0.826373768	0.591076805	0.000328209
cg07698804	17	80829309		TBCD	Body	chr17:80829533-80829871	0.802754207	0.489750101	0.000341679
cg05205074	14	107259880	rs72692491				0.836240568	0.612217592	0.000390641
cg22889918	10	115024317					0.803220853	0.58702169	0.000428232
cg09217898	3	194412522				chr3:194406440-194409045	0.703645179	0.489406931	0.000443094
cg14557690	9	139978293		UAP1L1	3'UTR	chr9:139980377-139981887	0.727119392	0.507319429	0.000457698
cg20666585	15	31523186					0.790867595	0.572289543	0.000463867
cg02947214	14	105998115				chr14:105992499-	0.770810947	0.555945272	0.000542735

cg22498365	17	80829028	rs80351902	TBCD	Body	105996414 chr17:80829533-80829871	0.776995507	0.570564741	0.000558925
cg04856396	17	3589140		P2RX5	Body		0.509217569	0.253318434	0.000567716
cg08428188	16	85363174				chr16:85363116-85363360	0.70680781	0.496606179	0.000567716
cg12216435	3	128995479		COPG	Body	chr3:128997382-128997600	0.678712886	0.47556429	0.000602173
cg09161446	2	240697468					0.786541085	0.57668179	0.000603999
cg13823257	9	35666823				chr9:35665059-35665364	0.818064315	0.603457737	0.000636575
cg22826986	8	11351507		BLK	TSS200		0.802237372	0.585772463	0.000643057
cg20618826	10	81148896	rs7070039	ZCCHC24	Body	chr10:81145958-81146197	0.780802498	0.569147457	0.000659677
cg00760938	2	217357159					0.80450797	0.60295626	0.000668745
cg13478045	9	137477320					0.584721077	0.36842537	0.000705366
cg21248060	7	1039957	rs13307692	C7orf50	Body	chr7:1039874-1040096	0.758380743	0.534611551	0.000719507
cg17995557	10	126289971		LHPP	Body		0.822159664	0.60917926	0.000719507
cg04819180	17	80829157		TBCD	Body	chr17:80829533-80829871	0.725634854	0.517042083	0.000744471
cg03302822	10	99928067		C10orf28	5'UTR		0.771537622	0.570368504	0.000744551
cg08155625	21	44851244	rs128543				0.677904538	0.421528489	0.00076619
cg13482010	3	122703782		SEMA5B	5'UTR		0.780875378	0.578454506	0.000767403
cg11767757	21	40145404		NCRNA00114	TSS200		0.749155236	0.511955962	0.000770889
cg07721872	16	87735256		LOC100129637	Body	chr16:87736652-87736907	0.868617369	0.660675206	0.000773608
cg27183818	1	85764021					0.811973951	0.593985244	0.000890155
cg12499316	2	106242236	rs5833142				0.771006774	0.533883239	0.000898054
cg01644640	2	40147801					0.712014356	0.507172948	0.000944063
cg10203922	4	145566200		HHIP	TSS1500	chr4:145566242-145567413	0.767410328	0.54593163	0.000976336
cg19905757	15	68924127		CORO2B	Body		0.754806515	0.54827783	0.000979307
cg03160526	17	80928410		B3GNTL1	Body	chr17:80929062-80929354	0.545554143	0.335908057	0.00097974
cg00225902	3	66444122	rs6790222	LRIG1	Body		0.82762836	0.602374941	0.000996165
cg27031754	5	54185940					0.816393035	0.598399305	0.001007835
cg03917666	10	3977608					0.838525654	0.638138042	0.001010255
cg15730481	7	1126318		GPER	TSS200	chr7:1126578-1126910	0.820362421	0.607411376	0.001053131
cg01869288	15	31523257					0.800851525	0.572392608	0.001068137
cg25712005	7	73443113		ELN	Body		0.528281027	0.315662452	0.001104327
cg12604490	6	99373813		FBXL4	Body		0.809473878	0.602091883	0.001113155

cg12074150	2	34128722					0.67981837	0.111835655	0.001183333
cg22175624	17	80829261		TBCD	Body	chr17:80829533-80829871	0.752437011	0.505416093	0.001357488
cg18792536	7	76145562		UPK3B	3'UTR	chr7:76145396-76145781	0.609557128	0.229357363	0.001452346
cg18184053	10	5882739				chr10:5882303-5882635	0.805979196	0.603774219	0.001533401
cg09267773	16	79802467				chr16:79804241-79804694	0.594195302	0.385514893	0.001538662
cg19344626	19	16830749		NWD1	TSS200		0.771143085	0.556898907	0.001697152
cg02402436	6	31540051	rs36221309	LTA	TSS200; 1stExon		0.709425886	0.47440368	0.001794972
cg23549902	7	5184155				chr7:5183878-5184223	0.613861492	0.324047929	0.001828181
cg04450994	6	3318592		SLC22A23	Body		0.772718716	0.554552921	0.001923657
cg25134567	17	64782369		PRKCA; MIR634	Body; TSS1500		0.68883819	0.471634806	0.002021368
cg08122386	6	26195954				chr6:26197070-26197537	0.405423319	0.201239862	0.00215555
cg16345226	19	16771394	rs7255269	TMEM38A; C19orf42	TSS1500; TSS1500	chr19:16770636-16772632	0.478691698	0.272345625	0.00217775
cg16308790	18	77225973		NFATC1	Body	chr18:77227273-77227727	0.689006656	0.449899883	0.002301073
cg15572235	7	5183992				chr7:5183878-5184223	0.583623875	0.36441395	0.002430667
cg09621572	6	31539973	rs36221306	LTA	TSS200; 1stExon		0.870576116	0.645791093	0.002460991
cg25637655	6	29911542	rs29028728	HLA-A	Body	chr6:29910202-29911367	0.811436206	0.394058208	0.002464741
cg24893378	5	59837426		PART1	Body		0.779968144	0.535326737	0.002659079
cg24361265	15	44068668		ELL3	Body	chr15:44068586-44069792	0.755333787	0.506206426	0.002945802
cg26021007	17	45283040					0.59857918	0.396748507	0.002954727
cg23222617	14	89653143		FOXP3	Body		0.732887413	0.527467932	0.002967723
cg14437551	6	31539986	rs36221306	LTA	1stExon; 5'UTR		0.88225313	0.639637318	0.003004427
cg03181300	6	26195995				chr6:26197070-26197537	0.3403027	0.100254647	0.00315135
cg23201527	20	61783225					0.687652539	0.46752257	0.003324305
cg26271001	12	5313518					0.738330128	0.534685028	0.003327177
cg09371409	7	1632025					0.692411614	0.476229271	0.003327902
cg18309183	3	134373303				chr3:134369464-134370242	0.788821718	0.558974904	0.003337985
cg26021304	6	29645117		ZFP57	TSS200		0.446258828	0.243928516	0.004060497
cg11946459	6	29911558	rs29028728	HLA-A	Body	chr6:29910202-29911367	0.82670911	0.447680267	0.004231586
cg05407200	11	93586725	rs10831109			chr11:93583374-	0.72657479	0.273714802	0.004482278

cg25644740	6	29894152		HCG4P6	TSS1500	93583717 chr6:29894140- 29895117	0.460839618	0.115910047	0.004616232
cg23214071	6	32627784	rs1049123	HLA-DQB1	3'UTR		0.716086979	0.514737988	0.005484736
cg25393284	2	942653				chr2:945686- 946774	0.806200401	0.604574692	0.006503847
cg00028022	1	246984573					0.717884381	0.481479074	0.00687231
cg22335228	20	62405407		ZBTB46	Body	chr20:62406463- 62407179	0.590677846	0.320495262	0.006958369
cg08121686	1	2981840		FLJ42875	Body	chr1:2979275- 2980758	0.784164766	0.487266539	0.007729507
cg15201545	4	675137		MYL5	Body	chr4:674843- 676091	0.482798616	0.133096854	0.008255433
cg22109827	7	30727326					0.737250617	0.332540591	0.008286682
cg13179472	11	11177888					0.772560721	0.565170034	0.008929947
cg22933800	6	32605704		HLA-DQA1	Body		0.651626657	0.355955843	0.009666902
cg10776061	19	12768390		MAN2B1	Body	chr19:12767749- 12767980	0.697105565	0.295237553	0.009834346
cg06611532	13	114900021				chr13:114897358- 114898516	0.801739615	0.534567757	0.009929493
cg15848685	6	30882641	rs6926224	VAR52	5'UTR	chr6:30881533- 30882296	0.517936179	0.275419244	0.009939759

Table 3.9 CpGs showing >20% hypermethylation differences in CML-CCyR samples with respect to healthy control bone marrow (FDR < 0.01)

Illimina Probe ID	Chromosome	MAPINFO probe location	SNP	UCSC Gene Name	UCSC Gene position	UCSC CpG Island	BM Healthy Control	BM Remission	q-value
cg07772516	15	52107742		TMOD2	3'UTR		0.298042295	0.538672527	4.08E-05
cg12940181	2	145353012					0.365071486	0.596948606	0.00031329
cg13599613	9	130524902		SH2D3C	TSS1500; Body		0.347976497	0.586649021	0.000396669
cg13528479	6	36462140		STK38	3'UTR		0.564592686	0.826380048	0.000463867
cg04226002	11	113953462		ZBTB16	Body	chr11:113953620- 113953839	0.276157173	0.481974917	0.000481584
cg20357538	15	101777761		CHSY1	Body		0.455410314	0.661815016	0.000623707
cg22142142	15	90891614		GABARAPL3	Body	chr15:90895472- 90895990	0.540560542	0.855181497	0.000643057
cg02225720	17	3623472		ITGAE	Body	chr17:3626924- 3627839	0.219090841	0.42130859	0.000719507
cg10738119	7	75957040		YWHAG	3'UTR	chr7:75958865- 75959074	0.207069412	0.411776959	0.000721573
cg04319611	3	171784422		FNDC3B	5'UTR		0.255604333	0.519651813	0.000885517
cg17301248	12	112467203		NAA25	3'UTR		0.412526021	0.619015305	0.000898054
cg17164954	6	157345266	rs59609031	ARID1B	Body	chr6:157342060- 157343075	0.241446919	0.468513385	0.001023807
cg21671607	11	36616015		RAG2; C11orf74	5'UTR; TSS200		0.209134616	0.451538742	0.001053131
cg11327657	21	46388162		C21orf70	Body	chr21:46387051- 46387302	0.594969019	0.809333132	0.001066844

cg17478979	6	149772150		ZC3H12D	Body	chr6:149771836-149772855	0.234569114	0.486269294	0.001512182
cg07134666	6	29648400	rs17184563				0.600106996	0.928101037	0.001526968
cg26162326	7	75957061		YWHAG	3'UTR	chr7:75958865-75959074	0.253853181	0.475142037	0.001592928
cg01022219	18	13641735		C18orf1	Body	chr18:13641584-13642415	0.281776548	0.491370158	0.001728368
cg11747594	6	29648225					0.646953176	0.871058926	0.00179108
cg11383134	6	29648590					0.65040448	0.920280205	0.002086941
cg08354527	1	229252042					0.569491042	0.854006252	0.002100761
cg16885113	6	29648507					0.653450843	0.897290296	0.002107139
cg23296792	5	147777756		FBXO38	Body		0.358475507	0.592585582	0.002524069
cg24471254	7	100253792		ACTL6B	Body	chr7:100253782-100254150	0.464151063	0.664165093	0.002559004
cg20139683	12	133250118		POLE	Body	chr12:133249979-133250243	0.491315046	0.895612334	0.003111908
cg24536818	12	55371892		KIAA0748	5'UTR		0.197167481	0.412879582	0.003284581
cg03449857	6	29648623					0.605030686	0.854772078	0.003357043
cg20429104	18	74114570		ZNF516	Body	chr18:74114551-74114791	0.627283291	0.829814135	0.003497863
cg25390787	6	49681299		CRISP2	TSS200		0.606188724	0.820275196	0.003888569
cg06458771	6	29663039					0.55981261	0.781349877	0.004442285
cg08022281	6	29648345	rs2747429				0.528050572	0.743408827	0.004503425
cg16914890	2	242839615					0.608689531	0.842868913	0.004537632
cg21159128	1	54693933		SSBP3	Body		0.500849889	0.718182671	0.004641986
cg10648573	6	29648348	rs2747429				0.630604071	0.845786543	0.004831514
cg15570656	6	29648628					0.578515204	0.845922726	0.005230408
cg17404449	7	76221076		LOC100133091	Body		0.269182219	0.700838905	0.005509704
cg01561304	5	177913485	rs75977787	COL23A1	Body		0.493386715	0.710767716	0.006143343

cg24100841	6	29649024	rs75814207				0.567124552	0.800556637	0.006413401
cg20228636	6	29648525	rs2747431				0.61286365	0.851627706	0.006636652
cg19636627	6	29649084	rs3131847				0.536994564	0.824507627	0.007660037
cg14482998	6	111984370	rs9387025	FYN	Body		0.357044177	0.570338336	0.007781719
cg00588198	6	29648452	rs3129057				0.542183512	0.778116864	0.008088405
cg13835168	6	29648756	rs416560				0.520824877	0.755931011	0.009016258
cg09581911	12	4699232		DYRK4	TSS200		0.521087797	0.779415775	0.009825737

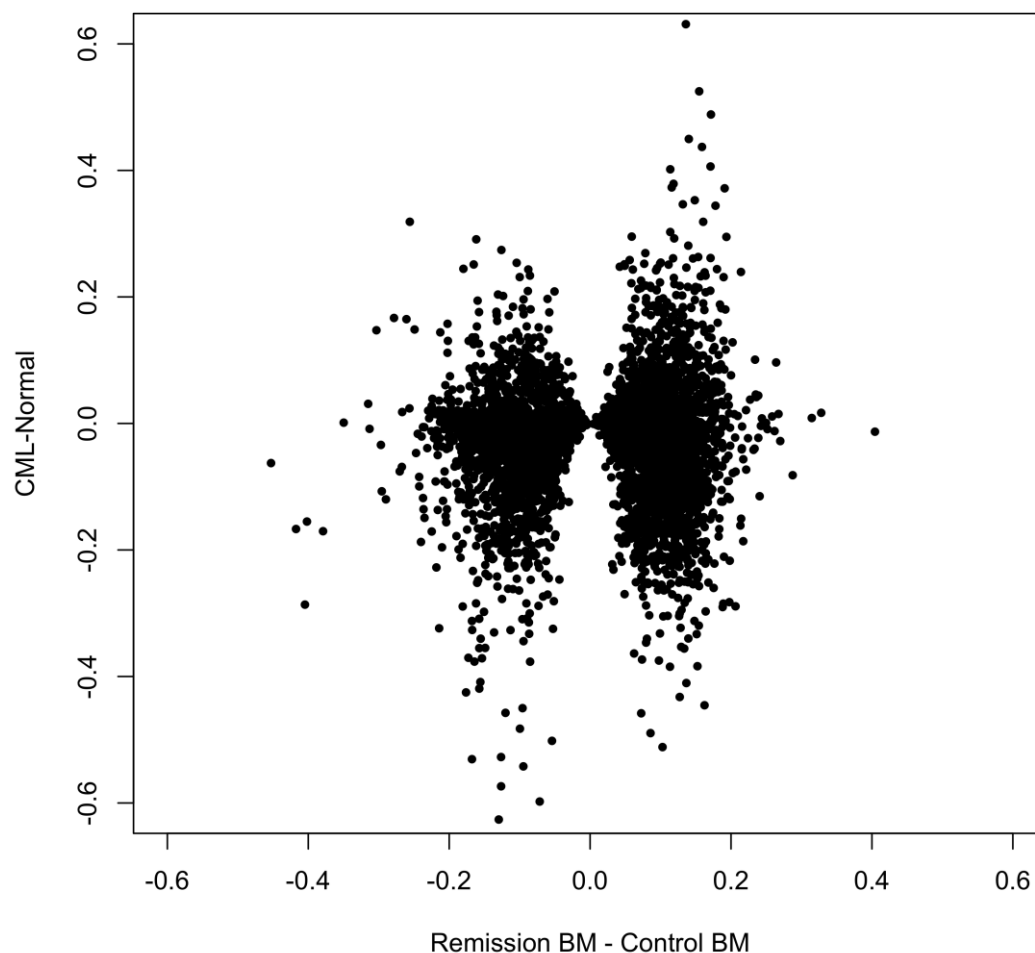


Figure 3.30 Scatter plot illustrating the beta values of the CML-CCyR bone marrow DMPs at diagnosis

The figure depicts graphically all 5,846 DMPs (adjusted p-value < 0.01) between CML-CCyR-BM and healthy control samples, (x -axis), plotted against the same probes for CML-CP differences, (y -axis). In CML-CP, 2,117 of these probes are also significantly different from healthy controls (adjusted p-value < 0.01) although directionality is not consistent with the CML-CCyR-BM DMPs.

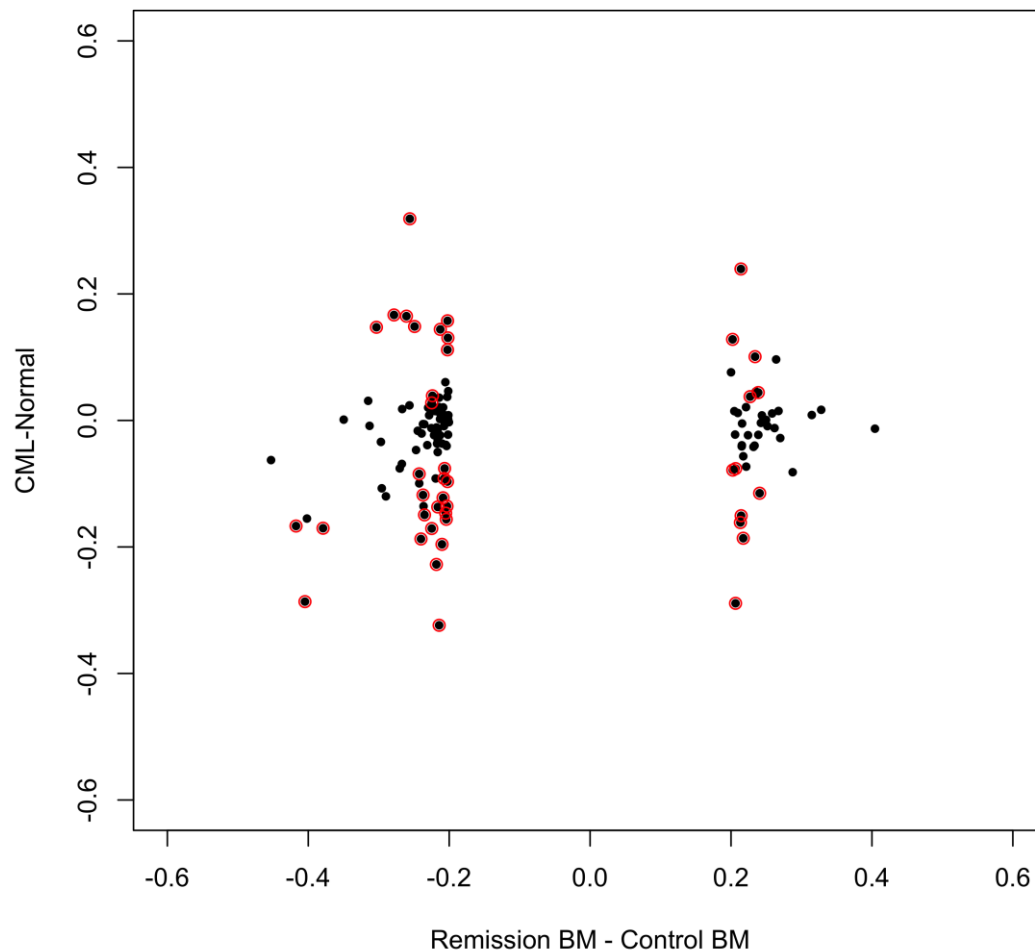


Figure 3.31 Scatter plot showing CpGs with >20% methylation differences in CML-CCyR-BM samples

A total of 137 CML-CCyR-BM DMPs have beta-value cut-off differences greater than 20% (adjusted p-value < 0.01). Such large differences are usually expected to be residual cancer marks however when these DMPs were applied to CML-CP samples only 34/137 (highlighted in red) are significant at diagnosis (adjusted p-value < 0.01) including 5 probes (4%) with > 20% differences i.e. either hyper- or hypomethylated (enrichment p-value = 0.001). Few of these probes show similar directionality between diagnosis and remission disease states suggesting that even those common to both could be the end result of different epigenetic processes.

3.5 Discussion

Despite the consistent presence of the *BCR-ABL1* fusion gene in Ph-positive CML, the clinical course of patient cohorts treated in the past with cytotoxic drugs and more recently with TKIs is heterogeneous. Some of this heterogeneity might be due to epigenetic changes acquired in the progeny of the original leukemia stem cell.

Although reports of aberrant DNA methylation occurring in BCR-ABL1⁺ CML have been described previously, genome-wide patterns of differential methylation have not been extensively investigated. Furthermore the relationship between the methylome, patient prognosis and therapy response is poorly understood.

The primary aim of this study was to gain a comprehensive overview of the DNA methylation landscape in CML haematopoietic stem and progenitor cells by comparing them to non-leukaemic, healthy equivalents using the HumanMethylation 450k BeadChip (HM450). In addition, the epigenomic profiles of CD34⁺ CML cells collected at the time of diagnosis were compared with paired, analogous profiles from patients who achieved complete cytogenetic remission on imatinib treatment, a methylation state that has previously never been reported. Finally, HM450 technology was investigated as a novel diagnostic tool for predicting methylation-based treatment response and outcome in CML.

This study identified and replicated (3,734 DMPs with >20% differences) consistent perturbations in the genome-wide methylation patterns of CD34⁺ cells from patients with newly diagnosed CML suggesting that epigenetic dysregulation could be involved in the leukaemic transformation process. In contrast to many other haematological malignancies, the CML DNA methylation signature was remarkably homogeneous; it differed from CD34⁺ cells from healthy donors and did not correlate with an individual patient's response to TKI therapy imatinib.

Genomic feature analysis revealed that CML-CP DMPs were significantly enriched in gene body and intergenic regions with an overall loss of methylation suggesting that in CML, DNA methylation is functionally relevant outside gene promoters. This is in agreement with Kulis *et al* who performed a large-scale analysis to characterize the

DNA methylome in normal B and CLL cells and highlighted widespread hypomethylation of gene body and enhancer sites during B cell differentiation and CLL evolution [294]. Aberrant DNA methylation in transcribed regions outside the promoter such as the gene body, could potentially disrupt the regulation of alternative splicing, retrotransposon elements, and other functional elements necessary in the maintenance of transcription [315, 316].

Using Marmalaid a meta-analysis tool, it was possible to compare the CML-CP methylation signature with a number of different tissue and tumour types including haematological malignancies from other studies. Here, it could be shown that in addition to retaining recognisable features with their non-malignant stem and progenitor cell counterparts, the CML-CP methylome shares striking similarities with other cancers and is made up of CML-unique, myeloid leukemia-specific and pan-cancer sub-signatures. Furthermore, some of the aberrantly methylated genes involved have previously been implicated in CML signaling pathways, including TGF-beta, Wnt, Jak-STAT and MAPK.

A variety of technologies are available for interrogating DNA methylation on a genome-wide scale including the HM450. Although less comprehensive than sequencing-based profiling methods e.g. whole genome bisulfite sequencing (WGBS) and methylated DNA immunoprecipitation-sequencing (MeDIP-seq), the HM450 array was selected for this study as a relatively cost-effective means of generating robust epigenomic data in two cohorts of CML patients on TKI therapy.

Furthermore, the data generated was supported by a well-defined analysis workflow that could be integrated with gene expression profiling (See Chapter 5) and compared with high quality HM450 experiments available from public repositories to further support data interpretation and analysis. This is one of the reasons why the HM450 is often selected as the platform of choice in epigenome-wide association studies (EWAS) [317].

Despite its many advantages and ongoing success, the HM450 does have some inherent design limitations, which cannot be resolved computationally. As a ‘genotyping’ microarray, the platform contains 485,512 probes (99% of all RefSeq genes) and interrogates 19,755 CGIs with some minor coverage of GGI shores, miRNA promoters as well as non-CpG sites [276]. Yet, overall, this represents less

than 2% of all CpG sites available for profiling in the human genome and means that some crucial regulatory regions such as enhancers are entirely neglected. Furthermore, by failing to detect allele-specific methylation, the Infinium assay underpinning the HM450 does not permit full elucidation of parent-of-origin effects or genomic imprinting.

At the time of CML-CCyR the majority of CD34⁺ haematopoietic stem and progenitor cells are by definition Ph⁻ negative. Therefore it is not surprising that the methylation patterns are different from CML-CP cells, the majority of which are Ph⁺. This supports the notion that BCR-ABL1 plays a key role in altering global methylation in CML but also poses additional questions about the differences found between CML-CCyR cells (bone marrow, unmobilised) versus healthy controls. Two possibilities exist; the most obvious relates to the pharmacodynamic effects of long-term TKI therapy on the methylome. Further studies are needed given that the length of exposure time to TKIs in CML is approaching 20 years for some patients and is likely to exceed this in the future.

Certainly in other chronic diseases, commonly used pharmacological agents can cause persistent epigenetic changes. For example both hydralazine, a vasodilator and procainamide, a sodium channel blocker inhibit DNA methylation through the interference of DNA methyltransferase activity, triggering a lupus-like autoimmune disease [318]. Furthermore, tamoxifen an anti-oestrogen used in the treatment of breast cancer has been shown to induce rapid, irreversible chromatin remodeling and cause changes in microRNA expression [319, 320]. One relatively simple and ethical method of testing this hypothesis in CML would be to perform serial bone marrow aspirates on patients recruited into therapy discontinuation studies. DNA methylation tends to be a dynamic epigenetic modification and over time, it might be expected that the methylation signature defining CML-CCyR would be ‘erased’.

One further intriguing explanation for the methylation differences observed between CML-CCyR and healthy controls is a yet undefined additional (epi)genetic lesion predisposing to the development of disease.

While the existence of genetic alterations is a common feature of tumour cells and well recognised, epigenetic lesions affecting genes are more difficult to define. The loss of the DNA repair activity of O6-methylguanine-DNA methyltransferase

(MGMT) in the absence of apparent genetic mutations has been observed in a number of cancers including gliomas, colorectal carcinomas, lymphomas and non-small cell lung tumours [321-323].

In fact, data suggests that transcriptional silencing of MGMT by promoter hypermethylation results in the accumulation of many G to A transitions of key proto-oncogenes including the proto-oncogene *K-ras*. Esteller and colleagues demonstrated that 71% of colorectal tumours with G to A mutations in *K-ras* are deficient in MGMT repair due to epigenetic silencing [324]. This is in contrast to other tumours with alternative non-G to A change *K-ras* mutations where the MGMT promoter remains unmethylated in 70% of cases. Further supporting the epigenetic lesion model, the detection of aberrant MGMT methylation in small adenomas precedes the appearance of K-ras mutations.

The timings corresponding to disease onset and clinical presentation/diagnosis undoubtedly vary between individuals, yet the consistency of the CML-CP DNA methylation signature was strikingly robust. Despite this, it was not possible to validate a TKI therapy response signal. As previously discussed, possible limitations relate to the study sample size, the geographical collection of cells (single, local centre vs multiple nationwide sites), the limit genomic coverage of the HM450 or the fact that a methylation signature based on TKI therapy response was not present. A larger study using the HM450 or a similar sized study utilising the newer MethylationEPIC BeadChip (interrogates over 850,000 methylation sites) or a less biased, whole-genome approach such as WGBS are potential options to investigate the issue in the future.

Although the possibility of G-CSF-induced methylation and gene expression changes was considered at the study design stage, the fact that both CML-CCyR and healthy control groups received G-CSF removed concerns of treatment bias. Furthermore, no published data was evident to account for the effects of G-CSF in CML patients on TKI therapy. This is likely because patients with leukaemia undergoing procedures such as autologous stem cell harvests are in the position where the serious risks associated with their diagnoses outweigh those from exposure to a short course of G-CSF. Data is however available for the effects of G-CSF on healthy volunteers as part

of bone marrow donor registries safety monitoring. Here, adverse risks associated with any drug carries considerable ethical concerns.

Nagler and colleagues considered the effects of G-CSF on peripheral blood lymphocytes in healthy stem cell donors and proposed that administration affected chromosome integrity resulting in cancer-associated asynchronous replication, which appeared to be related to altered DNA methylation [325]. However, the methylation data from this study did not support this finding and instead showed that mobilised CD34⁺ healthy control cells clustered consistently with age-matched unmobilised CD34⁺ bone marrow samples as well as CD34⁺ cord blood from TCGA data accessed through Marmal-aid.

For reasons that are unclear, G-CSF in combination with imatinib appears to affect the DNA methylation profiles of CML-CCyR samples. This complicated issue needs to be addressed further using a primary cell culture system since the process of autologous stem cell ‘back-up’ strategies are infrequently used in the era of TKIs and it would be unethical to give a healthy subject imatinib prior to peripheral blood stem cell donation.

In conclusion, perturbations in the genome-wide methylation patterns of CD34⁺ cells have been identified from patients with CML suggesting that epigenetic dysregulation could be involved in the leukemic transformation process. Conversely, the consistency with which these methylation aberrations occur provides no explanation for the clinical heterogeneity observed in this leukaemia. Whether any of these changes are of pathological significance remains to be determined in future studies.

4 The effects of *BCR-ABL1* conditional expression on the genome-wide DNA methylation patterns of U937p210BCR-ABL/c6 cells

4.1 Introduction

The work from Chapter 3 shows that a global methylation pattern exists in primary CD34⁺ CML-CP cells that are different from paired CML-CCyR cells and normal CD34⁺ counterparts. Importantly this was validated in an independent cohort. The fact that imatinib responders and non-responders have similar DNA methylation profiles that cannot easily be distinguished by DMPs actually strengthens one hypothesis that BCR-ABL1 is in some way able to cooperate with other pathways/mechanisms to cause universal, widespread epigenetic changes that can possibly contribute to leukaemia formation and progression.

To this extent, *in vitro* cell culture experiments were planned using U937p210BCR-ABL/c6 cells with conditionally inducible p210^{BCR-ABL} expression (Tet-On) as a CML-CP model [275, 326]. U937 human monocytic cell lines are derived from the pleural effusion of a patient with generalised histiocytic lymphoma [327]. In the induced state U937p210BCR-ABL/c6 (c6) cells display a complex hyperdiploid karyotype and feature one copy of the *BCR-ABL1* (b3a2) transgene on the short arm of chromosome 11 (11p) with protein tyrosine kinase levels sevenfold lower than those found in K562 cells.

To investigate the effect of BCR-ABL1 on centrosome aberrations in CML, Giehl and colleagues performed long-term culture experiments using U937p210BCR-ABL/c6 cells to model early CML-CP *in vitro* [326]. They showed that BCR-ABL1 expression induces centrosomal changes in a time-dependent manner. Partial reversibility of the proliferative phenotype was achieved through both BCR-ABL1 turn off and oncoprotein expression inhibition by Abl-directed siRNA. On administration, imatinib did not reverse the aberrant phenotype but actually inhibited

the normalizing effect of BCR-ABL1 expression turn off.

Selection of U937p210BCR-ABL/c6 cells was based on the considerable lack of other cell lines or animal engraftment models that reliably represent early CML-CP. The most commonly utilised CML-derived cell lines are immortalised CML-BC cells with highly proliferative sub-clones and excessive copy numbers of *BCR-ABL1* e.g. K562, LAMA-84 and KCL-22. An additional advantage that U937p210BCR-ABL/c6 cells offered over other non-conditional models relates to the early effects that BCR-ABL1 may have on the methylome at the time of oncogene induction. It was felt that these could be studied in a more controlled manner while reducing the risk of selecting for secondary mutations of other co-operating/key genes.

Furthermore, the HM450 is a human-specific DNA methylation array making it unsuitable for comparative analysis with model organisms such as mice, since a substantial proportion of probe sequences fail to align uniquely with the reference human genome during downstream mapping analysis [328]. This ruled out other widely utilised CML model systems such as Ba/F3, a murine interleukin-3 dependent pro-B cell line [11, 329].

4.2 Hypothesis

BCR-ABL1 is capable of initiating genome-wide DNA methylation perturbations in a CML-CP cell line model which is reversible upon abrogation

4.3 Aims

1. To establish the transforming potential of P210^{BCR-ABL1} on DNA methylation in the conditional U937p210BCR-ABL/c6 cell line
2. To define the conditional expression of *BCR-ABL1* at the mRNA and protein level in U937p210BCR-ABL/c6 cells to better understand any detectable DNA methylation changes
3. To examine the effects of imatinib (IM) therapy on the abrogation of any BCR-ABL1-mediated DNA methylation changes in U937p210BCR-ABL/c6 cells
4. To attempt to translate any relevant findings to primary CD34⁺ CML cells and in so doing, explain the DNA methylation changes observed in Chapter 3.

4.4 Methods and overall experimental design

The c6 clone was propagated, as described in Chapter 2 Materials and Methods. Untransfected U937 cells were used as a negative control. Initially, U937 and c6 cell lines were tested with respect to their identity, karyotype and proliferation rate. Cells were cultured for 7 days following BCR-ABL1 induction by doxycycline and triplicate cultures were performed for all experimental conditions.

To assess the antagonistic influence of TKIs on any possible observed BCR-ABL1-mediated DNA methylation induced changes; some groups were exposed to imatinib for up to 14 days either from day 0 or 7 days into the culture experiments. Cells from all the conditions were harvested and DNA was extracted for HM450 analysis. In addition, RNA and protein lysates were isolated for quantification of *BCR-ABL1* transcript levels by real-time PCR (QT-PCR) and for western blot analysis respectively.

Prior to culture, cells were tested for transgene inducibility. Densitometric analysis of BCR-ABL1 protein showed that the optimal induction of p210^{BCR-ABL} tyrosine kinase activity by doxycycline (Dox) stimulation in 'c6-On' cells occurred after 72 hours of treatment (Figure 4.1). Cells were therefore treated with Dox on day -3 and analysed on day 7 (approximately two cell cycle rounds) allowing for potential methylation changes to take place whilst ensuring optimal cell viability. Additionally, c6 and U937 cells were exposed to imatinib either from day 1 or day 7 and cultured for a further 7 to 14 days respectively to test whether any methylation changes detectable were reversible upon treatment. The culture conditions of cell groups are listed in (Table 4.1 Culture conditions of U937p210BCR-ABL1/c6 cells) and the experimental design is summarised in (Figure 4.2)

Table 4.1 Culture conditions of U937p210BCR-ABL1/c6 cells

1. U937+Dox (harvested d7)
2. c6+Dox (harvested d7)
3. c6-Dox (harvested d7)
4. U937+Dox+IM (IM added d1, harvested d7)
5. c6+Dox+IM d7 (IM added d1, harvested d7)
- D1. U937+Dox+IM (IM added d7, harvested d21)
- D2. c6+Dox+IM (IM added d7, harvested d21)
- D3. c6-Dox+IM (IM added d7, harvested d21)

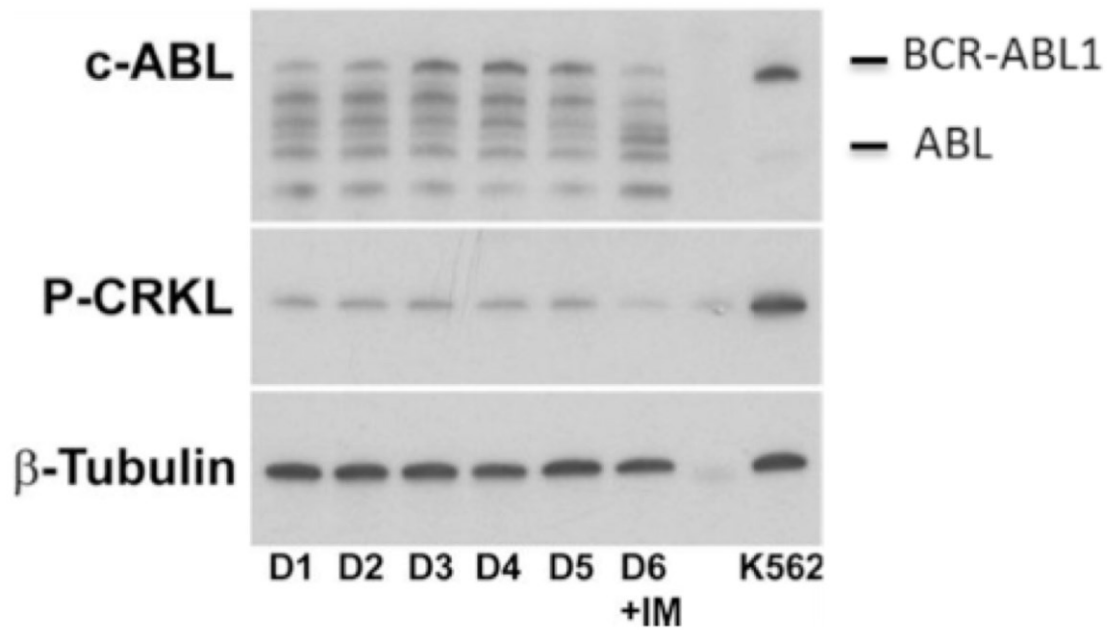


Figure 4.1 Western blotting showing BCR-ABL1 induction in c6 cells upon doxycycline stimulation

Optimal production of BCR-ABL1 protein was reached on day 3 (D3) following induction with doxycycline. Addition of imatinib resulted in oncoprotein abrogation. K562, a human-derived CML blast crisis cell line was used as a positive control for p210^{BCR-ABL1}. CRKL is a signal transducing adapter protein and a downstream substrate of BCR-ABL1. Tyrosine-phosphorylation of CRKL is used as a surrogate indicator of BCR-ABL1 activity.

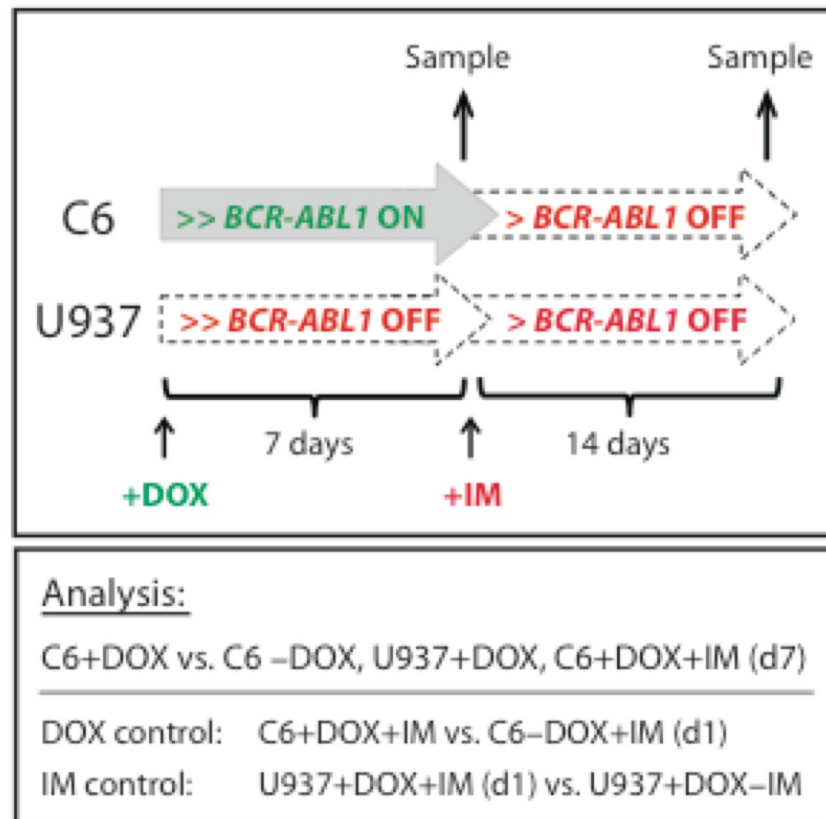


Figure 4.2 Experimental design

The purpose of this work was to compare the DNA methylation profiles of c6-On (samples 2 A-C) with both U937+Dox (samples 1 A-C) and c6-Off (3 A-C). The effect of IM treatment could be investigated between c6-On (2 A-C) and c6-On+IM (5 A-C or D2 B-C, note D2 A culture was infected and eliminated prior to methylation profiling) using U937+Dox (1 A-C) and U937+Dox+IM (4 A-C) as controls. The initial abrogation of BCR-ABL1 induction could be investigated by comparing c6-On (2 A-C) and c6-On+IM (5 A-C) (groups in brackets correspond to culture conditions from Table 4.1 Culture conditions of U937p210BCR-ABL1/c6 cells).

4.5 Results

4.5.1 BCR-ABL1 induction by QT-PCR

Following the addition of doxycycline to c6 cells, BCR-ABL1 transcript ratios on day 3 by qRT-PCR were $91.8 \pm 1.8\%$ ($n=6$) (Figure 4.3). This was somewhat higher but still comparable to 3 primary CD34⁺ CML samples from newly diagnosed patients where the average ratio was 62% ($p > 0.05$). As expected, p210^{BCR-ABL} protein was undetectable in bcr-abl-negative U937 cells. However, the p210^{BCR-ABL} cDNA construct under Tet-On promoter control demonstrated considerable “leakiness”, since c6 cells without doxycycline induction i.e. c6-Off had a $62.3 \pm 4.9\%$ ($n=6$) BCR-ABL1/ABL transcript ratio.

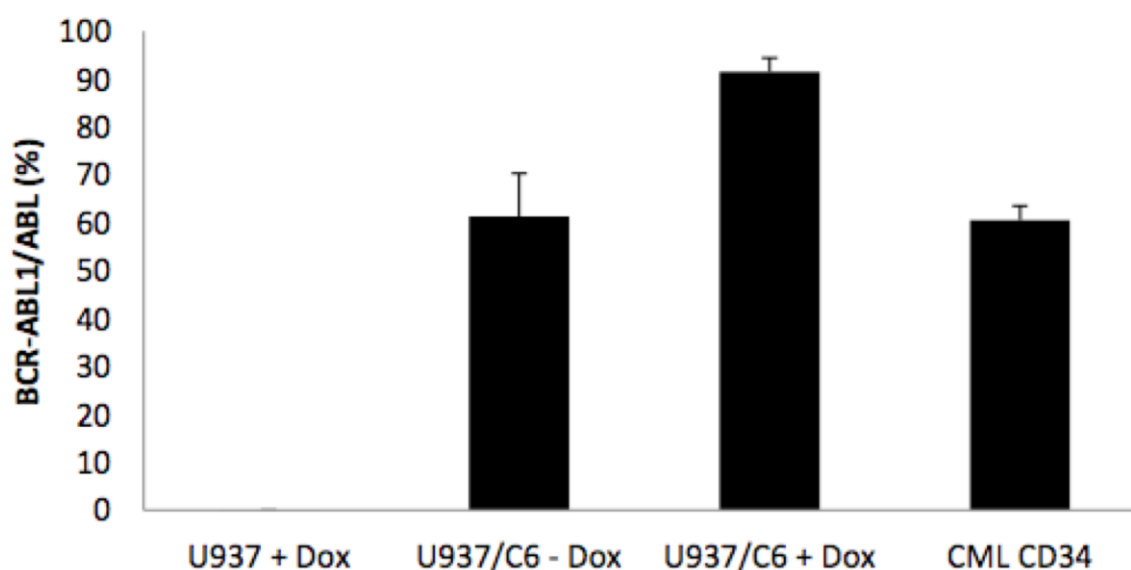


Figure 4.3 *BCR-ABL1* mRNA expression ratio (%) by QT-PCR

Tet-On promoter ‘leakiness’ is evident in c6-Off cells (U937/c6-Dox) where BCR-ABL1 transcripts are detectable at levels comparable with CD34⁺ CML primary cells. This finding was unexpected.

4.5.2 BCR-ABL1 protein induction by Western blotting

To see how this translated into protein expression, Western blotting was performed using an anti-c-abl antibody (OC20) to detect BCR-ABL1 (MW: 210kDa) and c-ABL (MW: 145kDa). Also, BCR-ABL1 protein activity was evaluated with both an anti-p-Tyrosine antibody (4G10), and anti-p-CrkL antibody where levels of CRKL phosphorylation can be used as an alternative, indirect marker of p210^{BCR-ABL} function [330, 331]. Anti-Lamin B1 was used as a nuclear loading control.

As expected, U937 cells demonstrated no BCR-ABL1 expression (c-Abl) or activity (P-Tyr and p-CRKL) under any culture conditions (Figure 4.4). c6-On cells showed high phospho-BCR-ABL1, which was almost entirely abrogated by treatment with IM regardless of whether exposure was on d1 or d7 (2.8-fold decrease by densitometry). In all the IM treated c6 groups, weak but constitutive activity of p-CRKL existed possibly due to CRKL associating and interacting with other protein complexes e.g. SOS and PI-3K.

In parallel to the unanticipated BCR-ABL1 mRNA expression in c6-Off cells, Tet-On system leakiness was evident at the protein level, which translated to a 3.5-fold lower densitometric value when compared with c6-On. Similarly, activity measured by phospho-BCR-ABL1 was detectable in the c6-Off state but at 2.1-fold lower levels than c6-On. Furthermore, in the absence of IM, all c6 groups demonstrated elevated levels of p-CrkL activity from baseline with c6-On cells showing a 1.3-fold increase by densitometry compared with the c6-Off state.

Overall, the results showed that the U937p210BCR-ABL/c6 model was suboptimal due to leaky expression of the Tet system under non-induced conditions. However it did provide the only opportunity for studying the *in vitro* relationship between p210^{BCR-ABL} activity, DNA methylation and IM treatment in the context of CML-CP. To characterise any induced DNA methylation alterations and identify DMRs, cell triplicates from each of the culture conditions were spun down, DNA was extracted, bisulfite converted and hybridized to the HM450 array according to the Methods described in Chapter 2.

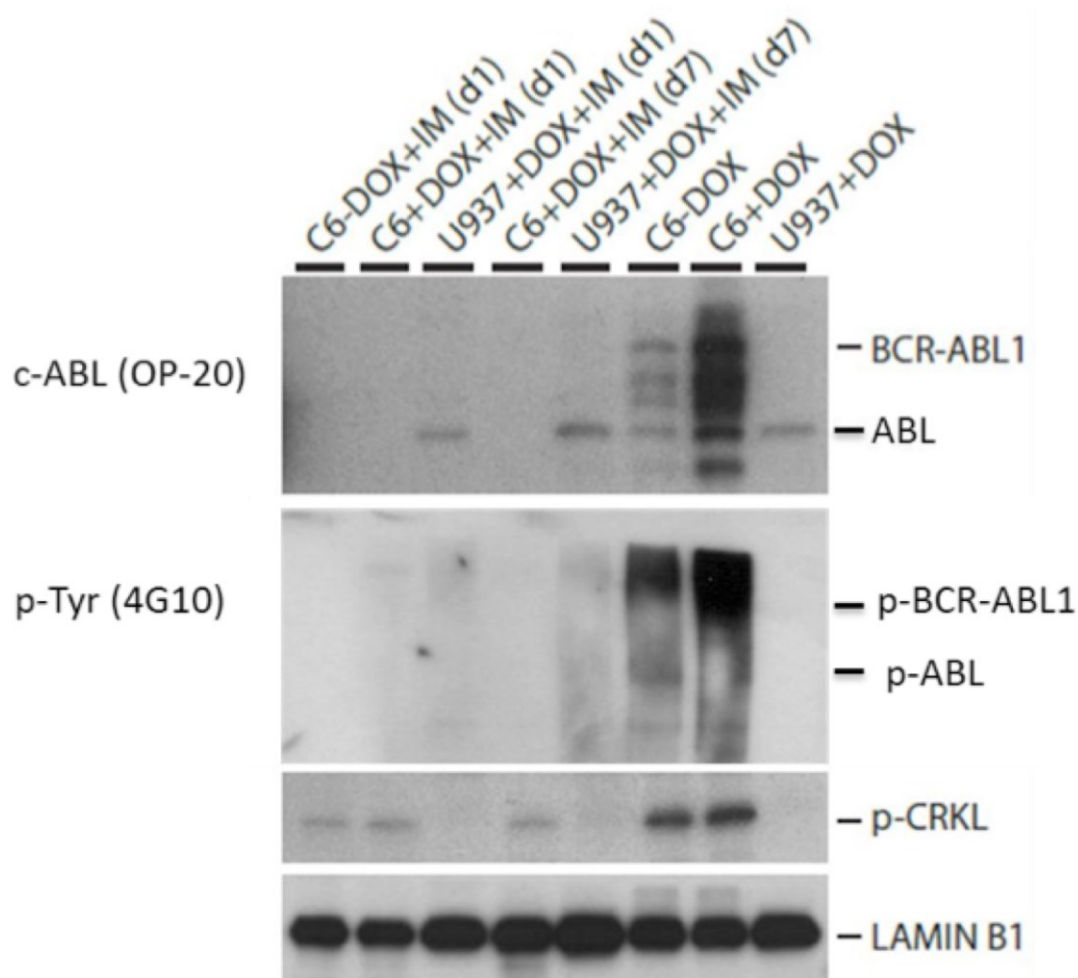


Figure 4.4 Western blotting showing BCR-ABL1 (OC20), p-BCR-ABL1 (p-Tyr [4G10]) and p-CrkL protein expression. Lamin B1 provided a loading control.

Cell line conditions are indicated (DOX= doxycycline, IM= imatinib).

4.5.3 DNA methylation using HM450

The anticipated outcome from DNA methylation studies was that a set of DMPs would be identified between cells lacking or possessing functioning BCR-ABL1 kinase activity. In this way it would be possible to associate the oncogene with any consequential DNA methylation events. These DMPs were not expected to duplicate those separating CML-CP patients from healthy controls, merely reflect them, in the knowledge that the genetic backgrounds of the chosen cell lines are markedly different from CD34⁺ primary cells.

4.5.3.1 DNA quality

Following visualisation on agarose gels, all DNA samples showed high integrity with no evidence of RNA contamination. Genomic DNA samples (500 ng) by Qubit quantification were bisulphite converted as previously described and the conversion efficiency was checked by qPCR.

4.5.3.2 HM450 Processing, QC and analysis

Samples were randomly assigned to 96-well plates together with further human primary CML samples from a different experiment. In addition, some intra-batch sample duplicates were run. Batch effects were less of a concern during this experiment since all samples were processed and hybridized together.

All samples were processed using the protocol previously described for HM450 array hybridisation, which was performed at the UCL Genomics facility. Sample-independent and -dependent QC (data not shown but available) demonstrated satisfactory control probe performance. Data QC, processing and analysis was performed according to the methods set out in Chapter 2, Materials and Methods.

Data filtering was performed on .idat files. Since the experimental samples originated from cell lines, the sex chromosome constitution was the same for each. U937 cells are derived from a male patient although the Y chromosome is deleted and the X chromosome is usually duplicated (i.e. X, +X, -Y). For the purpose of this work, sex chromosomes were not filtered at the pre-processing stage, neither were SNP-associated probes excluded. However, those demonstrating cross-hybridisation were removed as standard. While the correct positioning of the samples could not be checked against sex, interspersed primary samples on the plate, were used as a crude

point of cross-reference. The final number of probes used for subsequent analysis was 461,075.

As a first comparison of global methylation patterns, β -values (DNA methylation values between 0 and 1, representing 0-100% methylation as previously outlined) across all samples in the U937 and c6 cell dataset were derived from the filtered .idat files.

4.5.3.3 Identifying DMPs between U937 and c6 cells

Unsupervised hierarchical clustering of the variable probes in this dataset (those showing β -value differences of 10% or more) highlighted the genome-wide DNA methylation differences between the two cell lines with c6 cells looking closer to CD14⁺ monocytes than the U937, from which they were derived. Most concerning, culture conditions for each group, had no effect on the global methylation patterns e.g. c6-On cells are indistinguishable from c6-Off and similarly, imatinib treated samples could not be identified as a subgroup (Figure 4.5).

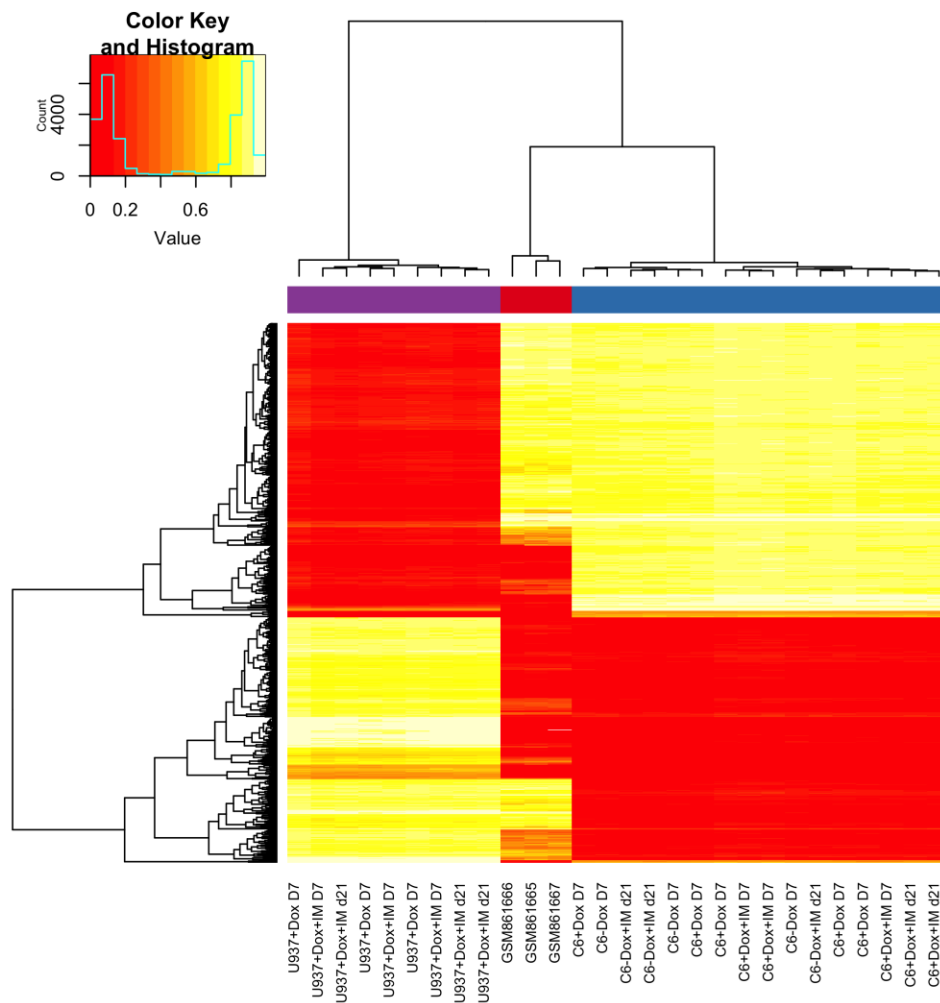


Figure 4.5 unsupervised hierarchical clustering (scaled by row) of probes showing variable DNA methylation β values of more than 10% in U937, c6 and CD14+ human monocytes.

(Dox= doxycycline, IM= imatinib, GSM86= CD14+ monocyte)

Samples separate by cell type without any evidence that culture condition affect methylation. c6 cells (blue) display methylation patterns across probes that are closer to CD14+ cells (red) primary human monocytes despite being derived from U937 cells (purple) and containing a BCR-ABL1 inducible transgene. Darker shade represents increasing degree of methylation (i.e. yellow, $0 < \beta < 0.5$ < red, 1).

These differences between U937 and c6 rather than the culture conditions/BCR-ABL1 status was maintained following unsupervised clustering on the 100 most highly variable DMPs (Figure 4.6 and Figure 4.7).

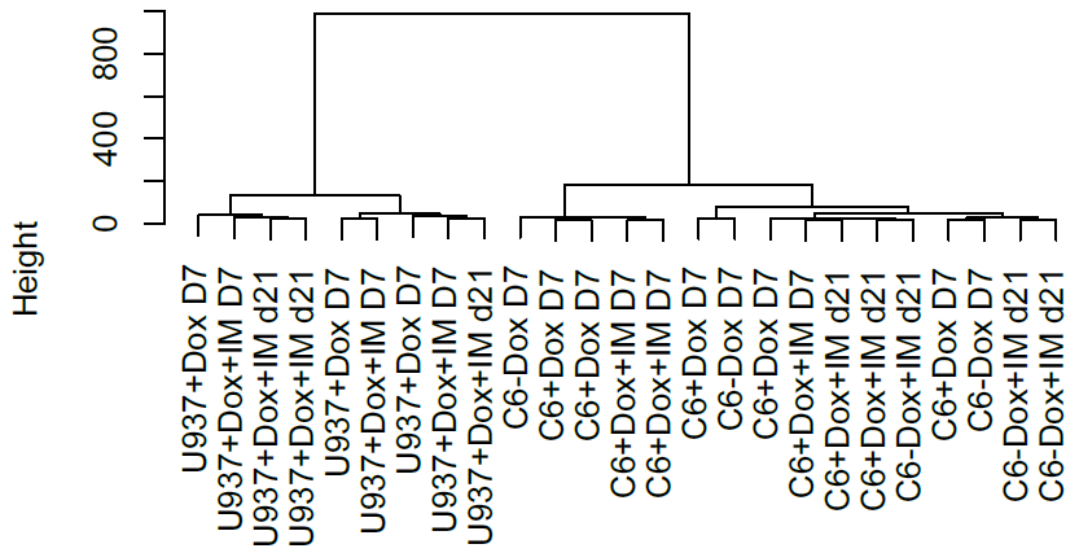


Figure 4.6 Unsupervised hierarchical cluster dendrogram on the top 100 DMPs

The greatest β -value differences separate samples on the basis of cell type rather than experimental conditions. Among c6 cells, subgroup variations are not statistically significant as evidenced by the lack of segregation into smaller clusters e.g. no differences between c6+Dox (c6-On) D7 versus c6-Dox (c6-Off) D7 groups. (Dox= doxycycline, IM= imatinib).

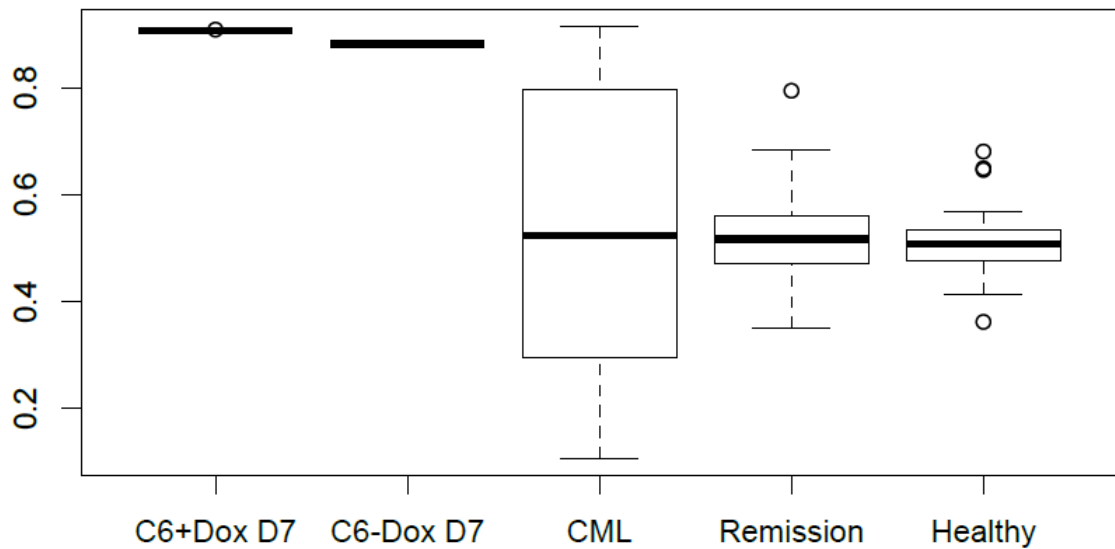


Figure 4.7 Comparison of the average DNA methylation levels (β -value) (y-axis) across the top 100 c6-On (+Dox) and c6-Off (-Dox) DMPs.

Regardless of BCR-ABL1 status, few detectable differences between the c6 sub groups are apparent, with all samples showing marked hypermethylation. CML-CP (CML), CML-CCyR (remission) and healthy CD34+ bone marrow control (healthy) samples were included for comparative purposes. The median methylation values across these probes for primary cell samples are almost identical (β -value~ 0.5) but with a greater degree of heterogeneity seen in the CML-CP group, a finding that likely reflects ‘cancer’ changes. (Dox= doxycycline).

4.6 Discussion

Two issues emerged in the early stages of analysis. The first concern was that the epigenetic backgrounds of U937 and c6 cells were considerably different despite c6 clones being derived from U937 cells. These ‘cell type’ methylation differences appeared to be far greater than the BCR-ABL1-driven leukaemia effects that were being investigated.

Induction of the *BCR-ABL1* oncogene by doxycycline resulted in no discernable differences in DNA methylation patterns in the c6-On cells. The fact that c6-Off cells were leaky at the mRNA and protein level, emphasized the value that analogous epigenetic backgrounds would have provided by allowing comparisons between U937+Dox D7 and C6+Dox D7 as an alternative.

The second problem related to this was that in c6-On cells as well as leaky c6-Off samples, BCR-ABL1 abrogation by imatinib appeared to have undetectable consequences on the methylation status. Although somewhat unexpected and in complete contrast to the anticipated outcome indicated from the primary cell data in Chapter 3, this was likely to be due to the overall ‘cell type’ background differences, which were substantially altered to begin with. The value of further analysis using U937/c6-On/Off modeling as a tool for defining the effects of BCR-ABL1 on DNA methylation in CML-CP was unlikely to be informative. For this reason, plans for additional analysis were discontinued.

Recently, Amabile and colleagues, used cellular reprogramming to generate leukaemia-induced pluripotent stem (LiPS) cells from K562 and KBM-7 (CML blast crisis) cell lines as well as human primary CML bone marrow cells [332]. Utilising Reduced Representation Bisulfite Sequencing (RRBS), the group showed that LiPS clones gained DNA methylation patterns that were comparable with but not identical to ES cell lines and CD34⁺ iPS, supporting the hypothesis for cell-of-origin-specific ‘epigenetic memory’ in iPSCs [333, 334]. Notably, reprogrammed primary CML cells still retained the *BCR-ABL1* oncogene. This finding is consistent with previous work by Carette and others who, in generating and characterising LiPSCs from the aforementioned KBM-7 human-derived CML cell lines, elegantly highlighted that

although the parental cells were entirely oncogene addicted to BCR-ABL1, reprogrammed cells lost this trait and were completely resistant to imatinib [335], a result that was attributed to a specific epigenetic undifferentiated cell state. Similarly, Amabile *et al*, showed that reprogramming restores myeloid differentiation of CML cells even in the presence of BCR-ABL1 by reactivation of the critical differentiation-associated gene CEBPA [332].

Most relevant to the work in this thesis, the same group sought to investigate the functional relationship between the presence or absence of *BCR-ABL1* and DNA methylation in CML. Using a BCR-ABL1 conditional transgenic mouse model [336], they were able to show that induced mice developed moderate but significant CGI hypermethylation, which crucially was reversible upon *BCR-ABL1* abrogation at the genetic (repression with tetracycline or inhibition by IM) or epigenetic level (by 5-Azacytidine, AZA). This convincingly highlights that a single oncogene is capable of initiating genome-wide DNA methylation perturbations in a well-defined murine CML model.

Among the 500 most significant leukaemia-specific DMRs associated with gene promoters, most were hypermethylated with respect to the derived LiPS cells and significantly enriched for functional categories associated with development, differentiation and signaling. Furthermore, overlapping these DMRs with epigenomic features showed they were enriched for specific histone marks such as enhancer-associated H3K4me1 and promoter-associated H3K4me3.

In conclusion, although the choice of cell line used in this study was carefully selected based on the need for a human cell line with inducible P210^{BCR-ABL1}, the experiments in Chapter 4 failed to reveal the transforming potential of this oncogene on the cellular DNA methylation state likely due to its leaky On/Off system. However, recent work by Amabile and colleagues suggests a strong association does exist and further work using alternative cell lines and animal models is justified and necessary.

Future strategies to better understand global methylation changes in CML might involve engineering lentiviral delivery of clustered regularly interspersed short palindromic repeats (CRISPR)-Cas9 to generate the *BCR-ABL1* rearrangement using single guide RNA (sgRNA) targeting *BCR* and *ABL1* genes in a human cell line. Alternatively, a conditional p210 BCR-ABL1 transgenic animal model could be used

to induce/inhibit oncogene expression followed by BS-Seq in order to avoid the need for a human array based technology like the HM450.

5 Investigating the impact of aberrant DNA methylation on gene expression in CML

5.1 Introduction

Whole-genome expression profiling is a powerful tool for studying transcriptomics, the science of measuring the collection of genes that are expressed or transcribed from genomic DNA, including mRNAs, small RNAs and non-coding RNAs (ncRNAs).

From a translational science perspective, transcriptomics makes it possible to classify or risk-stratify disease in any number of ways depending on the biological features or pathological processes being investigated.

There have been a number of reports utilising gene expression profiling in the context of hematological malignancies that have been validated and used successfully to accurately predict for response to therapies and guide disease outcomes in leukemia [337, 338].

In CML, the main challenges associated with gene expression profiling (GEP) studies relate to the lack of consistent results between investigators despite robust laboratory methodology. The reasons for this are multi-factorial and include challenges such as biological and experimental heterogeneity.

One source of biological heterogeneity includes cell type selection e.g. peripheral blood mononuclear cells (PBMCs) versus bone marrow CD34⁺ cells. From a practical perspective, the more readily available and accessible cells are, the greater the potential for a technology such as GEP to be adapted for use in clinical practice. However, at the time of CML-CP presentation, a newly diagnosed patient displays excessive numbers of myeloid cells at various stages of maturation in the peripheral blood. Under these circumstances, measurable transcripts would potentially reflect this and complicate the interpretation of other subtle signatures such as those related to treatment response, particularly when study sample sizes are small. Furthermore as previously highlighted in this work, the definitions of response to imatinib have evolved over time, complicating comparisons between studies to a greater extent.

Similarly, technical variation associated with sample preparation, the choice of profiling platform (e.g. cDNA microarray versus oligonucleotide arrays) and the bioinformatic pipelines used in data analysis can contribute to differences in derived gene sets.

Nine groups to date have carried out GEP on CD34⁺ progenitor cells in CML [339-347]. Yong and others collected CD34⁺ cells at diagnosis from two distinct groups of CML patients, one of whom transformed to CML-CP within 3 years of diagnosis while the other remained stable in CML-CP for at least 7 years [342]. A total of 20 genes were differentially expressed among the patients with the aggressive disease phenotype compared to those with a more indolent disease course. Low expression of CD7 and high expression of proteinase 3 or elastase were predictors of longer survival in CML.

Zheng and colleagues used a 14,500-gene oligonucleotide array to compare the gene expression differences between peripheral blood CD34⁺ CML-CP (n= 11) and CML-BC (n= 9) cells [343]. In total, 114 genes were differentially expressed between the two groups with 34 of these indicating at least a two-fold difference in expression. Among the patients with CML-BC, SOCS2, CD52, CD34, HLA-DRA, -DRB and -DPA were significantly upregulated. Conversely, the transcription factors JUNB, FOS, KLF4 and CEBPB were underexpressed.

In addition to analyses based around the identification of biomarkers for disease progression, response to TKIs or prognosis, a few studies have address the gene expression differences between CD34⁺ CML cells and their normal counterparts.

In one study, Kronenwett isolated CD34⁺ cells from the BM of CML patients and normal individuals [339]. Using custom-designed cDNA arrays covering 1,185 genes, the study identified 158 of these that were dysregulated in CML relative to healthy controls. Genes associated with proliferation and cell cycle progression, including CDK4, as well as downstream effectors of BCR-ABL1, such as the RAS family members and *JAK2* were among the notable genes upregulated in CML cells. Conversely, genes significantly downregulated in CD34⁺ CML cells included the cyclin-dependent kinase inhibitors (CDKI), *CDKN1A/p21* and *CDKN2D/p19* and a number of chemokine- and cytokine-related genes such as *IL10*, the chemokine receptor 4 (*CXCR4*), G-CSF receptor (*CSF3R*) as well as platelet-derived growth

factor- α (*PDGFA*). Interestingly, the study also examined expression variations between bone marrow and peripheral blood-derived CML CD34⁺ cells. Both groups shared close expression homology with their normal counterparts and displayed no additional differences based on the source of cell collection.

Gene expression signatures of bone marrow-derived primitive CML stem (CD34⁺ CD38⁻) and progenitor cells (CMPs, GMPs and MEPs) were investigated by Bruns and colleagues [346]. This work highlighted the expression similarities between CML CMPs and leukaemia stem cells (LSC) and demonstrated significant downregulation of a number of cell adhesion proteins and chemokine receptors including *CXCR4* between CML LSCs and normal HSCs suggesting defective adhesion and migratory capabilities in the former.

Finally, one study to date has considered gene expression differences between CD34⁺ CML-CCyR samples and healthy bone marrow donors [340]. Neumann and colleagues failed to detect any expression differences across the two groups, findings that may not be definitive and require elucidation given the small sample sizes represented and the fact that imatinib has been shown to have activity against wild-type tyrosine kinases [45].

Few studies in the field of CML including those discussed so far have considered the role of epigenetic regulation, specifically DNA methylation on the various derived gene expression signatures. The data that exists focuses on targeted genes under specific conditions [348, 349].

Therefore, the work presented in this Chapter was designed to drawn on the findings from Chapter 3 and investigate whether aberrant DNA methylation in CD34⁺ CML cells affected the corresponding gene expression patterns.

5.2 Hypothesis

DNA Methylation in CML is a crucial epigenetic modification at the gene transcription regulation level. Alterations at promoter-associated CGIs are anti-correlated with the expression of key genes in CML.

5.3 Aims

The aims of this chapter are

1. To investigate gene expression pattern differences between CD34⁺ CML cells versus their healthy CD34⁺ counterparts.
2. To identify the genes and pathways involved in the maintenance of CD34⁺ CML cells.
3. To investigate a possible role for aberrant DNA methylation in regulating gene expression during CML by determining the correlation between DNA methylation and gene expression in CD34⁺ CML cells.

5.4 Methods

5.4.1 The Illumina HumanHT-12 v4 Expression BeadChip (HT-12)

For full details, please see Chapter 3, Materials and Methods. Briefly, Illumina's Direct Hybridization Assay protocol was used with the HT-12 platform (12 sample BeadChip format) according to the manufacturer's instructions [350].

A BeadChip comprises a series of rectangular BeadArray sections or 'strips' on a slide. The BeadArray is made up of randomly positioned silica beads. Each bead holds many thousands of copies of a particular 50mer oligonucleotide probe sequence that is randomly duplicated more than 15 times on additional beads across each array [351]. This provides robust internal technical replication and minimises positional artefact. Overall, the array contains more than 47,000 probes and targets over 31,000 annotated genes per sample, derived from sources including the National Center for Biotechnology Information Reference Sequence (NCBI) RefSeq Release 38 (November 7, 2009).

Following RNA isolation from pre-selected CD34⁺ cells, the quantification and integrity of purified DNase-treated RNA samples were assessed using 4200 TapeStation system (Agilent). RNA was amplified, biotin-labeled, and hybridised to HT-12 arrays (Figure 5.1). Slides were scanned using Illumina iScan system where fluorescence emission by Cy3 is quantitatively detected. All array processing was performed in the core Genomics Facility, King's College London.

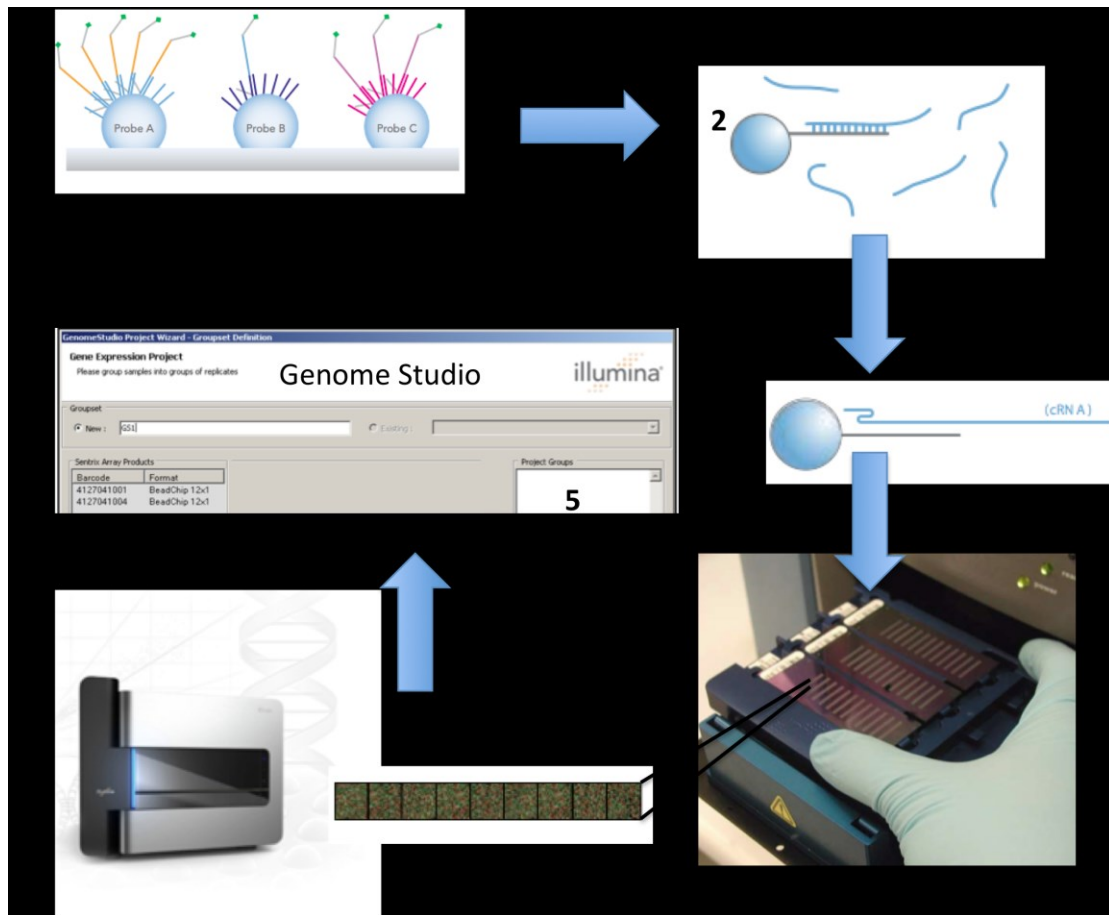


Figure 5.1 Illumina whole-genome gene expression direct hybridization assay

Following first- and second-strand reverse transcription plus an additional single *in vitro* amplification to incorporate biotin-labeled nucleotides, the cRNA strand is hybridised overnight to the HT-12 BeadChip containing complementary gene-specific sequences (1). This is followed with washing to remove unhybridised sample (2). Probe-bound cRNA (3) is loaded for differential detection scanning on the Illumina iScan (4) after which HT-12 data is uploaded as raw files or extracted by GenomeStudio/BeadStudio software for downstream analysis.

Adapted from [350]

5.4.2 HT-12 data processing and analysis

5.4.2.1 Data quality control and normalisation

As previously outlined (Section 2.2.11.2), acquired raw/bead-level data were analysed with the Bioconductor package *lumi*. *Lumi* is based on the statistical programming language, R (version 3.3.0) with functions for processing and annotating BeadArray expression data [352] (Figure 5.2).

Intrinsic to its design, the HT-12 contains many sample-independent as well as dependent quality control (QC) probes. Independent metrics relate to hybridisation, washing and staining performance (e.g. biotin controls) while sample-dependent ones measure sample quality and labeling (e.g. housekeeping gene controls). Prior to downstream analysis and normalisation, data were filtered on internal control probes and background intensity; those with a detection confidence below 80% in more than 20% of the samples were removed from further analysis. Probes with a signal-to-noise ratio (SNR) below 10 (calculated as a fraction of the 95th and 5th quantiles of all pixel intensities on the image analysis) were also excluded along with unreliable probes that either failed to demonstrate unique genomic locations or mapped to non-transcribed regions.

Raw data were transformed using the variance stabilising transform (VST) function and normalised by Robust Spline Normalisation (RSN). This algorithm combines the advantages of quantile normalisation including computational efficiency and preservation of gene rank order with the continuous transformation benefits of loess normalisation [353].

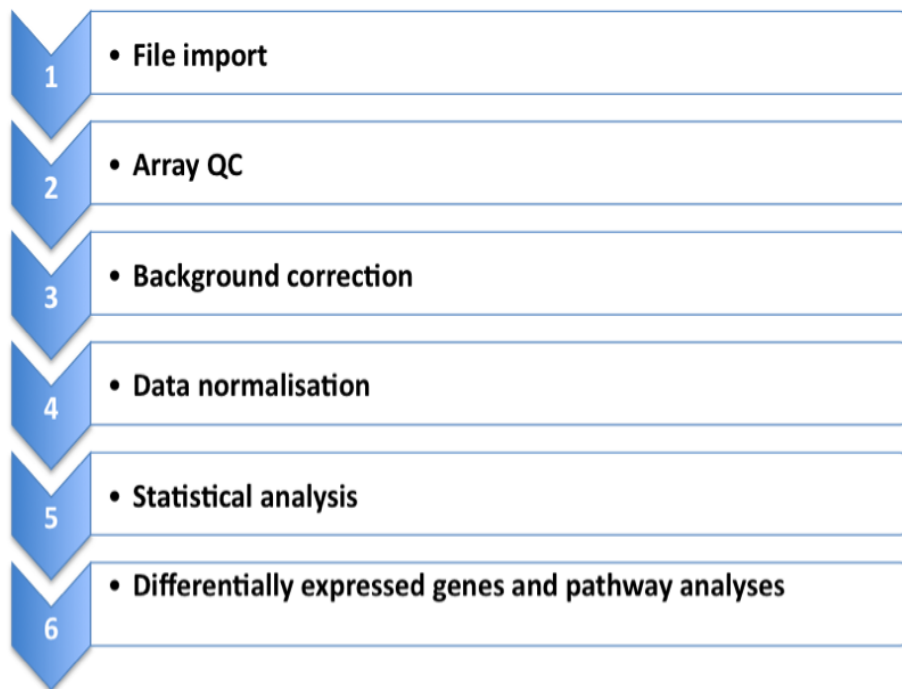


Figure 5.2 HT-12 array pre-processing steps

5.4.2.2 Statistical and differential expression analysis

Differentially expressed genes between CML-CP and healthy control samples were identified based on a moderated t-test (moderated standard errors across genes by Bayesian model) using an alternative Bioconductor open source tool, *limma* [282]. Multiple hypotheses testing were controlled using the Benjamini-Hochberg method [278]. Genes that satisfied the significance criteria of a false discovery rate (FDR)-adjusted P-value < 0.01 and log ratio higher than 1.2-fold were further analysed for the enrichment of gene ontology (GO) [287] and subjected to KEGG pathway analysis using the full list of genes included on the array background [286].

5.4.2.3 Correlation between DNA methylation and gene expression

The correlation between β -value and transformed gene expression data were evaluated using the Spearman rank-order correlation coefficient (Spearman's correlation), a nonparametric measure of the strength and direction of association. Statistical significance of differentially methylated positions was calculated by permuting the data 10,000 times and comparing the correlation coefficient in the unpermuted data to the permuted co-efficient. In each dataset, the permuted P-values were once again adjusted for multiple testing using the Benjamini-Hochberg approach for controlling multiple testing.

5.5 Results

5.5.1 Patient samples

The samples used to generate gene expression data were from the Hammersmith cohort (Section 3.4.1). In contrast to the DNA methylation work, it was not possible to validate the findings in an independent cohort. The reasons for this included the fact that there was no remaining material available for use from the SPIRIT 2 validation cohort while time limitations (including application for ethical approval), prevented access to alternative CML biobanks. This will be discussed further at the end of this chapter.

Following preprocessing and limited by sample availability, RNA from 28 of the paired CML-CP-R samples [CML-CP-R6 removed due to RNA integrity number (RIN) 6.5] and 24 CML-CCyR samples (CML-CCyR7 and -11 removed due to low RINs, CML-CCyR9 and -12 failed QC during pre-processing while there was no remaining sample for CML-CCyR16) were available for analysis (23 of the original 29 paired samples). In all, 16 of 17 CML-CP-NR samples were included (no remaining sample for CML-CPNR2) along with 22 healthy controls, 15 samples of which were common to both HM450 and HT-12 arrays (Table 3.2).

5.5.2 Overall strategy

Since a robust epigenetic response signature could not be established and validation of RNA samples was not possible for the reasons outlined above (Section 5.5.1), the focus for differential expression analyses was between CML-CP (combined CML-CP-R and CML-CP-NR) and healthy controls.

In addition, there were concerns that the combined effects of G-CSF mobilisation and imatinib therapy might impact on CML-CCyR cells at the transcription level and cause gene expression profile inconsistencies. Given analogous, unmobilised, remission bone marrow samples were unavailable for comparison, testing of this group was restricted to preliminary clustering only.

5.5.3 Hierarchical cluster analysis of gene expression profiles between CD34⁺ CML-CP, CML-CCyR and healthy control cells

In order to evaluate gene expression differences between groups, unsupervised hierarchical cluster analyses were performed for CD34⁺ CML-CP, CML-CCyR and healthy controls (Figure 5.3). This identified three distinct clusters: (1) a homogenous group of CD34⁺ CML-CP cells, independent of response to imatinib (i.e. no clear separation between CML-CP-R and CML-CP-NR) (2) CD34⁺ healthy control cells (G-CSF mobilised), and (3) a group containing CD34⁺ CML-CCyR cells (G-CSF mobilised). One CML-CP sample (CML-CP-R30) clustered with CML-CCyR patients without any clear biological or technical reason. One explanation might relate to increased levels of background noise associated with gene expression arrays, although this should have largely been compensated for during the preprocessing steps.

Strikingly, gene expression profiles from mobilised healthy controls clustered more closely with CML-CP cells. In the same way, CML-CP samples were more distinct from paired, mobilised CML-CCyR cells than healthy controls. This divergence is likely to be therapy-driven rather than disease-related i.e. short-term G-CSF in combination with continuing imatinib therapy in CML-CCyR patients at the time of peripheral blood stem cell harvest.

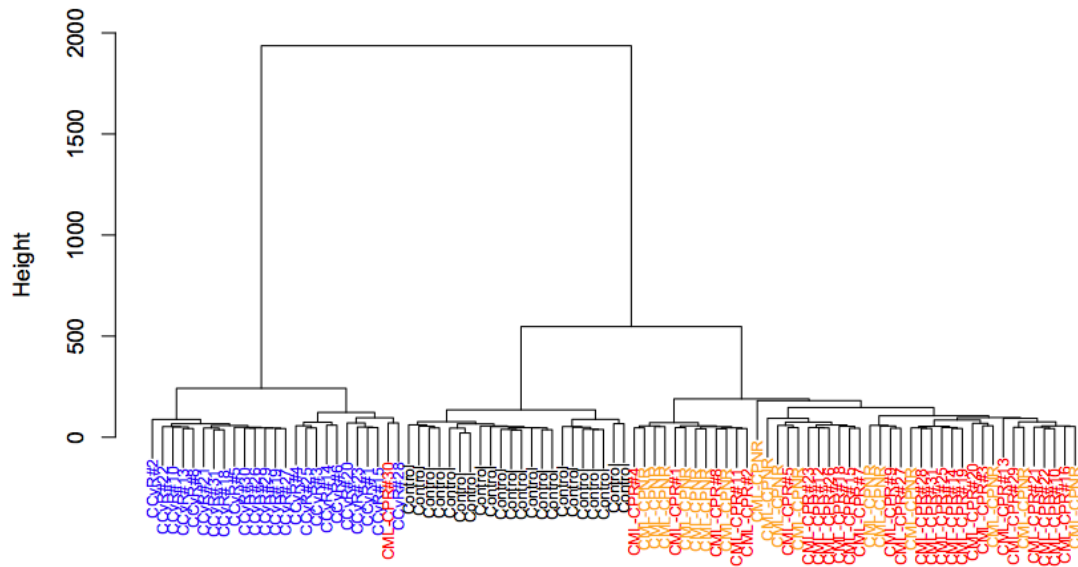


Figure 5.3 Unsupervised hierarchical cluster dendrogram of HT-12 gene expression data

Whole-genome gene expression profiles of CD34⁺ cells separate patients and controls into three distinct clusters; CML-CCyR (blue, mobilised), Healthy controls (black, mobilised) and CML-CP (orange and red). The clustering of CML-CCyR cells is likely to be driven in part by the combined effects of G-CSF and imatinib at the time of peripheral blood stem cell collection.

5.5.4 Gene expression signature of CD34⁺ cells from patients with CML-CP

Microarray-based gene expression profiling was used to compare global gene expression differences in peripheral blood CD34⁺ cells from newly diagnosed, untreated CML-CP patients and mobilised healthy controls.

Of 47,231 genes, a total 9,415 (3,632 under- and 5,783 overexpressed in CML) were significantly differentially expressed using a conservative FDR-adjusted *p*-value threshold (*q*-value <0.01). Considering only those with log fold change (FC) values greater than 1.2, of these, 200 (5.5%) genes were down regulated in CML relative to healthy controls while 158 (2.7%) demonstrated up-regulation. However, one disadvantage of using fold-change in this setting is the risk of neglecting differentially expressed genes with large differences but small ratios, leading to a high miss rate at high intensities.

Amongst these genes, Growth arrest specific 2 (*GAS2*) (*q*-value= 2.018e-15; log-fold change 2.39) was over-expressed in CML-CP. *GAS2* is a caspase-3 substrate and component of the microfilament system that regulates cell shape during apoptosis. In addition, *GAS2* can alter cellular susceptibility to p53-dependent apoptosis through binding the calmodulin-like region of m-calpain and directly inhibiting calpain activity [354, 355]. The calpains themselves are regulatory, (rather than digestive) calcium-dependent proteases whose complete biological roles are still poorly understood but include processes such as cell mobility and cell cycle progression [356].

In the blood, calpains have been reported to regulate numerous stages of the phagocyte response, including neutrophil migration, degranulation, and chemotaxis, while other calpain members are expressed during macrophage activation [357-359]. In this study, the genes encoding both m-calpain family members, calpain 1 and 2 were differentially underexpressed in CML-CP patients including the large distinct heterodimeric subunits *CAPN1* (*q*-value= 0.00012; log-fold change -0.6449) and *CAPN2* (*q*-value= 1.31e-06; log-fold change -0.7232) along with their identical small subunit *CAPNS1* (previously known as *CAPN4*) (*q*-value= 6.59e-05; log-fold change -0.5058, *q*-value= 8.64e-06; log-fold change -0.4144, alternative splice variants). The transcript changes were all significant using multiple testing with FC expression values greater than 1.2, possibly reflecting the effect of increased *GAS2* expression.

Crucially, β -catenin is a substrate of the calpain system where increased calpain protease activity leads to the degradation of β -catenin [360]. In this way, GAS2 can modulate the levels, action, and localisation of β -catenin through endogenous calpain regulation in the absence of external stimuli. Given that increased β -catenin activity correlates with leukemia stem cell expansion and disease progression in CML [95, 96], Huang and colleagues demonstrated a GAS2-dependent increase in the stability and activity of β -catenin in BCR-ABL1⁺ myeloid progenitor cells. This was reversible on either GAS2 knockdown or overexpression [361] highlighting the GAS2-calpain axis as a potentially significant pathway in the regulation of β -catenin in CML. *CTNBL1*, the gene encoding β -catenin was not differentially expressed in CML-CP samples which is not unexpected given that β -catenin levels in myeloid progenitors from patients with CML only rise at the time of disease progression (CML-AP and -BC) [96].

Janssen and colleagues have reported deregulation of GAS2 in CML [362]. They demonstrated increased expression levels at the time of disease progression to CML-BC. Radich *et al* also described *GAS2* overexpression in CD34⁺ CML-BC patients relative to CML-CP cells [363] while another microarray-based gene expression study showed that *GAS2* expression was higher in CD34⁺ CML-CP patients than healthy control bone marrow [344]. More recently, Zhou and others confirmed these findings [364]. They used RNA interference (RNAi) to investigate the functional role of GAS2 in CML-CP cells and established that knock-down of GAS2 resulted in the inhibition of CML cell growth. Furthermore, *in vitro*, targeting of GAS2 sensitised K562 TKI-resistant cell lines to imatinib.

The transcriptional target of mitogen-activated protein kinase (MAPK) pathways, *MYC* was significantly upregulated in CML-CP cells (q -value= 9.38e-05; log fold change 1.59), supporting a key role for *MYC* activation in BCR-ABL1-induced transformation [365]. This confirms the observation by Diaz-Blanco and others who previously reported *MYC* overexpression between CML patients and healthy control CD34⁺ cells [344]. Using different statistical methods and less stringent significance criteria to those set in this study, the same group also showed that the positive regulator of *MYC*, heterogenous nuclear ribonucleoprotein K (*HNRPK*), was also upregulated in CML-CP leading to the conclusion that HNRPK supports MAPK pathway-dependent activation of *MYC*. In this current study, *HNRPK* expression was

also differentially overexpressed at the time of CML-CP (q -value= 0.0039), however the log fold change was small, 0.071.

Similarly, and in agreement with the findings of Diaz-Blanco, the neuroepithelial cell-transforming gene 1 (*NET1*), a target of transforming growth factor beta 1 (TGF- β 1) was upregulated in CD34⁺ CML-CP cells (q -value= 2.34e-07; log-fold change 1.58). NET1 initiates activation of the alternative JNK MAPK signaling pathway [366] and therefore, its upregulation might contribute to the MAPK-mediated mitogenic activation of CML CD34⁺ cells.

Of the genes showing lower expression levels in CML-CP patients compared with healthy controls, the nuclear receptor subfamily 4A member 2 (*NR4A2*) was significantly reduced (q -value= 0.00013; log-fold change -1.35 and q -value= 0.00046; log-fold change -1.54, alternative splice variants). This is in alignment with a study by Bruns and colleagues who describe reduced *NR4A2* transcriptional activity in primitive CML-CP cells [346].

Gene functions of the CML-CP gene set were organised by KEGG functional annotation and corrected for multiple hypotheses testing using the Benjamini-Hochberg method (FDR 1%). Over- and under-expressed genes were analysed collectively rather than separately to reflect the complex interaction of genes in any biological network. A total of 30 significant pathways were derived (Table 5.1). While the genes themselves varied, many of the biological processes overlapped with those generated during DNA methylation analysis (Section 3.4.5.3) going some way to support the validity of these findings as opposed to random enrichment of categories. Of particular relevance, pathways involved in CML (hsa05220, fold enrichment 22.2X; p -value: 7.61e-07), cancer (hsa05200, fold enrichment 17.2X; p -value: 5.20e-22), the cell cycle (hsa04110, fold enrichment 38.4X; p -value: 4.91e-26), haematopoietic cell lineage (hsa04640, fold enrichment 31.8X; p -value: 3.07e-13) and MAPK signaling (hsa04010, fold enrichment 7.8X; p -value: 3.07e-05) were significantly enriched and deregulated in CML-CP.

Table 5.1 KEGG functional annotation of genes aberrantly expressed in CD34⁺ CML-CP cells

KEGG pathway	Examples of annotated genes in CML-CP	Fold enrichment	<i>p</i> -value
hsa04110: Cell cycle	E2F2, MCM4, MCM5, WEE1, CDC25A, CCNE1, CDKN1A, CCNB2, CDKN1B, CCND2, BUB1, ABL1, CCNA2, GADD45A, MYC	38.4	4.91e-26
hsa05200: Pathways in cancer	E2F2, FOXO1, BCL2L1, CCNE1, CASP3, SLC2A1, MYC, CSF2RA, FLT3, BIRC5, BIRC3, STAT3, DAPK1, CDKN1A, CDKN1B, LAMA5, JAK1, ABL1	17.2	5.20e-22
hsa05322: Systemic lupus erythematosus	HIST1H2AC, HIST2H2AA3 and AA4, HIST1H2BC, HIST2H2BE, FCGR2A, HLA-DPB1, HLA-DOA	33.9	6.42e-17
hsa04060: Cytokine-cytokine receptor interaction	CCL3, TNFSF4, TNFRSF25, FLT3, LEPR, TNFSF14, CCL5, LEP, IL12RB2, CCL23, CCL3L1, CCL3L3, IL12A, MPL, CSF2RA	15.1	3.20e-14
hsa04640: Haematopoietic cell lineage	FLT3, ITGAM, CD38, CD37, CD36, CD44, TFRC, DNMT, CD33, EPOR, CSF2RA, CD7, ITGA2B	31.8	3.07e-13
hsa04630: Jak-STAT signaling pathway	LEPR, CBL, BCL2L1, STAT3, IL12RB2, LEP, STAT4, SPRY1, CCND2, IL12A, JAK1, EPOR, MPL, MYC, CSF2RA	19.9	1.04e-12
hsa04514: Cell adhesion molecules (CAMs)	PTPRC, F11R, SELL, CD99, CDH1, CDH2, HLA-E, CLDN15, HLA-F, HLA-DPB1, HLA-DOA	20.2	5.58e-11
hsa04530: Tight junction	PRKCZ, RRAS, PRKCH, ACTN1, PRKCD, CLDN15	18.7	1.04e-09
hsa05222: Small cell lung cancer	E2F2, CCNE1, CKS1B, CDKN1B, LAMA5, RXRA, BCL2L1, BIRC3, MYC, ITGA2B	24.7	4.95e-09
hsa03030: DNA replication	POLE2, LIG1, POLE, PCNA, MCM2, MCM4, FEN1, MCM5	46.9	5.01e-09
hsa04612: Antigen processing and presentation	TAP1, HLA-DPB1, HLA-E, HLA-DOA, CANX, CD74, CTS1, TAPBP, HLA-F	22.3	1.03e-07
hsa04062: Chemokine signaling pathway	PRKCZ, CCL3, CCL23, CCL3L1, CCL3L3, PRKACB, CCL5, PRKCD, STAT3, PRKCB, GNG7, ELMO1	12.3	3.30e-07
hsa04115: p53 signaling pathway	CCNE1, CDKN1A, CASP3, CCNB2, CCND2, CHEK1, THBS1, GADD45A	24.5	4.37e-07
hsa04510: Focal adhesion	CCND2, LAMA5, ITGB7, ACTN1, ZYX, PDGFD, BIRC3, THBS1, FLNB, ITGA2B, PRKCB	11.4	5.66e-07
hsa05220: Chronic myeloid leukemia	E2F2, CDKN1A, CDKN1B, GAB2, CBL, BCL2L1, ABL1, MYC	22.2	7.61e-07
hsa04810: Regulation of actin cytoskeleton	CHRM3, ARHGEF6, ITGB7, RRAS2, CYFIP2, RRAS, ACTN1, PDGFD, ITGAM, ITGA2B, F2R	10.6	9.81e-07
hsa04666: Fc gamma R-mediated phagocytosis	LAT, PTPRC, GAB2, MARCKS, FCGR2A, PRKCD, PLA2G4B, PRKCB	17.8	3.04e-06
hsa03410: Base excision repair	HMGB1L1, POLE2, LIG1, POLE, PCNA, FEN1	36.2	5.12e-06
hsa04920: Adipocytokine	LEP, CPT1B, CD36, LEPR, RXRA, SLC2A1, STAT3	21.8	6.63e-06

signaling pathway			
hsa04114: Oocyte meiosis	CCNE1, CCNB2, BUB1, FBXO5, CDC20, AURKA, PTTG1, PRKACB	15.1	8.19e-06
hsa04270: Vascular smooth muscle	PRKCH, PRKACB, PRKCD, PLA2G4B, PPP1R14A, PRKCB	14.7	9.36e-06
hsa04142: Lysosome	AP1S2, LAPTM5, IDS, TPP1, CTSO, CTSG, CTS1, CTSF	14.0	1.20e-05
hsa05219: Bladder cancer	E2F2, CDKN1A, CDH1, THBS1, MYC, DAPK1	29.3	1.19e-05
hsa02010: ABC transporters	ABCA7, TAP1, ABCB1, ABCA1, ABCB6, ABCC5	28.0	1.44e-05
hsa04020: Calcium signaling pathway	SLC8A3, CD38, ATP2B4, CHRM3, RYR3, PRKACB, ITPKA, PRKCB, F2R	10.6	1.46e-05
hsa04144: Endocytosis	PRKCZ, RAB31, AP2B1, TFRC, CBL, RAB22A, HLA-E, F2R, HLA-F	10.3	1.66e-05
hsa04512: ECM-receptor interaction	CD36, CD44, LAMA5, ITGB7, THBS1, ITGA2B, HMMR	17.5	1.68e-05
hsa04540: Gap junction	TUBB6, TUBA4A, GUCY1A3, PRKG2, PDGFD, PRKACB, PRKCB	16.3	2.45e-05
hsa04080: Neuroactive ligand-receptor interaction	LEP, P2RY8, PTGER2, PTGER4, CHRM3, RXFP1, LEPR, F2RL1, CTSG, F2R	8.0	2.47e-05
hsa04010: MAPK signaling pathway	DUSP3, CASP3, RRAS2, RRAS, PRKACB, GADD45A, PLA2G4B, MYC, FLNB, PRKCB	7.8	3.07e-05

5.5.5 Differentially methylated positions (DMPs) as regulators of gene expression in CML-CP

To investigate the epigenetic features associated with CML leukaemogenesis, DNA methylation β -value data between CML-CP and healthy controls were integrated with log2 transformed transcription data.

For each DMP, the correlation between β -value and log2 transformed gene expression was evaluated using the Spearman's correlation coefficient. Statistical significance was calculated by permuting the data 10,000 times and comparing the correlation coefficient in the unpermuted data to the permuted coefficients. The permuted P-values were adjusted using the Benjamini-Hochberg multiple testing method for controlling FDR.

Overall, no statistically significant association between CML-CP promoter, or gene body DMPs and expression could be demonstrated, irrespective of whether the DMPs became more or less methylated. This indicated that in a substantial proportion of genes, DNA methylation alterations did not directly translate to major changes in expression.

Next, expression differences for genes with >20% methylation differences between CML-CP and healthy controls at promoter CGIs (within 1.5 kb of a TSS) were plotted. A significant negative-correlation was demonstrated in 31 of the 42 genes and 56 of the 69 DMPs tested (Spearman's correlation coefficient = -0.70, p -value = $1.4e-12$) (Table 5.2).

As confirmation of biological validity, the majority of the genes were represented more than once either as multiple probes from HM450 or alternative splice variants represented on the HT-12 e.g. a total of six HM450 probes were identified within the promoter region of *FAM83A* all of which were significantly hypomethylated and relatively over expressed.

Strikingly, among the list, promoter hypomethylation of *GAS2* in CML-CP patients relative to healthy controls was inversely correlated with transcriptional upregulation (described in section 5.5.4) supporting the recent findings of Perez and colleagues

who reported *GAS2* promoter hypomethylation in 56% percent of samples from untreated patients with Ph-negative myeloproliferative neoplasms (MPNs) using HM27 array [367]. Although gene expression was not measured as part of this study, it still goes some way in offering a potential epigenetic mechanism underlying *GAS2* augmented expression and implicates this gene as a potentially relevant pro-leukaemic biomarker in myeloid malignancies.

Work by Huang *et al* suggests the interferon consensus sequence binding protein (ICSBP) cooperates with Tel and HDAC3 to repress *GAS2* transcription in myeloid progenitor cells [361]. From the data generated in this chapter, *HDAC3*, the gene encoding the enzyme histone deacetylase 3, with promoter transcription repressor activity was the only one of these three genes overexpressed in CML-CP relative to healthy controls with a significant *q*-value but small fold-change below 1.2 (*q*-value= 1.04e-05; log-fold change 0.1754). It is possible that these small but significant changes do have some biological relevance and might indicate that *HDAC3* overexpression is a homeostatic response to compensate for elevated *GAS2* transcript levels and concomitant hypomethylation. Similarly, it is conceivable that the significant overexpression difference of the histone-lysine methyltransferase, enhancer of zeste homolog 2 (*EZH2*) gene (*q*-value= 8.78e-07; log-fold change 0.9631) is a finding consistent with malignancy in general or a regulatory process to counterbalance the hypomethylation of *GAS2* via silencing of histones H3 thorough lysine 27 methylation [361]. However such possibilities are entirely speculative and further work is essential not only to validate these findings but also to elucidate the epigenetic mechanisms through which they occur.

HOXA5, a regulator of embryonic and hematopoietic development and *TP53INP1* a tumor suppressor gene both demonstrated promoter hypermethylation coupled with relative gene under-expression. It is notable that *HOXA5* inactivation through promoter hypermethylation has been previously observed in leukemia and in particular during progression to blast crisis in CML [238]. Over expression of *TP53INP1* induces p53-mediated apoptosis and G1 cell cycle arrest [368], which suggests that promoter hypermethylation and subsequent gene repression would lead to loss of cellular homeostasis and abnormal proliferation. *TP53INP1* promoter hypermethylation has been detected in samples from patients with polycythaemia vera [367].

Gene Symbol	CMLCP-Normal Methylation (beta-value difference)	CMLCP-Normal Expression (log2 transformed difference)	Illumina Methylation Probe ID	Illumina expression Probe_ID	GEO Accession/version
CCL5	-0.2580	2.0542	cg08656816	ILMN_1773352	NM_002985.2
CCL5	-0.2580	2.1376	cg08656816	ILMN_2098126	NM_002985.2
CD33	-0.2449	1.6604	cg24809544	ILMN_1747622	NM_001772.3
CLDN15	0.2452	-1.0490	cg07651914	ILMN_1682226	NM_014343.1
CLDN15	0.3273	-1.0490	cg02512860	ILMN_1682226	NM_014343.1
CLDN15	0.2452	-0.7001	cg07651914	ILMN_2405667	NM_138429.1
CLDN15	0.3273	-0.7001	cg02512860	ILMN_2405667	NM_138429.1
CLEC3B	0.2237	-0.9317	cg02235659	ILMN_1682176	NM_003278.1
CMTM5	-0.2490	0.7822	cg10149582	ILMN_1775373	NM_001037288.1
CMTM5	-0.2189	0.7822	cg00174500	ILMN_1775373	NM_001037288.1
CRHBP	0.3055	-2.6924	cg25913647	ILMN_1761312	NM_001882.3
CRHBP	0.2858	-2.6924	cg06495038	ILMN_1761312	NM_001882.3
CRHBP	0.2156	-2.6924	cg16545105	ILMN_1761312	NM_001882.3
FAM38B	-0.2217	1.6150	cg12673429	ILMN_1665686	NM_022068.1
FAM83A	-0.3568	0.7822	cg22384356	ILMN_2239774	NM_032899.4
FAM83A	-0.4320	0.7822	cg23067535	ILMN_2239774	NM_032899.4
FAM83A	-0.3776	0.7822	cg03531754	ILMN_2239774	NM_032899.4
FAM83A	-0.2164	0.7822	cg02715629	ILMN_2239774	NM_032899.4
FAM83A	-0.3563	0.7822	cg24833277	ILMN_2239774	NM_032899.4
FAM83A	-0.3493	0.7822	cg19924352	ILMN_2239774	NM_032899.4
FCGR2A	-0.2126	0.7891	cg06350097	ILMN_2289849	NM_021642.2
FCGR2A	-0.2126	1.1309	cg06350097	ILMN_1666932	NM_021642.2
FREM1	-0.2751	1.0464	cg14356190	ILMN_1715452	NM_144966.4
FREM1	-0.2061	1.0464	cg00232500	ILMN_1715452	NM_144966.4
GAS2	-0.6024	2.7241	cg06493930	ILMN_1804569	NM_005256.2
GAS2	-0.2953	2.7241	cg15585294	ILMN_1804569	NM_005256.2

GAS2	-0.5019	2.7241	cg16363146	ILMN_1804569	NM_005256.2
GAS2	-0.4425	2.7241	cg12575928	ILMN_1804569	NM_005256.2
GAS2	-0.2064	2.7241	cg03547797	ILMN_1804569	NM_005256.2
GAS2	-0.3884	2.7241	cg02026246	ILMN_1804569	NM_005256.2
HBBP1	-0.2308	1.1907	cg05133853	ILMN_1653814	NR_001589.1
HOXA5	0.3714	-0.8897	cg13694927	ILMN_1753613	NM_019102.2
HOXA5	0.3375	-0.8897	cg01748892	ILMN_1753613	NM_019102.2
HOXA5	0.3239	-0.8897	cg03744763	ILMN_1753613	NM_019102.2
ITGA2B	-0.2849	1.9799	cg25809905	ILMN_1721888	NM_000419.2
ITGB7	-0.2342	0.8907	cg26689077	ILMN_1777519	NM_000889.1
ITGB7	-0.2036	0.8907	cg15889057	ILMN_1777519	NM_000889.1
KIAA0664	-0.2037	0.7302	cg00668150	ILMN_1770719	NM_015229.3
LEPR	-0.5560	0.7954	cg16265717	ILMN_1731966	NM_001003679.1
LEPR	-0.5560	1.7564	cg16265717	ILMN_2234956	NM_001003679.1
MN1	0.2165	-1.0730	cg12872311	ILMN_2111237	NM_002430.2
MN1	0.2165	-0.8699	cg12872311	ILMN_1732197	NM_002430.2
MPL	0.2563	-1.5951	cg22607339	ILMN_1776640	NM_005373.1
MSRB3	0.3632	-1.7165	cg12488187	ILMN_1676088	NM_198080.2
MSRB3	0.3632	-1.6862	cg12488187	ILMN_2332553	NM_198080.2
MSRB3	0.3632	-1.1167	cg12488187	ILMN_1779333	NM_001031679.1
MX1	0.2584	-1.4910	cg26312951	ILMN_1662358	NM_002462.2
MYL4	-0.2411	1.3166	cg05294455	ILMN_1684255	NM_002476.2
PRG2	-0.2604	1.6565	cg10426076	ILMN_1729314	NM_002728.4
PROM1	0.3407	-1.8934	cg22867816	ILMN_1786720	NM_006017.1
PROM1	0.3575	-1.8934	cg13117948	ILMN_1786720	NM_006017.1
PROM1	0.2974	-1.8934	cg07656173	ILMN_1786720	NM_006017.1
RCSD1	0.2141	-1.0792	cg04275362	ILMN_1749006	NM_052862.2
SEPP1	0.4253	-1.6099	cg25310824	ILMN_1785071	NM_005410.2
SEPP1	0.2668	-1.6099	cg08626131	ILMN_1785071	NM_005410.2
SLC29A1	-0.2216	0.7603	cg01993576	ILMN_1723971	NM_001078174.1

STAT4	-0.3248	1.2018	cg04285318	ILMN_1785202	NM_003151.2
TMEM136	0.4563	-1.1675	cg02610360	ILMN_1815346	NM_174926.1
TNS3	0.3749	-1.1118	cg23895495	ILMN_1667893	NM_022748.10
TNS3	0.3158	-1.1118	cg11203616	ILMN_1667893	NM_022748.10
TNS3	0.2004	-1.1118	cg17858880	ILMN_1667893	NM_022748.10
TP53INP1	0.2015	-0.8463	cg01824466	ILMN_2214197	NM_033285.2
TP53INP1	0.2031	-0.8463	cg13393036	ILMN_2214197	NM_033285.2

Table 5.2 Differentially methylated CML-CP promoter CGIs demonstrating anti-correlation with gene expression changes (Spearman's correlation coefficient = -0.70, p -value = 1.4e-12).

5.6 Discussion

There are few reports investigating the possible influence of genome-wide DNA methylation patterns on gene expression in CML. Thus, transcriptome profiles of CML-CP patients were generated and integrated with methylomic data from paired samples to identify functionally relevant epigenetic events associated with BCR-ABL1-driven cellular transformation.

For the majority of genes, DNA methylation alterations did not translate into significant expression changes suggesting a high rate of passenger epimutations arising independently of the transcriptome. However, a total of 31 genes displayed anti-correlated promoter DMPs and gene expression differences pointing towards epigenetic dysregulation.

Although there are no other published studies to date, which have considered the CML epigenome in this way, there are a number of shortcomings with this work which can be addressed with further experiments but must be taken into account when interpreting the data and drawing any conclusions.

Whereas the methylation data was validated in an independent cohort, this was not possible for the gene expression study, the principle reason being a lack of remaining material from the SPIRIT 2 trial samples. Alternative samples for validation were identified but the application for further ethics approval was outside the realistic timeline for work to be carried out and will therefore form the basis for a future study.

While biological validation is necessary to prove that the differences generated are reproducible and applicable to other cohorts of CML-CP patients a more timely issue relates to the technical validation of the HT-12 platform.

The HT-12 array offers a high-throughput, cost-effective means of measuring genome-wide expression levels. Nevertheless, as with any microarray-based technology, the large number of genes analysed inevitably leads to chance fluctuations and a high coefficient of variation resulting in anomalous false positive and negative candidate targets [369]. Additionally, the relatively narrow dynamic range of the fluorescent spot intensities (10^2 - 10^3) translates into unavoidable

limitations with respect to sensitivity such as the failure to detect real, small differences (low expressors) when samples contain low cell numbers [370].

Technical validation of the differentially expressed genes identified using real-time reverse transcription-polymerase chain reaction (RT-PCR) is underway on a total of 75 genes (including 46 genes over- and 29 under expressed) identified as being differentially expressed between CML-CP and healthy controls. Genes were selected based on their potential importance in relevant CML signaling pathways and statistical significance (i.e. $q < 0.01$) rather than fold change expression in order that the results are not biased towards large fold-change differences. This will be achieved using the TaqMan 384-well microfluidic card using 18S rRNA, GAPDH, and RPLP0 as housekeeping genes and performed in triplicate for each CML-CP sample and an RNA pool of healthy control samples.

A serious concern arising from this work relates to the use of G-CSF mobilised peripheral blood stem cells (PBSCs). In Chapter 3, it was possible to show that DNA methylation differences exist in a proportion of CML-CCyR patients on imatinib, which were unlikely to be disease related and somehow associated with the combined use of long-term TKI therapy and short-term exposure to G-CSF during peripheral blood stem cell (PBSC) harvest. These differences were not apparent in analogous unmobilised bone marrow cells from the same patients despite remaining on TKI inhibitors; neither could they be detected in healthy control donors who had never received imatinib/TKI therapy but who had undergone G-CSF PBSC collection.

While published data for the effects of G-CSF on the human methylome is lacking, some evidence is available with respect to gene expression, since G-CSF is the most commonly used cytokine for PBSC mobilisation from healthy donors for the purpose of allo-SCT.

Steidl *et al* examined the expression levels of 1185 genes in CD34⁺ cells from a total of 18 healthy volunteers consisting of 10 G-CSF mobilised peripheral blood stem cell donors (PBSCs) and 8 non-mobilised bone marrow samples using the Atlas Human 1.2 I cDNA array [371]. Hierarchical clustering separated the gene expression profiles of bone marrow and circulating CD34⁺ cells in to two groups with the bone marrow-derived cells showing greater homogeneity between samples. A total of 65 genes were significantly differentially expressed between the transcriptomes of the two groups.

Genes with cell cycle and DNA synthesis activity were over-represented in the unmobilised bone marrow cells, while mobilised PBSCs showed increased expression of genes driving apoptosis and differentiation arrest including GATA2. Despite the small sample size, the group concluded that due to the differing environments, the gene expression signatures of mobilised CD34⁺ cells might not fully reflect the signature of the HSCs in the BM niche.

Recently, Báez and colleagues analysed the immediate (day +5) and longer term (days +30 and +365) effects of G-CSF on gene and microRNA (miRNA) expression profiles in CD34⁺ cells from 6 healthy allo-SCT donors with the aim of defining the type and duration of detectable differences between paired pre- and post-exposure samples [372]. The group showed that miRNA expression differences were greatest on day +30 following G-CSF treatment when 179 of the 384 miRNAs measured were significantly altered. Overall, 6 miRNAs were over-expressed from the point of exposure out to 1 year compared with no differences in control patients who were assessed at the same time points after injection with vehicle (normal saline). In addition, 4136 genes were identified as differentially expressed in pre-mobilised CD34⁺ cells versus those obtained 5 days following G-CSF exposure. At 30 days, 3061 (74%) of these remained altered while 2424 (59%) maintained their aberrant expression levels from day 5 to 1 year after the treatment. Functional ingenuity pathway analyses of these genes showed involvement of several biological processes such as cancer, gene expression, protein synthesis and haematopoiesis although due to the brief statistical methodology outlined, it is unclear whether a genome-wide correct P-value or multiple hypotheses testing was applied to this small study population.

While both of these publications highlighted source-dependent gene expression variations in CD34⁺ cells, the genes identified were not consistent between experiments. The reasons for this are likely to be a combination of any of the factors discussed earlier inherent to array-based transcriptomics as well as the small sample sizes analysed and the type/version of the platform used e.g. older studies employing earlier generation microarray technology were less specific and sensitive.

In terms of timing, the sample processing and RNA extraction for this study overlapped with the DNA methylation data analysis from Chapter 3. The combined effects of imatinib and G-CSF on the methylation patterns in CML-CCyR cells were

identified after the prepared RNA samples were hybridised to the HT-12 arrays. Hierarchical clustering of the gene expression data separated the CML-CCyR samples away from both the CML-CP and healthy controls, which clustered closer together. At the time, the reasons for this were still unclear because the bone marrow aspirate samples collected to investigate methylation discrepancies had not been processed. With respect to the collection of unmobilised bone marrow, patients donated on the basis of their methylation patterns i.e. CD14⁺-like or CD34⁺-like but also according to whether they were enrolled on to the DESTINY (De-Escalation and Stopping Treatment of Imatinib, Nilotinib or sprYcel in Chronic Myeloid Leukaemia) trial (ClinicalTrials.gov Identifier: NCT01804985). For these patients a bone marrow aspiration procedure was performed at study entry for molecular monitoring and biomarker discovery. For the purposes of this experiment, patients consented to the additional donation of 15 mLs of material in the same procedure. On average the volume of aspirate returned approximately $0.5-1 \times 10^6$ CD34⁺ cells. While the cell quantity was optimal for use with HM450, numbers were too low for additional RNA extraction.

Similarly, obtaining sufficient numbers of CD34⁺ cells for RNA extraction in frozen, healthy donor bone marrow aspirate samples was not possible. One solution to address this problem using the 7 samples available might have been to share them between two groups e.g. 3 for DNA extraction and methylation studies and 4 for RNA extraction for gene expression work, although this would have affected the overall, statistical power of the analyses by reducing the number of biological replicates.

In practice, once the methylation changes were shown to be drug-related, and the gene expression clustering suggested equivalent effects were also occurring at the RNA level, further analysis involving remission samples was abandoned.

For the changes between CD34⁺ CML-CP and mobilised healthy control cells to be investigated, all analyses in this chapter assumed that any small differences between G-CSF mobilised PBSCs and bone marrow from healthy controls had minimal influence on the overall CML-CP expression signature itself. In support of this, many of the genes identified as transcriptionally dysregulated have previously been described in CML or had biological functions relevant to the disease pathogenesis.

To validate the healthy control expression signature, future work may well include identifying stored or collecting fresh normal bone marrow samples for transcriptome profiling followed by technical validation of significant hits using QRT-PCR. Alternatively, with the increasing popularity of the HT-12 array, it might be possible to mine publically available data for any experimentally relevant raw HT-12 expression values.

In CML this will prove quite challenging since only one other group has used the HT-12 array to date. Kok *et al*, recently reported and validated a predictive gene signature to identify CML-CP patients at diagnosis, at high risk of early molecular response failure and subsequent inferior clinical outcomes [373]. In this study, healthy control samples were not used, the TKIs included imatinib and nilotinib rather than dasatinib and moreover, the source of cells were PBMCs rather than selected CD34⁺ cells which would affect the concordance of gene expression patterns substantially between data sets [374].

DNA methylation is an essential epigenetic process at the gene transcription level, yet it is becoming increasingly clear that the precise role of this modification in regulating gene expression at promoter-associated CGIs is complex and context dependent. While some studies have reported anti-correlation of methylation levels with gene expression at these locations [375, 376] others describe absent or at best a weak relationship between the two [121, 377]. Moreover, cancer-associated DNA methylation changes during malignant transformation are not always accompanied by related gene expression changes [378].

In summary, the purpose of this chapter was firstly to examine the global gene expression patterns of CD34⁺ progenitor cells from patients with CML-CP using microarray-based profiling and compare these with normal CD34⁺ cells from healthy controls. Subsequently, transcriptome changes were intersected with derived methylome data from the same samples in order to identify perturbed epigenetic events driving putative pathways in leukaemogenesis.

Overall, these analyses demonstrate that CML-CP is associated with significant gene expression alterations as well as dysregulated methylation although; these changes seem to provide orthogonal information about the disease state. However, a core set

of promoter DMPs were identified that likely play a role in modulating transcript levels and should be validated in future prospective studies.

6 Conclusion and Future Directions

6.1 Conclusion

CML represents a paradigm in the therapeutic targeting of human cancers and is arguably one of the most widely studied malignancies in terms of its underlying biology and pathophysiology. Understanding the molecular properties of CML stem cells could deliver new insights into the process of malignant transformation and disease progression in less well-characterised cancers. This work speculated that aberrant epigenetic programming and in particular abnormal DNA methylation might contribute to the disease pathogenesis or underlie the variations in outcome and responses to imatinib. Therefore, an integrated epigenomic study was performed on CD34⁺ cells from CML patients to see whether the DNA methylation and transcription profiles would reflect disease-specific differences as a consequence of the t(9;22) and to investigate the extent to which epigenetic profiling could predict for response to TKI therapy.

In conclusion, consistent perturbations in the genome-wide methylation patterns of CD34⁺ cells from patients with CML were identified suggesting that epigenetic dysregulation could be involved in the leukemic transformation process. Conversely, the consistency with which these methylation aberrations occur provides no explanation for the clinical heterogeneity observed in this leukemia. By comparing the pattern of DNA methylation in samples with a number of different tissues and tumour types including hematological malignancies, it was found that in addition to retaining recognisable features with their non-malignant stem and progenitor cell counterparts, CD34⁺ CML-CP cells share striking methylation similarities with other leukaemias and solid cancers. Overall correlation between genome-wide gene expression and DNA methylation was weak but further sub-analysis revealed significant anti-correlation at promoter sites suggesting a functional role for DNA methylation in CML. Whether any of these changes are of pathological significance remains to be determined in future studies.

6.2 Future work

6.2.1 Genome-wide mapping of histone post-translational modifications (PTMs) in CD34⁺ CML cells using low-cell ChIP-seq

From this work, a number of genes were identified whose expression correlates strongly with DNA methylation. Amongst the biological pathways associated with these, those relating to nucleosome assembly, chromatin organisation and DNA packaging were highlighted. The lack of correlation at the remaining genomic loci, possibly implicates the interplay of other epigenetic marks such as histone modifications and chromatin modeling.

The role of histone post-translational modifications (PTMs) such as acetylation and methylation in regulating gene expression has been investigated in the course of normal development and tumourigenesis (reviewed in section 1.7.2.5). However, there are relatively few studies examining the genome-wide distribution of histone modifications in leukaemia stem cells and almost no published data available specifically in CML.

To establish the patterns of histone modifications and understand the complex relationship between chromatin remodeling and gene expression, further experiments are planned to conduct a comprehensive, unbiased co-localization analysis of histone post-translational modifications in CML progenitor cells. Using CD34⁺ cells from patients with untreated CML and comparing these to analogous cells from healthy controls, the genome-wide landscape of two histone marks associated with transcriptional activation (H3K4me3 and H3K36me3), two associated with transcriptional silencing (H3K27me3 and H3K9me3) and two linked with enhancer regulatory elements (H3K27ac and H3K4me1) will be determined using chromatin immunoprecipitation coupled with next generation deep sequencing (ChIP-seq) adjusted for low cell numbers (from 10,000 cells).

Chromatin immunoprecipitation (ChIP) provides an unbiased means of identifying crucial genomic DNA fragments that co-precipitate with any DNA-binding proteins including transcription factors and modified histones proteins *in vivo*. The assay will allow the identification of all the DNA segments in the CML genome physically

associated with the aforementioned histones under investigation, which can then be integrated with DNA methylation analysis and gene expression data.

For the purposes of this work, a cross-linked ChIP (X-ChIP) assay approach will be used where DNA-bound protein complexes are reversibly cross-linked by formaldehyde and sheared by sonication as opposed to native ChIP (N-ChIP), which uses unfixed chromatin that is fragmented by nuclease digestion. The full methods are not shown but available. The low-cell ChIP method has been optimised in murine embryonic stem cells by Dr Carolina Gemma, a postdoctoral researcher formerly in Vardhman Raykan's lab, Blizard Institute, QMUL, currently at the Imperial Centre for Translational and Experimental Medicine, Imperial College London.

Briefly, following thawing, CD34⁺ immunomagnetic selection and DNA cross-linking of primary cells, chromatin is sonicated into fragments of 200 to 300 base pairs. This is followed by immunoprecipitation, using specific antibodies against the aforementioned proteins of interest. The success of a ChIP-seq project depends crucially on strong enrichment of the chromatin specifically bound by the protein under study. Carolina Gemma and Alexandra Bazeos have tested a number of antibodies and chosen those with consistently high enrichment of DNA at a known binding site when compared with the DNA immunoprecipitated by a nonspecific control anti-IgG antibody as a negative control. Library preparation requires the incorporation of indexes to permit multiplexing of samples during sequencing, the addition of linkers to facilitate binding to the sequencing platform and involves a PCR amplification step, during which time the number of cycles are minimised to avoid clonal amplification and bias. Following amplification, the correct size DNA fragments are selected and sequenced. Finally, the short sequenced fragments (known as tags) are computationally mapped by alignment to a reference genome and regions of enriched tag counts are identified, a step known as peak calling. Further bioinformatic work will include the comparison of this dataset to our previously generated DNA methylation and gene expression work as mentioned.

The interaction of BCR-ABL1 with chromatin remains elusive. In the same way that JAK2 tyrosine kinase has been shown to phosphorylate histones and alter chromatin structure in AML [379] it would be interesting to determine the genomic locations of ABL and BCR-ABL1 using ChIP-seq. It is believed that while ABL tyrosine kinase is

located in both the nucleus and cytoplasm, binding directly to DNA (via three nuclear localization sequences as well as a DNA binding domain) to influence histone acetylation, BCR-ABL1 is predominantly based in the cytoplasm and retains the capacity to translocate to the nucleus and interact directly with chromatin under certain conditions [380].

From a translational research perspective, histone deacetylase (HDAC) inhibitors are emerging as an exciting new class of anticancer agents for the treatment of solid and haematological malignancies. Zhang et al. have demonstrated that treatment with HDAC inhibitors combined with standard of care TKI therapy effectively induces apoptosis in quiescent CML stem cells resistant to elimination with single agent imatinib in a transgenic mouse model of CML [381]. In this way it would appear that HDAC inhibition might represent a promising anti-leukemic approach for imatinib-refractory patients in the clinic.

6.2.2 Somatic mutations and alterations in chromatin-modifying enzymes in CML

Mutations in epigenetic modifiers are emerging as a class of mutations critical in the transformation and progression of a number of acute leukaemias. Moreover, many leukaemia-associated fusion proteins are able to recruit transcription factors with chromatin- and cytosine modifying activities that contribute to the oncogenic properties of chimeric proteins including ASXL1, DNMT3A, EP300, EZH2, IDH1/2, MLL1/2, and TET2 [382]. So far, chronic myelomonocytic leukemia (CMML), a myelodysplastic/myeloproliferative disorder appears to be the haematological malignancy with the highest rate of epigenetic modifier gene mutations although few studies have measured the rates of these mutations in CML [383, 384].

The work in this thesis has shown that *EZH2*, *IDH1*, *MLL4* and *DNMT1* all display statistically significant expression differences in newly diagnosed CML patients along with genes encoding histones and *HOXA* genes (themselves regulated by EZH2).

To define the frequency of recurrent mutations in chromatin modifiers in CML, DNA from CD34⁺ cells of 26 CML-CP-R and 26 CML-CP-NR patients (including 19 patients from the Hammersmith cohort and 7 from the SPIRIT 2 study), are being sequenced using the Ion Torrent platform using methods detailed in section 2.2.9.

A total of 71 genes were selected for analyses based on literature searches of those previously reported as mutated in leukaemia as well as those differentially expressed from this study including 11 genes involved in DNA methylation (DNMT1, TRDMT1, DNMT3A, DNMT3B, DNMT3L, LSD1, TET1, TET2, IDH1, IDH2 and IDH3B).

Dr George Nteliopoulos is currently leading this project and preliminary data is pending.

References

1. Bennett, J., *Case of hypertrophy of the spleen and liver in which death took place from suppuration of the blood* Edinb Med Surg J, 1845. **64**: p. 413-423.
2. Virchow, R., *Weisses Blut*. Froriep's Notzien, 1845. **36**: p. 151-156.
3. Naegli, O., *On red bone marrow and myeloblasts [in German]*. Dtsch Med Wochenschr 1900. **18**: p. 287-290.
4. Cowan Doyle, A., *NOTES ON A CASE OF LEUCOCYTHÆMIA*. The Lancet 1882. **119 (originally Volume 1)**(3056): p. 490.
5. Nowell, P.a.H., *D A minute chromosome in human chronic granulocytic leukemia*. Science, 1960. **132**: p. 1497-1501.
6. Rowley, J., *A new consistent chromosomal abnormality in chronic myeloid leukemia identified by quinacrine fluorescence and Giemsa staining*. Nature, 1973. **243**: p. 290-293.
7. de Klein, A., et al., *A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia*. Nature, 1982. **300**(5894): p. 765-7.
8. Groffen, J., et al., *Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22*. Cell, 1984. **36**(1): p. 93-9.
9. Canaani, E., et al., *Altered transcription of an oncogene in chronic myeloid leukaemia*. Lancet, 1984. **1**(8377): p. 593-5.
10. Ben-Neriah, Y., et al., *The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene*. Science, 1986. **233**(4760): p. 212-4.
11. Daley, G.Q. and D. Baltimore, *Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein*. Proc Natl Acad Sci U S A, 1988. **85**(23): p. 9312-6.
12. Daley, G.Q., R.A. Van Etten, and D. Baltimore, *Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome*. Science, 1990. **247**(4944): p. 824-30.
13. Galton, D.A.G., *The use of myleran in chronic myeloid leukaemia*. The Lancet, 1953. **261 (Originally published as Volume 1)**(6753): p. 203-252.
14. Kennedy, B. and J. Yarbrow, *Metabolic and therapeutic effects of hydroxyurea in chronic myelogenous leukemia*. Trans Assoc Am Physicians, 1965. **78**: p. 391-399.
15. Talpaz, M., et al., *Hematologic remission and cytogenetic improvement induced by recombinant human interferon alpha A in chronic myelogenous leukemia*. N Engl J Med, 1986. **314**(17): p. 1065-9.
16. Goldman, J.M., et al., *Buffy coat autografts for patients with chronic granulocytic leukaemia in transformation*. Blut, 1981. **42**(3): p. 149-55.
17. Goldman, J.M., et al., *Bone marrow transplantation for patients with chronic myeloid leukemia*. N Engl J Med, 1986. **314**(4): p. 202-7.
18. Davies, S.M., et al., *Equivalent outcomes in patients with chronic myelogenous leukemia after early transplantation of phenotypically matched bone marrow from related or unrelated donors*. Am J Med, 2001. **110**(5): p. 339-46.

19. Druker, B.J., et al., *Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells*. Nat Med, 1996. **2**(5): p. 561-6.
20. Druker, B.J., et al., *Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome*. N Engl J Med, 2001. **344**(14): p. 1038-42.
21. Hughes, T.P., et al., *Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia*. N Engl J Med, 2003. **349**(15): p. 1423-32.
22. Druker, B.J., et al., *Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia*. N Engl J Med, 2006. **355**(23): p. 2408-17.
23. Institute, N.C. *Surveillance, Epidemiology, and End Results (SEER) programme* <http://seer.cancer.gov/statfacts/html/cmyle.html>.
24. Phekoo, K.J., et al., *The incidence and outcome of myeloid malignancies in 2,112 adult patients in southeast England*. Haematologica, 2006. **91**(10): p. 1400-4.
25. Rohrbacher, M. and J. Hasford, *Epidemiology of chronic myeloid leukaemia (CML)*. Best Pract Res Clin Haematol, 2009. **22**(3): p. 295-302.
26. Huang, X., J. Cortes, and H. Kantarjian, *Estimations of the increasing prevalence and plateau prevalence of chronic myeloid leukemia in the era of tyrosine kinase inhibitor therapy*. Cancer, 2012. **118**(12): p. 3123-7.
27. Kim, D.H., et al., *A genome-wide association study identifies novel loci associated with susceptibility to chronic myeloid leukemia*. Blood, 2011. **117**(25): p. 6906-11.
28. Hemminki, K. and Y. Jiang, *Familial myeloid leukemias from the Swedish Family-Cancer Database*. Leuk Res, 2002. **26**(6): p. 611-3.
29. Finch, S.C., *Radiation-induced leukemia: lessons from history*. Best Pract Res Clin Haematol, 2007. **20**(1): p. 109-18.
30. Ichimaru, M., et al., *Atomic bomb and leukemia*. J Radiat Res, 1991. **32 Suppl 2**: p. 14-9.
31. Mehlman, M.A., *Dangerous and cancer-causing properties of products and chemicals in the oil-refining and petrochemical industry--Part XXII: Health hazards from exposure to gasoline containing methyl tertiary butyl ether: study of New Jersey residents*. Toxicol Ind Health, 1996. **12**(5): p. 613-27.
32. Giles, F.J., et al., *Accelerated and blastic phases of chronic myelogenous leukemia*. Hematol Oncol Clin North Am, 2004. **18**(3): p. 753-74, xii.
33. Savage, D.G., R.M. Szydlo, and J.M. Goldman, *Clinical features at diagnosis in 430 patients with chronic myeloid leukaemia seen at a referral centre over a 16-year period*. Br J Haematol, 1997. **96**(1): p. 111-6.
34. Hughes, T., et al., *Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results*. Blood, 2006. **108**(1): p. 28-37.
35. SH Swerdlow, E.C., NL Harris, *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. WHO Classification of Tumours (4th edn.). Vol. Vol 2. 2008: World Health Organization, Geneva.
36. Baccarani, M., et al., *Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet*. Blood, 2006. **108**(6): p. 1809-20.

37. Sokal, J.E., et al., *Prognostic discrimination in "good-risk" chronic granulocytic leukemia*. Blood, 1984. **63**(4): p. 789-99.
38. Hasford, J., et al., *A new prognostic score for survival of patients with chronic myeloid leukemia treated with interferon alfa*. Writing Committee for the Collaborative CML Prognostic Factors Project Group. J Natl Cancer Inst, 1998. **90**(11): p. 850-8.
39. http://www.leukemianet.org/content/leukemias/cml/euro_and_sokal_score/index_eng.html.
40. Hasford, J., et al., *Predicting complete cytogenetic response and subsequent progression-free survival in 2060 patients with CML on imatinib treatment: the EUTOS score*. Blood, 2011. **118**(3): p. 686-92.
41. http://www.leukemia-net.org/content/leukemias/cml/eutos_score/index_eng.html.
42. Buchdunger, E., et al., *Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative*. Cancer Res, 1996. **56**(1): p. 100-4.
43. Buchdunger, E., et al., *Abl protein-tyrosine kinase inhibitor STI571 inhibits in vitro signal transduction mediated by c-kit and platelet-derived growth factor receptors*. J Pharmacol Exp Ther, 2000. **295**(1): p. 139-45.
44. Goldman, J., et al., *Biology of CML stem cells: the basis for clinical heterogeneity?* Leukemia Supplements, 2012. **1**: p. Leukemia Supplements.
45. Kantarjian, H., et al., *Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia*. N Engl J Med, 2002. **346**(9): p. 645-52.
46. Hochhaus, A., et al., *Six-year follow-up of patients receiving imatinib for the first-line treatment of chronic myeloid leukemia*. Leukemia, 2009. **23**(6): p. 1054-61.
47. Marin, D., et al., *European LeukemiaNet criteria for failure or suboptimal response reliably identify patients with CML in early chronic phase treated with imatinib whose eventual outcome is poor*. Blood, 2008. **112**(12): p. 4437-44.
48. Baccarani, M., et al., *European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013*. Blood, 2013. **122**(6): p. 872-84.
49. Marin, D., et al., *Assessment of BCR-ABL1 transcript levels at 3 months is the only requirement for predicting outcome for patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors*. J Clin Oncol, 2012. **30**(3): p. 232-8.
50. Jabbour, E., et al., *Frequency and clinical significance of BCR-ABL mutations in patients with chronic myeloid leukemia treated with imatinib mesylate*. Leukemia, 2006. **20**(10): p. 1767-73.
51. Soverini, S., et al., *Contribution of ABL kinase domain mutations to imatinib resistance in different subsets of Philadelphia-positive patients: by the GIMEMA Working Party on Chronic Myeloid Leukemia*. Clin Cancer Res, 2006. **12**(24): p. 7374-9.
52. Shah, N.P., et al., *Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia*. Cancer Cell, 2002. **2**(2): p. 117-25.

53. Soverini, S., et al., *BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet*. Blood, 2011. **118**(5): p. 1208-15.
54. Corbin, A.S., et al., *Several Bcr-Abl kinase domain mutants associated with imatinib mesylate resistance remain sensitive to imatinib*. Blood, 2003. **101**(11): p. 4611-4.
55. Sherbenou, D.W., et al., *Mutations of the BCR-ABL-kinase domain occur in a minority of patients with stable complete cytogenetic response to imatinib*. Leukemia, 2007. **21**(3): p. 489-93.
56. Cortes, J., et al., *Phase 2 study of subcutaneous omacetaxine mepesuccinate after TKI failure in patients with chronic-phase CML with T315I mutation*. Blood, 2012. **120**(13): p. 2573-80.
57. Nicolini, F.E., et al., *Epidemiologic study on survival of chronic myeloid leukemia and Ph(+) acute lymphoblastic leukemia patients with BCR-ABL T315I mutation*. Blood, 2009. **114**(26): p. 5271-8.
58. Johansson, B., T. Fioretos, and F. Mitelman, *Cytogenetic and molecular genetic evolution of chronic myeloid leukemia*. Acta Haematol, 2002. **107**(2): p. 76-94.
59. Hochhaus, A., et al., *Molecular and chromosomal mechanisms of resistance to imatinib (STI571) therapy*. Leukemia, 2002. **16**(11): p. 2190-6.
60. Thomas, J., et al., *Active transport of imatinib into and out of cells: implications for drug resistance*. Blood, 2004. **104**(12): p. 3739-45.
61. Marin, D., et al., *Adherence is the critical factor for achieving molecular responses in patients with chronic myeloid leukemia who achieve complete cytogenetic responses on imatinib*. J Clin Oncol, 2010. **28**(14): p. 2381-8.
62. Bazeos, A., et al., *hOCT1 transcript levels and single nucleotide polymorphisms as predictive factors for response to imatinib in chronic myeloid leukemia*. Leukemia, 2010. **24**(6): p. 1243-5.
63. White, D.L., et al., *OCT-1-mediated influx is a key determinant of the intracellular uptake of imatinib but not nilotinib (AMN107): reduced OCT-1 activity is the cause of low in vitro sensitivity to imatinib*. Blood, 2006. **108**(2): p. 697-704.
64. Giannoudis, A., et al., *Effective dasatinib uptake may occur without human organic cation transporter 1 (hOCT1): implications for the treatment of imatinib-resistant chronic myeloid leukemia*. Blood, 2008. **112**(8): p. 3348-54.
65. Mahon, F.X., et al., *MDR1 gene overexpression confers resistance to imatinib mesylate in leukemia cell line models*. Blood, 2003. **101**(6): p. 2368-73.
66. Koren-Michowitz, M., et al., *Imatinib plasma trough levels in chronic myeloid leukaemia: results of a multicentre study CSTI571AIL1ITGLIVEC*. Hematol Oncol, 2012. **30**(4): p. 200-5.
67. Larson, R.A., et al., *Imatinib pharmacokinetics and its correlation with response and safety in chronic-phase chronic myeloid leukemia: a subanalysis of the IRIS study*. Blood, 2008. **111**(8): p. 4022-8.
68. Sawyers, C.L., *Signal transduction pathways involved in BCR-ABL transformation*. Baillieres Clin Haematol, 1997. **10**(2): p. 223-31.
69. Samanta, A.K., et al., *Jak2 inhibition deactivates Lyn kinase through the SET-PP2A-SHP1 pathway, causing apoptosis in drug-resistant cells from chronic myelogenous leukemia patients*. Oncogene, 2009. **28**(14): p. 1669-81.

70. Danhauser-Riedl, S., et al., *Activation of Src kinases p53/56lyn and p59hck by p210bcr/abl in myeloid cells*. Cancer Res, 1996. **56**(15): p. 3589-96.
71. Warmuth, M., et al., *The Src family kinase Hck interacts with Bcr-Abl by a kinase-independent mechanism and phosphorylates the Grb2-binding site of Bcr*. J Biol Chem, 1997. **272**(52): p. 33260-70.
72. Hu, Y., et al., *Targeting multiple kinase pathways in leukemic progenitors and stem cells is essential for improved treatment of Ph⁺ leukemia in mice*. Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16870-5.
73. Wu, J., et al., *Association between imatinib-resistant BCR-ABL mutation-negative leukemia and persistent activation of LYN kinase*. J Natl Cancer Inst, 2008. **100**(13): p. 926-39.
74. de Lavallade, H., *Chronic myeloid leukaemia*. Medicine, 2013. **41**(5): p. 275-277.
75. Faderl, S., et al., *The biology of chronic myeloid leukemia*. N Engl J Med, 1999. **341**(3): p. 164-72.
76. Deininger, M.W., J.M. Goldman, and J.V. Melo, *The molecular biology of chronic myeloid leukemia*. Blood, 2000. **96**(10): p. 3343-56.
77. Melo, J.V., *The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype*. Blood, 1996. **88**(7): p. 2375-84.
78. Pane, F., et al., *Neutrophilic-chronic myeloid leukemia: a distinct disease with a specific molecular marker (BCR/ABL with C3/A2 junction)*. Blood, 1996. **88**(7): p. 2410-4.
79. Zhang, X., et al., *The NH(2)-terminal coiled-coil domain and tyrosine 177 play important roles in induction of a myeloproliferative disease in mice by Bcr-Abl*. Mol Cell Biol, 2001. **21**(3): p. 840-53.
80. Ren, S.Y., et al., *Phosphatidylinositol 3-kinase p85{alpha} subunit-dependent interaction with BCR/ABL-related fusion tyrosine kinases: molecular mechanisms and biological consequences*. Mol Cell Biol, 2005. **25**(18): p. 8001-8.
81. Sattler, M., et al., *Critical role for Gab2 in transformation by BCR/ABL*. Cancer Cell, 2002. **1**(5): p. 479-92.
82. Skorski, T., et al., *Phosphatidylinositol-3 kinase activity is regulated by BCR/ABL and is required for the growth of Philadelphia chromosome-positive cells*. Blood, 1995. **86**(2): p. 726-36.
83. Dameshek, W., *Some speculations on the myeloproliferative syndromes*. Blood, 1951. **6**(4): p. 372-5.
84. Whang, J., et al., *THE DISTRIBUTION OF THE PHILADELPHIA CHROMOSOME IN PATIENTS WITH CHRONIC MYELOGENOUS LEUKEMIA*. Blood, 1963. **22**: p. 664-73.
85. Bonnet, D. and J.E. Dick, *Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell*. Nat Med, 1997. **3**(7): p. 730-7.
86. Levin, R.H., et al., *PERSISTENT MITOSIS OF TRANSFUSED HOMOLOGOUS LEUKOCYTES IN CHILDREN RECEIVING ANTILEUKEMIC THERAPY*. Science, 1963. **142**(3597): p. 1305-11.
87. Dazzi, F., et al., *Normal and chronic phase CML hematopoietic cells repopulate NOD/SCID bone marrow with different kinetics and cell lineage representation*. Hematol J, 2000. **1**(5): p. 307-15.

88. Petzer, A.L., et al., *Characterization of primitive subpopulations of normal and leukemic cells present in the blood of patients with newly diagnosed as well as established chronic myeloid leukemia*. Blood, 1996. **88**(6): p. 2162-71.
89. Chomel, J.C., et al., *Leukemic stem cell persistence in chronic myeloid leukemia patients with sustained undetectable molecular residual disease*. Blood, 2011. **118**(13): p. 3657-60.
90. Chu, S., et al., *Persistence of leukemia stem cells in chronic myelogenous leukemia patients in prolonged remission with imatinib treatment*. Blood, 2011. **118**(20): p. 5565-72.
91. Rousselot, P., et al., *Loss of major molecular response as a trigger for restarting tyrosine kinase inhibitor therapy in patients with chronic-phase chronic myelogenous leukemia who have stopped imatinib after durable undetectable disease*. J Clin Oncol, 2014. **32**(5): p. 424-30.
92. Corbin, A.S., et al., *Human chronic myeloid leukemia stem cells are insensitive to imatinib despite inhibition of BCR-ABL activity*. J Clin Invest, 2011. **121**(1): p. 396-409.
93. Clevers, H. and R. Nusse, *Wnt/beta-catenin signaling and disease*. Cell, 2012. **149**(6): p. 1192-205.
94. Hu, Y., et al., *beta-Catenin is essential for survival of leukemic stem cells insensitive to kinase inhibition in mice with BCR-ABL-induced chronic myeloid leukemia*. Leukemia, 2009. **23**(1): p. 109-16.
95. Zhao, C., et al., *Loss of beta-catenin impairs the renewal of normal and CML stem cells in vivo*. Cancer Cell, 2007. **12**(6): p. 528-41.
96. Jamieson, C.H., et al., *Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML*. N Engl J Med, 2004. **351**(7): p. 657-67.
97. Duman-Scheel, M., et al., *Hedgehog regulates cell growth and proliferation by inducing Cyclin D and Cyclin E*. Nature, 2002. **417**(6886): p. 299-304.
98. Ingham, P.W. and A.P. McMahon, *Hedgehog signaling in animal development: paradigms and principles*. Genes Dev, 2001. **15**(23): p. 3059-87.
99. Lee, J., et al., *Gli1 is a target of Sonic hedgehog that induces ventral neural tube development*. Development, 1997. **124**(13): p. 2537-52.
100. Dierks, C., et al., *Expansion of Bcr-Abl-positive leukemic stem cells is dependent on Hedgehog pathway activation*. Cancer Cell, 2008. **14**(3): p. 238-49.
101. Calabretta, B. and D. Perrotti, *The biology of CML blast crisis*. Blood, 2004. **103**(11): p. 4010-22.
102. Zhang, S.J., et al., *Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia*. Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2076-81.
103. Quintas-Cardama, A. and J. Cortes, *Molecular biology of bcr-abl-positive chronic myeloid leukemia*. Blood, 2009. **113**(8): p. 1619-30.
104. Waddington, C.H., *The epigenotype. 1942*. Int J Epidemiol, 2012. **41**(1): p. 10-3.
105. Bird, A., *DNA methylation patterns and epigenetic memory*. Genes Dev, 2002. **16**(1): p. 6-21.
106. Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development*. Nature, 2007. **447**(7143): p. 425-32.

107. Cantone, I. and A.G. Fisher, *Epigenetic programming and reprogramming during development*. Nat Struct Mol Biol, 2013. **20**(3): p. 282-9.
108. Meissner, A., *Epigenetic modifications in pluripotent and differentiated cells*. Nat Biotechnol, 2010. **28**(10): p. 1079-88.
109. Rando, T.A. and H.Y. Chang, *Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock*. Cell, 2012. **148**(1-2): p. 46-57.
110. Dinant, C., A.B. Houtsmuller, and W. Vermeulen, *Chromatin structure and DNA damage repair*. Epigenetics Chromatin, 2008. **1**(1): p. 9.
111. Griffith, J.S. and H.R. Mahler, *DNA ticketing theory of memory*. Nature, 1969. **223**(5206): p. 580-2.
112. Riggs, A.D., *X inactivation, differentiation, and DNA methylation*. Cytogenet Cell Genet, 1975. **14**(1): p. 9-25.
113. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development*. Science, 1975. **187**(4173): p. 226-32.
114. Sager, R. and R. Kitchin, *Selective silencing of eukaryotic DNA*. Science, 1975. **189**(4201): p. 426-33.
115. Holliday, R., *A new theory of carcinogenesis*. Br J Cancer, 1979. **40**(4): p. 513-22.
116. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer*. Cell, 2007. **128**(4): p. 683-92.
117. Baylin, S.B., et al., *Alterations in DNA methylation: a fundamental aspect of neoplasia*. Adv Cancer Res, 1998. **72**: p. 141-96.
118. Costello, J.F., et al., *Aberrant CpG-island methylation has non-random and tumour-type-specific patterns*. Nat Genet, 2000. **24**(2): p. 132-8.
119. Esteller, M., *Epigenetics in cancer*. N Engl J Med, 2008. **358**(11): p. 1148-59.
120. Herman, J.G. and S.B. Baylin, *Gene silencing in cancer in association with promoter hypermethylation*. N Engl J Med, 2003. **349**(21): p. 2042-54.
121. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome*. Nat Genet, 2007. **39**(4): p. 457-66.
122. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences*. Nature, 2009. **462**(7271): p. 315-22.
123. Ramsahoye, B.H., et al., *Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a*. Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5237-42.
124. Yan, J., J.R. Zierath, and R. Barres, *Evidence for non-CpG methylation in mammals*. Exp Cell Res, 2011. **317**(18): p. 2555-61.
125. Guo, J.U., et al., *Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain*. Nat Neurosci, 2014. **17**(2): p. 215-22.
126. White, G.P., et al., *Differential patterns of methylation of the IFN-gamma promoter at CpG and non-CpG sites underlie differences in IFN-gamma gene expression between human neonatal and adult CD45RO- T cells*. J Immunol, 2002. **168**(6): p. 2820-7.
127. Bird, A.P., *CpG-rich islands and the function of DNA methylation*. Nature, 1986. **321**(6067): p. 209-13.
128. Bestor, T.H., *Transposons reanimated in mice*. Cell, 2005. **122**(3): p. 322-5.
129. Xu, G.L., et al., *Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene*. Nature, 1999. **402**(6758): p. 187-91.

130. Wigler, M., D. Levy, and M. Perucho, *The somatic replication of DNA methylation*. Cell, 1981. **24**(1): p. 33-40.
131. Stein, R., et al., *Clonal inheritance of the pattern of DNA methylation in mouse cells*. Proc Natl Acad Sci U S A, 1982. **79**(1): p. 61-5.
132. Bestor, T., et al., *Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases*. J Mol Biol, 1988. **203**(4): p. 971-83.
133. Bestor, T.H., *The DNA methyltransferases of mammals*. Hum Mol Genet, 2000. **9**(16): p. 2395-402.
134. Robertson, K.D., et al., *The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors*. Nucleic Acids Res, 1999. **27**(11): p. 2291-8.
135. Jackson-Grusby, L., et al., *Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation*. Nat Genet, 2001. **27**(1): p. 31-9.
136. Lei, H., et al., *De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells*. Development, 1996. **122**(10): p. 3195-205.
137. Chedin, F., *The DNMT3 family of mammalian de novo DNA methyltransferases*. Prog Mol Biol Transl Sci, 2011. **101**: p. 255-85.
138. Okano, M., S. Xie, and E. Li, *Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases*. Nat Genet, 1998. **19**(3): p. 219-20.
139. Yang, Y., et al., *CRL4B promotes tumorigenesis by coordinating with SUV39H1/HP1/DNMT3A in DNA methylation-based epigenetic silencing*. Oncogene, 2015. **34**(1): p. 104-18.
140. Fatemi, M., et al., *Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA*. Eur J Biochem, 2002. **269**(20): p. 4981-4.
141. Rhee, I., et al., *CpG methylation is maintained in human cancer cells lacking DNMT1*. Nature, 2000. **404**(6781): p. 1003-7.
142. Liao, H.F., et al., *Functions of DNA methyltransferase 3-like in germ cells and beyond*. Biol Cell, 2012. **104**(10): p. 571-87.
143. Gowher, H., et al., *Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L*. J Biol Chem, 2005. **280**(14): p. 13341-8.
144. Jeltsch, A., W. Nellen, and F. Lyko, *Two substrates are better than one: dual specificities for Dnmt2 methyltransferases*. Trends Biochem Sci, 2006. **31**(6): p. 306-8.
145. Alexandrov, A., et al., *Rapid tRNA decay can result from lack of nonessential modifications*. Mol Cell, 2006. **21**(1): p. 87-96.
146. Schaefer, M., et al., *RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage*. Genes Dev, 2010. **24**(15): p. 1590-5.
147. Hajkova, P., et al., *Epigenetic reprogramming in mouse primordial germ cells*. Mech Dev, 2002. **117**(1-2): p. 15-23.
148. Morgan, H.D., et al., *Epigenetic reprogramming in mammals*. Hum Mol Genet, 2005. **14 Spec No 1**: p. R47-58.
149. Ji, H., et al., *Comprehensive methylome map of lineage commitment from haematopoietic progenitors*. Nature, 2010. **467**(7313): p. 338-42.
150. Ito, S., et al., *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification*. Nature, 2010. **466**(7310): p. 1129-33.

151. Song, C.X., et al., *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine*. Nat Biotechnol, 2011. **29**(1): p. 68-72.
152. Globisch, D., et al., *Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates*. PLoS One, 2010. **5**(12): p. e15367.
153. Nabel, C.S., et al., *AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation*. Nat Chem Biol, 2012. **8**(9): p. 751-8.
154. Zhang, L., et al., *Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA*. Nat Chem Biol, 2012. **8**(4): p. 328-30.
155. Cortellino, S., et al., *Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair*. Cell, 2011. **146**(1): p. 67-79.
156. He, Y.F., et al., *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA*. Science, 2011. **333**(6047): p. 1303-7.
157. Docherty, L.E., et al., *Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes*. J Med Genet, 2014. **51**(4): p. 229-38.
158. Kornberg, R.D. and J.O. Thomas, *Chromatin structure; oligomers of the histones*. Science, 1974. **184**(4139): p. 865-8.
159. Oudet, P., M. Gross-Bellard, and P. Chambon, *Electron microscopic and biochemical evidence that chromatin structure is a repeating unit*. Cell, 1975. **4**(4): p. 281-300.
160. Eickbush, T.H. and E.N. Moudrianakis, *The histone core complex: an octamer assembled by two sets of protein-protein interactions*. Biochemistry, 1978. **17**(23): p. 4955-64.
161. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. Nature, 1997. **389**(6648): p. 251-60.
162. Whitlock, J.P., Jr. and R.T. Simpson, *Removal of histone H1 exposes a fifty base pair DNA segment between nucleosomes*. Biochemistry, 1976. **15**(15): p. 3307-14.
163. Bednar, J., et al., *Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin*. Proc Natl Acad Sci U S A, 1998. **95**(24): p. 14173-8.
164. van Steensel, B., *Chromatin: constructing the big picture*. EMBO J, 2011. **30**(10): p. 1885-95.
165. Grewal, S.I. and S. Jia, *Heterochromatin revisited*. Nat Rev Genet, 2007. **8**(1): p. 35-46.
166. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
167. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. Nature, 2000. **403**(6765): p. 41-5.
168. Cosgrove, M.S., J.D. Boeke, and C. Wolberger, *Regulated nucleosome mobility and the histone code*. Nat Struct Mol Biol, 2004. **11**(11): p. 1037-43.
169. Ye, J., et al., *Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly*. Mol Cell, 2005. **18**(1): p. 123-30.
170. Fenley, A.T., D.A. Adams, and A.V. Onufriev, *Charge state of the globular histone core controls stability of the nucleosome*. Biophys J, 2010. **99**(5): p. 1577-85.

171. Lee, J.S., E. Smith, and A. Shilatifard, *The language of histone crosstalk*. Cell, 2010. **142**(5): p. 682-5.
172. Bernstein, B.E., et al., *Methylation of histone H3 Lys 4 in coding regions of active genes*. Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8695-700.
173. Liang, G., et al., *Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome*. Proc Natl Acad Sci U S A, 2004. **101**(19): p. 7357-62.
174. Bannister, A.J., et al., *Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes*. J Biol Chem, 2005. **280**(18): p. 17732-6.
175. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast*. Cell, 2005. **122**(4): p. 517-27.
176. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing*. Science, 2002. **298**(5595): p. 1039-43.
177. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain*. Nature, 2001. **410**(6824): p. 120-4.
178. Kourmouli, N., et al., *Heterochromatin and tri-methylated lysine 20 of histone H4 in animals*. J Cell Sci, 2004. **117**(Pt 12): p. 2491-501.
179. Peltomaki, P., *Mutations and epimutations in the origin of cancer*. Exp Cell Res, 2012. **318**(4): p. 299-310.
180. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts*. Nature, 1983. **301**(5895): p. 89-92.
181. Esteller, M., *Cancer epigenomics: DNA methylomes and histone-modification maps*. Nat Rev Genet, 2007. **8**(4): p. 286-98.
182. Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells*. Nat Genet, 2005. **37**(8): p. 853-62.
183. Ng, H.H. and A. Bird, *DNA methylation and chromatin modification*. Curr Opin Genet Dev, 1999. **9**(2): p. 158-63.
184. Garrick, D., et al., *Repeat-induced gene silencing in mammals*. Nat Genet, 1998. **18**(1): p. 56-9.
185. Shann, Y.J., et al., *Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines*. Genome Res, 2008. **18**(5): p. 791-801.
186. Nishigaki, M., et al., *Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays*. Cancer Res, 2005. **65**(6): p. 2115-24.
187. Walsh, C.P., J.R. Chaillet, and T.H. Bestor, *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation*. Nat Genet, 1998. **20**(2): p. 116-7.
188. Eden, A., et al., *Chromosomal instability and tumors promoted by DNA hypomethylation*. Science, 2003. **300**(5618): p. 455.
189. Fraga, M.F., et al., *A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors*. Cancer Res, 2004. **64**(16): p. 5527-34.
190. Sassaman, D.M., et al., *Many human L1 elements are capable of retrotransposition*. Nat Genet, 1997. **16**(1): p. 37-43.
191. Whitelaw, E. and D.I. Martin, *Retrotransposons as epigenetic mediators of phenotypic variation in mammals*. Nat Genet, 2001. **27**(4): p. 361-5.

192. Dante, R., et al., *Methylation patterns of long interspersed repeated DNA and alphoid repetitive DNA from human cell lines and tumors*. Anticancer Res, 1992. **12**(2): p. 559-63.
193. Jurgens, B., B.J. Schmitz-Drager, and W.A. Schulz, *Hypomethylation of L1 LINE sequences prevailing in human urothelial carcinoma*. Cancer Res, 1996. **56**(24): p. 5698-703.
194. Santourlidis, S., et al., *High frequency of alterations in DNA methylation in adenocarcinoma of the prostate*. Prostate, 1999. **39**(3): p. 166-74.
195. Casau, A.E., et al., *Germ cell expression of an isolated human endogenous retroviral long terminal repeat of the HERV-K/HTDV family in transgenic mice*. J Virol, 1999. **73**(12): p. 9976-83.
196. Gotzinger, N., et al., *Regulation of human endogenous retrovirus-K Gag expression in teratocarcinoma cell lines and human tumours*. J Gen Virol, 1996. **77** (Pt 12): p. 2983-90.
197. Wang-Johanning, F., et al., *Expression of human endogenous retrovirus k envelope transcripts in human breast cancer*. Clin Cancer Res, 2001. **7**(6): p. 1553-60.
198. Li, Y., G. Meng, and Q.N. Guo, *Changes in genomic imprinting and gene expression associated with transformation in a model of human osteosarcoma*. Exp Mol Pathol, 2008. **84**(3): p. 234-9.
199. Cerrato, F., et al., *Different mechanisms cause imprinting defects at the IGF2/H19 locus in Beckwith-Wiedemann syndrome and Wilms' tumour*. Hum Mol Genet, 2008. **17**(10): p. 1427-35.
200. Feinberg, A.P., *Imprinting of a genomic domain of 11p15 and loss of imprinting in cancer: an introduction*. Cancer Res, 1999. **59**(7 Suppl): p. 1743s-1746s.
201. Feng, W., et al., *Correlation between CpG methylation profiles and hormone receptor status in breast cancers*. Breast Cancer Res, 2007. **9**(4): p. R57.
202. Feng, W., et al., *Imprinted tumor suppressor genes ARHI and PEG3 are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation*. Cancer, 2008. **112**(7): p. 1489-502.
203. Cui, H., et al., *Loss of IGF2 imprinting: a potential marker of colorectal cancer risk*. Science, 2003. **299**(5613): p. 1753-5.
204. Kaneda, A. and A.P. Feinberg, *Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk*. Cancer Res, 2005. **65**(24): p. 11236-40.
205. Ito, Y., et al., *Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer*. Hum Mol Genet, 2008. **17**(17): p. 2633-43.
206. Yin, D., et al., *Methylation, expression, and mutation analysis of the cell cycle control genes in human brain tumors*. Oncogene, 2002. **21**(54): p. 8372-8.
207. Gopisetty, G., K. Ramachandran, and R. Singal, *DNA methylation and apoptosis*. Mol Immunol, 2006. **43**(11): p. 1729-40.
208. Lahtz, C. and G.P. Pfeifer, *Epigenetic changes of DNA repair genes in cancer*. J Mol Cell Biol, 2011. **3**(1): p. 51-8.
209. van Vlodrop, I.J., et al., *Prognostic significance of Gremlin1 (GREM1) promoter CpG island hypermethylation in clear cell renal cell carcinoma*. Am J Pathol, 2010. **176**(2): p. 575-84.
210. Saito, T., et al., *Hypermethylation in promoter region of E-cadherin gene is associated with tumor dedifferentiation and myometrial invasion in endometrial carcinoma*. Cancer, 2003. **97**(4): p. 1002-9.

211. Das, P.M. and R. Singal, *DNA methylation and cancer*. J Clin Oncol, 2004. **22**(22): p. 4632-42.
212. Lee, W.H., et al., *Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis*. Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11733-7.
213. Melki, J.R., P.C. Vincent, and S.J. Clark, *Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia*. Cancer Res, 1999. **59**(15): p. 3730-40.
214. Eads, C.A., A.E. Nickel, and P.W. Laird, *Complete genetic suppression of polyp formation and reduction of CpG-island hypermethylation in Apc(Min/+) Dnmt1-hypomorphic Mice*. Cancer Res, 2002. **62**(5): p. 1296-9.
215. Rhee, I., et al., *DNMT1 and DNMT3b cooperate to silence genes in human cancer cells*. Nature, 2002. **416**(6880): p. 552-6.
216. Rountree, M.R., K.E. Bachman, and S.B. Baylin, *DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci*. Nat Genet, 2000. **25**(3): p. 269-77.
217. Fuks, F., et al., *DNA methyltransferase Dnmt1 associates with histone deacetylase activity*. Nat Genet, 2000. **24**(1): p. 88-91.
218. Robertson, K.D., et al., *DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters*. Nat Genet, 2000. **25**(3): p. 338-42.
219. Bachman, K.E., M.R. Rountree, and S.B. Baylin, *Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin*. J Biol Chem, 2001. **276**(34): p. 32282-7.
220. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome*. Cell, 2007. **128**(4): p. 669-81.
221. Fraga, M.F., et al., *Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer*. Nat Genet, 2005. **37**(4): p. 391-400.
222. Yasui, W., et al., *Histone acetylation and gastrointestinal carcinogenesis*. Ann N Y Acad Sci, 2003. **983**: p. 220-31.
223. Jenuwein, T. and C.D. Allis, *Translating the histone code*. Science, 2001. **293**(5532): p. 1074-80.
224. Ballestar, E., et al., *Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer*. EMBO J, 2003. **22**(23): p. 6335-45.
225. Seligson, D.B., et al., *Global histone modification patterns predict risk of prostate cancer recurrence*. Nature, 2005. **435**(7046): p. 1262-6.
226. Richon, V.M., et al., *Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation*. Proc Natl Acad Sci U S A, 2000. **97**(18): p. 10014-9.
227. Gui, C.Y., et al., *Histone deacetylase (HDAC) inhibitor activation of p21WAF1 involves changes in promoter-associated proteins, including HDAC1*. Proc Natl Acad Sci U S A, 2004. **101**(5): p. 1241-6.
228. Brehm, A., et al., *Retinoblastoma protein recruits histone deacetylase to repress transcription*. Nature, 1998. **391**(6667): p. 597-601.
229. Luo, J., et al., *Acetylation of p53 augments its site-specific DNA binding both in vitro and in vivo*. Proc Natl Acad Sci U S A, 2004. **101**(8): p. 2259-64.
230. Luo, R.X., A.A. Postigo, and D.C. Dean, *Rb interacts with histone deacetylase to repress transcription*. Cell, 1998. **92**(4): p. 463-73.

231. Chen, L., et al., *Duration of nuclear NF-kappaB action regulated by reversible acetylation*. Science, 2001. **293**(5535): p. 1653-7.
232. Cohen, H.Y., et al., *Acetylation of the C terminus of Ku70 by CBP and PCAF controls Bax-mediated apoptosis*. Mol Cell, 2004. **13**(5): p. 627-38.
233. Scroggins, B.T., et al., *An acetylation site in the middle domain of Hsp90 regulates chaperone function*. Mol Cell, 2007. **25**(1): p. 151-9.
234. Zion, M., et al., *Progressive de novo DNA methylation at the bcr-abl locus in the course of chronic myelogenous leukemia*. Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10722-6.
235. Ben-Yehuda, D., et al., *Molecular follow-up of disease progression and interferon therapy in chronic myelocytic leukemia*. Blood, 1997. **90**(12): p. 4918-23.
236. Asimakopoulou, F.A., et al., *ABL1 methylation is a distinct molecular event associated with clonal evolution of chronic myeloid leukemia*. Blood, 1999. **94**(7): p. 2452-60.
237. Issa, J.P., et al., *Methylation of the ABL1 promoter in chronic myelogenous leukemia: lack of prognostic significance*. Blood, 1999. **93**(6): p. 2075-80.
238. Stratthdee, G., et al., *Inactivation of HOXA genes by hypermethylation in myeloid and lymphoid malignancy is frequent and associated with poor prognosis*. Clin Cancer Res, 2007. **13**(17): p. 5048-55.
239. Cillo, C., et al., *Homeobox genes in normal and malignant cells*. J Cell Physiol, 2001. **188**(2): p. 161-9.
240. Mills, K.I., et al., *Increasing methylation of the calcitonin gene during disease progression in sequential samples from CML patients*. Leuk Res, 1996. **20**(9): p. 771-5.
241. Issa, J.P., et al., *HIC1 hypermethylation is a late event in hematopoietic neoplasms*. Cancer Res, 1997. **57**(9): p. 1678-81.
242. Issa, J.P., et al., *The estrogen receptor CpG island is methylated in most hematopoietic neoplasms*. Cancer Res, 1996. **56**(5): p. 973-77.
243. Roman-Gomez, J., et al., *Epigenetic regulation of PRAME gene in chronic myeloid leukemia*. Leuk Res, 2007. **31**(11): p. 1521-8.
244. Nguyen, T.T., et al., *Quantitative measure of c-abl and p15 methylation in chronic myelogenous leukemia: biological implications*. Blood, 2000. **95**(9): p. 2990-2.
245. Nagy, E., et al., *Frequent methylation of p16INK4A and p14ARF genes implicated in the evolution of chronic myeloid leukaemia from its chronic to accelerated phase*. Eur J Cancer, 2003. **39**(16): p. 2298-305.
246. Dunwell, T., et al., *A genome-wide screen identifies frequently methylated genes in haematological and epithelial cancers*. Mol Cancer, 2010. **9**: p. 44.
247. Jelinek, J., et al., *Aberrant DNA methylation is associated with disease progression, resistance to imatinib and shortened survival in chronic myelogenous leukemia*. PLoS One, 2011. **6**(7): p. e22110.
248. Labi, V., et al., *Haematopoietic stem cell survival and transplantation efficacy is limited by the BH3-only proteins Bim and Bmf*. EMBO Mol Med, 2013. **5**(1): p. 122-36.
249. Luciano, F., et al., *Phosphorylation of Bim-EL by Erk1/2 on serine 69 promotes its degradation via the proteasome pathway and regulates its proapoptotic function*. Oncogene, 2003. **22**(43): p. 6785-93.

250. San Jose-Eneriz, E., et al., *Epigenetic down-regulation of BIM expression is associated with reduced optimal responses to imatinib treatment in chronic myeloid leukaemia*. Eur J Cancer, 2009. **45**(10): p. 1877-89.
251. Leo, E., et al., *DNA hypermethylation promotes the low expression of pro-apoptotic BCL2L1 associated with BCR-ABL1 fusion gene of chronic myeloid leukaemia*. Br J Haematol, 2012. **159**(3): p. 373-6.
252. Kuroda, J., et al., *Bim and Bad mediate imatinib-induced killing of Bcr/Abl+ leukemic cells, and resistance due to their loss is overcome by a BH3 mimetic*. Proc Natl Acad Sci U S A, 2006. **103**(40): p. 14907-12.
253. Mancini, M., et al., *Cytoplasmatic compartmentalization by Bcr-Abl promotes TET2 loss-of-function in chronic myeloid leukemia*. J Cell Biochem, 2012. **113**(8): p. 2765-74.
254. Ghoshal, K., et al., *5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal*. Mol Cell Biol, 2005. **25**(11): p. 4727-41.
255. Stresemann, C. and F. Lyko, *Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine*. Int J Cancer, 2008. **123**(1): p. 8-13.
256. Issa, J.P., et al., *Phase II study of low-dose decitabine in patients with chronic myelogenous leukemia resistant to imatinib mesylate*. J Clin Oncol, 2005. **23**(17): p. 3948-56.
257. Oki, Y., et al., *Phase II study of low-dose decitabine in combination with imatinib mesylate in patients with accelerated or myeloid blastic phase of chronic myelogenous leukemia*. Cancer, 2007. **109**(5): p. 899-906.
258. Song, M.S., L. Salmena, and P.P. Pandolfi, *The functions and regulation of the PTEN tumour suppressor*. Nat Rev Mol Cell Biol, 2012. **13**(5): p. 283-96.
259. Li, J., et al., *PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer*. Science, 1997. **275**(5308): p. 1943-7.
260. Tang, J.M., et al., *Phosphorylated Akt overexpression and loss of PTEN expression in non-small cell lung cancer confers poor prognosis*. Lung Cancer, 2006. **51**(2): p. 181-91.
261. Dahia, P.L., et al., *PTEN is inversely correlated with the cell survival factor Akt/PKB and is inactivated via multiple mechanisms in haematological malignancies*. Hum Mol Genet, 1999. **8**(2): p. 185-93.
262. Nishioka, C., et al., *Imatinib causes epigenetic alterations of PTEN gene via upregulation of DNA methyltransferases and polycomb group proteins*. Blood Cancer J, 2011. **1**(12): p. e48.
263. Nishioka, C., et al., *Long-term exposure of leukemia cells to multi-targeted tyrosine kinase inhibitor induces activations of AKT, ERK and STAT5 signaling via epigenetic silencing of the PTEN gene*. Leukemia, 2010. **24**(9): p. 1631-40.
264. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions*. Cell, 2009. **136**(2): p. 215-33.
265. Neviani, P., et al., *FTY720, a new alternative for treating blast crisis chronic myelogenous leukemia and Philadelphia chromosome-positive acute lymphocytic leukemia*. J Clin Invest, 2007. **117**(9): p. 2408-21.
266. Bueno, M.J., et al., *Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression*. Cancer Cell, 2008. **13**(6): p. 496-506.

267. Croce, C.M., *Causes and consequences of microRNA dysregulation in cancer*. Nat Rev Genet, 2009. **10**(10): p. 704-14.
268. Chen, C.Z., et al., *MicroRNAs modulate hematopoietic lineage differentiation*. Science, 2004. **303**(5654): p. 83-6.
269. Eiring, A.M., et al., *miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts*. Cell, 2010. **140**(5): p. 652-65.
270. Guo, S. and D.T. Scadden, *A microRNA regulating adult hematopoietic stem cells*. Cell Cycle, 2010. **9**(18): p. 3637-8.
271. Agirre, X., et al., *Down-regulation of hsa-miR-10a in chronic myeloid leukemia CD34+ cells increases USF2-mediated cell growth*. Mol Cancer Res, 2008. **6**(12): p. 1830-40.
272. Venturini, L., et al., *Expression of the miR-17-92 polycistron in chronic myeloid leukemia (CML) CD34+ cells*. Blood, 2007. **109**(10): p. 4399-405.
273. Grafodatskaya, D., et al., *EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines*. Genomics, 2010. **95**(2): p. 73-83.
274. Vilahur, N., et al., *Storage conditions and stability of global DNA methylation in placental tissue*. Epigenomics, 2013. **5**(3): p. 341-8.
275. Hakansson, P., et al., *Establishment and phenotypic characterization of human U937 cells with inducible P210 BCR/ABL expression reveals upregulation of CEACAM1 (CD66a)*. Leukemia, 2004. **18**(3): p. 538-47.
276. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution*. Genomics, 2011. **98**(4): p. 288-95.
277. Sandoval, J., et al., *Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome*. Epigenetics, 2011. **6**(6): p. 692-702.
278. Benjamini, Y.a.H., Y, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society, 1995. **57**(1): p. 289-300.
279. *The R Project for Statistical Computing* www.r-project.org/.
280. minfi- Bioconductor
www.bioconductor.org/packages/release/bioc/html/minfi.html.
281. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis*. BMC Bioinformatics, 2010. **11**: p. 587.
282. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Res, 2015. **43**(7): p. e47.
283. *The UCSC Genome Browser* www.genome.ucsc.edu/.
284. Kent, W.J., *BLAT--the BLAST-like alignment tool*. Genome Res, 2002. **12**(4): p. 656-64.
285. Huang, D.W., et al., *The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists*. Genome Biol, 2007. **8**(9): p. R183.
286. KEGG: Kyoto Encyclopedia of Genes and Genomes
<http://www.genome.jp/kegg/>.
287. *The Gene Ontology Consortium* <http://geneontology.org/>.
288. MARMAL-AID <http://marmal-aid.org/>.
289. Lowe, R. and V.K. Rakyan, *Marmal-aid--a database for Infinium HumanMethylation450*. BMC Bioinformatics, 2013. **14**: p. 359.
290. *The Cancer Genome Atlas (TCGA)* <http://cancergenome.nih.gov/>.

291. Welch, J.S., et al., *The origin and evolution of mutations in acute myeloid leukemia*. Cell, 2012. **150**(2): p. 264-78.
292. Jan, M., et al., *Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia*. Sci Transl Med, 2012. **4**(149): p. 149ra118.
293. Figueroa, M.E., et al., *DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia*. Cancer Cell, 2010. **17**(1): p. 13-27.
294. Kulis, M., et al., *Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia*. Nat Genet, 2012. **44**(11): p. 1236-42.
295. Drummond, M.W., et al., *Mobilization of Ph chromosome-negative peripheral blood stem cells in chronic myeloid leukaemia patients with imatinib mesylate-induced complete cytogenetic remission*. Br J Haematol, 2003. **123**(3): p. 479-83.
296. Maksimovic, J., L. Gordon, and A. Oshlack, *SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips*. Genome Biol, 2012. **13**(6): p. R44.
297. Naeem, H., et al., *Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array*. BMC Genomics, 2014. **15**: p. 51.
298. Illumina. *Infinium HD Methylation SNP List*
http://support.illumina.com/downloads/infinium_hd_methylation_snp_list.html. 2013.
299. Ward, J., *Hierarchical Grouping to Optimize an Objective Function*. Journal of the American Statistical Association, 1963. **58**(301): p. pp. 236-244.
300. Bjornsson, H.T., et al., *Intra-individual change over time in DNA methylation with familial clustering*. JAMA, 2008. **299**(24): p. 2877-83.
301. Boks, M.P., et al., *The relationship of DNA methylation with age, gender and genotype in twins and healthy controls*. PLoS One, 2009. **4**(8): p. e6767.
302. *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia*. N Engl J Med, 2013. **368**(22): p. 2059-74.
303. Teschendorff, A.E., et al., *An epigenetic signature in peripheral blood predicts active ovarian cancer*. PLoS One, 2009. **4**(12): p. e8274.
304. Leek, J.T., et al., *Tackling the widespread and critical impact of batch effects in high-throughput data*. Nat Rev Genet, 2010. **11**(10): p. 733-9.
305. Ciriello, G., et al., *Emerging landscape of oncogenic signatures across human cancers*. Nat Genet, 2013. **45**(10): p. 1127-33.
306. Lin, Q. and W. Wagner, *Epigenetic Aging Signatures Are Coherently Modified in Cancer*. PLoS Genet, 2015. **11**(6): p. e1005334.
307. Allison E Ashley-Koch, P., *Hydroxyurea Induces Genome-Wide Epigenetic Changes In Sickle Cell Disease*
, in *AMERICAN SOCIETY OF HEMATOLOGY*2010: Orlando, FL, USA.
308. Shah, N.P., et al., *Overriding imatinib resistance with a novel ABL kinase inhibitor*. Science, 2004. **305**(5682): p. 399-401.
309. Talpaz, M., et al., *Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias*. N Engl J Med, 2006. **354**(24): p. 2531-41.
310. O'Brien, S., C. Hedgley, and L. Foroni, *SPIRIT 2: AN NCRI RANDOMISED STUDY COMPARING DASATINIB WITH IMATINIB IN PATIENTS WITH*

- NEWLY-DIAGNOSED CHRONIC MYELOID LEUKAEMIA- 2 YEAR FOLLOW UP, in *European Hematology Association* 2015: Milan, Italy.
311. Brdickova, N., et al., *LIME: a new membrane Raft-associated adaptor protein involved in CD4 and CD8 coreceptor signaling*. J Exp Med, 2003. **198**(10): p. 1453-62.
 312. Hur, E.M., et al., *LIME, a novel transmembrane adaptor protein, associates with p56lck and mediates T cell activation*. J Exp Med, 2003. **198**(10): p. 1463-73.
 313. Ahn, E., H. Lee, and Y. Yun, *LIME acts as a transmembrane adapter mediating BCR-dependent B-cell activation*. Blood, 2006. **107**(4): p. 1521-7.
 314. de Lavallade, H., et al., *Tyrosine kinase inhibitors impair B-cell immune responses in CML through off-target inhibition of kinases important for cell signaling*. Blood, 2013. **122**(2): p. 227-38.
 315. Maunakea, A.K., et al., *Conserved role of intragenic DNA methylation in regulating alternative promoters*. Nature, 2010. **466**(7303): p. 253-7.
 316. Wolff, E.M., et al., *Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer*. PLoS Genet, 2010. **6**(4): p. e1000917.
 317. Rakyan, V.K., et al., *Epigenome-wide association studies for common human diseases*. Nat Rev Genet, 2011. **12**(8): p. 529-41.
 318. Yung, R., et al., *Mechanisms of drug-induced lupus. II. T cells overexpressing lymphocyte function-associated antigen 1 become autoreactive and cause a lupuslike disease in syngeneic mice*. J Clin Invest, 1996. **97**(12): p. 2866-71.
 319. Badia, E., et al., *Rapid tamoxifen-induced inactivation of an estrogenic response is accompanied by a localized epigenetic modification but not by mutations*. Breast Cancer Res Treat, 1998. **47**(1): p. 71-81.
 320. Badia, E., et al., *Long-term hydroxytamoxifen treatment of an MCF-7-derived breast cancer cell line irreversibly inhibits the expression of estrogenic genes through chromatin remodeling*. Cancer Res, 2000. **60**(15): p. 4130-8.
 321. Pegg, A.E., M.E. Dolan, and R.C. Moschel, *Structure, function, and inhibition of O6-alkylguanine-DNA alkyltransferase*. Prog Nucleic Acid Res Mol Biol, 1995. **51**: p. 167-223.
 322. Qian, X.C. and T.P. Brent, *Methylation hot spots in the 5' flanking region denote silencing of the O6-methylguanine-DNA methyltransferase gene*. Cancer Res, 1997. **57**(17): p. 3672-7.
 323. Watts, G.S., et al., *Methylation of discrete regions of the O6-methylguanine DNA methyltransferase (MGMT) CpG island is associated with heterochromatinization of the MGMT transcription start site and silencing of the gene*. Mol Cell Biol, 1997. **17**(9): p. 5612-9.
 324. Sanchez-Cespedes, M., et al., *Gene promoter hypermethylation in tumors and serum of head and neck cancer patients*. Cancer Res, 2000. **60**(4): p. 892-5.
 325. Nagler, A., et al., *Granulocyte colony-stimulating factor generates epigenetic and genetic alterations in lymphocytes of normal volunteer donors of stem cells*. Exp Hematol, 2004. **32**(1): p. 122-30.
 326. Giehl, M., et al., *Expression of the p210BCR-ABL oncoprotein drives centrosomal hypertrophy and clonal evolution in human U937 cells*. Leukemia, 2007. **21**(9): p. 1971-6.
 327. Sundstrom, C. and K. Nilsson, *Establishment and characterization of a human histiocytic lymphoma cell line (U-937)*. Int J Cancer, 1976. **17**(5): p. 565-77.

328. Wong, N.C., et al., *Exploring the utility of human DNA methylation arrays for profiling mouse genomic DNA*. Genomics, 2013. **102**(1): p. 38-46.
329. Palacios, R. and M. Steinmetz, *Il-3-dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes in vivo*. Cell, 1985. **41**(3): p. 727-34.
330. White, D., et al., *In vitro sensitivity to imatinib-induced inhibition of ABL kinase activity is predictive of molecular response in patients with de novo CML*. Blood, 2005. **106**(7): p. 2520-6.
331. Hamilton, A., et al., *BCR-ABL activity and its response to drugs can be determined in CD34+ CML stem cells by CrkL phosphorylation status using flow cytometry*. Leukemia, 2006. **20**(6): p. 1035-9.
332. Amabile, G., et al., *Dissecting the role of aberrant DNA methylation in human leukaemia*. Nat Commun, 2015. **6**: p. 7091.
333. Kim, K., et al., *Epigenetic memory in induced pluripotent stem cells*. Nature, 2010. **467**(7313): p. 285-90.
334. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells*. Nature, 2011. **471**(7336): p. 68-73.
335. Carette, J.E., et al., *Generation of iPSCs from cultured human malignant cells*. Blood, 2010. **115**(20): p. 4039-42.
336. Koschmieder, S., et al., *Inducible chronic phase of myeloid leukemia with expansion of hematopoietic stem cells in a transgenic model of BCR-ABL leukemogenesis*. Blood, 2005. **105**(1): p. 324-34.
337. Bullinger, L., et al., *Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia*. N Engl J Med, 2004. **350**(16): p. 1605-16.
338. Decaux, O., et al., *Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the Intergrroupe Francophone du Myelome*. J Clin Oncol, 2008. **26**(29): p. 4798-805.
339. Kronenwett, R., et al., *Distinct molecular phenotype of malignant CD34(+) hematopoietic stem and progenitor cells in chronic myelogenous leukemia*. Oncogene, 2005. **24**(34): p. 5313-24.
340. Neumann, F., et al., *Gene expression profiling of Philadelphia chromosome (Ph)-negative CD34+ hematopoietic stem and progenitor cells of patients with Ph-positive CML in major molecular remission during therapy with imatinib*. Leukemia, 2005. **19**(3): p. 458-60.
341. Ohmine, K., et al., *Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells*. Oncogene, 2001. **20**(57): p. 8249-57.
342. Yong, A.S., et al., *Molecular profiling of CD34+ cells identifies low expression of CD7, along with high expression of proteinase 3 or elastase, as predictors of longer survival in patients with CML*. Blood, 2006. **107**(1): p. 205-12.
343. Zheng, C., et al., *Gene expression profiling of CD34+ cells identifies a molecular signature of chronic myeloid leukemia blast crisis*. Leukemia, 2006. **20**(6): p. 1028-34.
344. Diaz-Blanco, E., et al., *Molecular signature of CD34(+) hematopoietic stem and progenitor cells of patients with CML in chronic phase*. Leukemia, 2007. **21**(3): p. 494-504.

345. McWeeney, S.K., et al., *A gene expression signature of CD34+ cells to predict major cytogenetic response in chronic-phase chronic myeloid leukemia patients treated with imatinib*. Blood, 2010. **115**(2): p. 315-25.
346. Bruns, I., et al., *The hematopoietic stem cell in chronic phase CML is characterized by a transcriptional profile resembling normal myeloid progenitor cells and reflecting loss of quiescence*. Leukemia, 2009. **23**(5): p. 892-9.
347. Graham, S.M., et al., *Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal and chronic myeloid leukemia sources*. Stem Cells, 2007. **25**(12): p. 3111-20.
348. Yang, M.Y., et al., *JunB gene expression is inactivated by methylation in chronic myeloid leukemia*. Blood, 2003. **101**(8): p. 3205-11.
349. Roman-Gomez, J., et al., *Epigenetic regulation of human cancer/testis antigen gene, HAGE, in chronic myeloid leukemia*. Haematologica, 2007. **92**(2): p. 153-62.
350. Illumina. *Whole-Genome Gene Expression Direct Hybridization Assay Guide* https://support.illumina.com/content/dam/illumina-support/documents/myillumina/3466bf71-78bd-4842-8bfc-393a45d11874/wggex_direct_hybridization_assay_guide_11322355_a.pdf. June 2010.
351. Kuhn, K., et al., *A novel, high-performance random array platform for quantitative gene expression profiling*. Genome Res, 2004. **14**(11): p. 2347-56.
352. Du, P., W.A. Kibbe, and S.M. Lin, *lumi: a pipeline for processing Illumina microarray*. Bioinformatics, 2008. **24**(13): p. 1547-8.
353. Schmid, R., et al., *Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3*. BMC Genomics, 2010. **11**: p. 349.
354. Brancolini, C., S. Bottega, and C. Schneider, *Gas2, a growth arrest-specific protein, is a component of the microfilament network system*. J Cell Biol, 1992. **117**(6): p. 1251-61.
355. Benetti, R., et al., *The death substrate Gas2 binds m-calpain and increases susceptibility to p53-dependent apoptosis*. EMBO J, 2001. **20**(11): p. 2702-14.
356. Goll, D.E., et al., *The calpain system*. Physiol Rev, 2003. **83**(3): p. 731-801.
357. Pontremoli, S. and E. Melloni, *The role of intracellular proteinases in human neutrophil activation*. Revis Biol Celular, 1989. **20**: p. 161-77.
358. Pontremoli, S., et al., *Biochemical responses in activated human neutrophils mediated by protein kinase C and a Ca²⁺-requiring proteinase*. J Biol Chem, 1986. **261**(18): p. 8309-13.
359. Trost, M., et al., *The phagosomal proteome in interferon-gamma-activated macrophages*. Immunity, 2009. **30**(1): p. 143-54.
360. Benetti, R., et al., *The calpain system is involved in the constitutive regulation of beta-catenin signaling functions*. J Biol Chem, 2005. **280**(23): p. 22070-80.
361. Huang, W., et al., *Interferon consensus sequence binding protein (ICSBP) decreases beta-catenin activity in myeloid cells by repressing GAS2 transcription*. Mol Cell Biol, 2010. **30**(19): p. 4575-94.
362. Janssen, J.J., et al., *Identification of genes potentially involved in disease transformation of CML*. Leukemia, 2005. **19**(6): p. 998-1004.
363. Radich, J.P., et al., *Gene expression changes associated with progression and response in chronic myeloid leukemia*. Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2794-9.

364. Zhou, H., et al., *Growth arrest specific 2 is up-regulated in chronic myeloid leukemia cells and required for their growth*. PLoS One, 2014. **9**(1): p. e86195.
365. Sawyers, C.L., W. Callahan, and O.N. Witte, *Dominant negative MYC blocks transformation by ABL oncogenes*. Cell, 1992. **70**(6): p. 901-10.
366. Alberts, A.S. and R. Treisman, *Activation of RhoA and SAPK/JNK signalling pathways by the RhoA-specific exchange factor mNET1*. EMBO J, 1998. **17**(14): p. 4075-85.
367. Perez, C., et al., *Aberrant DNA methylation profile of chronic and transformed classicphiladelphia-negative myeloproliferative neoplasms*. Haematologica, 2013.
368. Tomasini, R., et al., *TP53INP1s and homeodomain-interacting protein kinase-2 (HIPK2) are partners in regulating p53 activity*. J Biol Chem, 2003. **278**(39): p. 37722-9.
369. Simon, R., M.D. Radmacher, and K. Dobbin, *Design of studies using DNA microarrays*. Genet Epidemiol, 2002. **23**(1): p. 21-36.
370. Gu, C.C., et al., *Role of gene expression microarray analysis in finding complex disease genes*. Genet Epidemiol, 2002. **23**(1): p. 37-56.
371. Steidl, U., et al., *Gene expression profiling identifies significant differences between the molecular phenotypes of bone marrow-derived and circulating human CD34+ hematopoietic stem cells*. Blood, 2002. **99**(6): p. 2037-44.
372. Baez, A., et al., *Granulocyte colony-stimulating factor produces long-term changes in gene and microRNA expression profiles in CD34+ cells from healthy donors*. Haematologica, 2014. **99**(2): p. 243-51.
373. C, K., L. T, and W. D, *A 20 Gene Expression Signature That Predicts Early Molecular Response Failure in Chronic Phase CML Patients Treated with Frontline Imatinib*, in American Society of Hematology 2015: Orlando, FL.
374. Szaniszló, P., et al., *Getting the right cells to the array: Gene expression microarray analysis of cell mixtures and sorted cells*. Cytometry A, 2004. **59**(2): p. 191-202.
375. Futscher, B.W., et al., *Role for DNA methylation in the control of cell type specific maspin expression*. Nat Genet, 2002. **31**(2): p. 175-9.
376. Song, F., et al., *Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression*. Proc Natl Acad Sci U S A, 2005. **102**(9): p. 3336-41.
377. Metivier, R., et al., *Cyclical DNA methylation of a transcriptionally active promoter*. Nature, 2008. **452**(7183): p. 45-50.
378. Sproul, D., et al., *Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns*. Genome Biol, 2012. **13**(10): p. R84.
379. Dawson, M.A., et al., *JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin*. Nature, 2009. **461**(7265): p. 819-22.
380. Brusa, G., et al., *P210 Bcr-abl tyrosine kinase interaction with histone deacetylase 1 modifies histone H4 acetylation and chromatin structure of chronic myeloid leukaemia haematopoietic progenitors*. Br J Haematol, 2006. **132**(3): p. 359-69.
381. Zhang, B., et al., *Effective targeting of quiescent chronic myelogenous leukemia stem cells by histone deacetylase inhibitors in combination with imatinib mesylate*. Cancer Cell, 2010. **17**(5): p. 427-42.

- 382. Mizuno, S., et al., *Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia*. Blood, 2001. **97**(5): p. 1172-9.
- 383. Ernst, T., et al., *Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders*. Nat Genet, 2010. **42**(8): p. 722-6.
- 384. Grossmann, V., et al., *Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in >80% of patients with TET2 and EZH2 being of high prognostic relevance*. Leukemia, 2011. **25**(5): p. 877-9.