Characterising the cell biology of leukemic stem cells in acute myeloid leukemia.

Terri Cornforth
Green Templeton College
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2013

Abstract

Acute Myeloid leukemia (AML) is an aggressive haematological malignancy that mainly affects the elderly. Relapse is common and is thought to be due to the presence of chemotherapy resistant leukemic stem cells (LSC). Within the CD34+ disease (>5% of the blast cells expressing CD34) , two subtypes have been identified; an LMPP/GMP-like expanded type and a MPP/CMP-like expanded type, the former is the most common, accounting for around 80% of CD34+ AML. Both the GMP-like and LMPP-like expanded populations show LSC activity. To improve our understanding of the disease and gain better insight in to how to develop treatments, the molecular basis of the disease needs to be investigated. I investigated miRNAs in the GMP/LMPP-like expanded AML. miRNAs are small non-coding RNAs involved in the regulation of mRNA. In recent years miRNAs have been shown to be implicated in many different diseases. To investigate the role miRNAs play in AML, miRNA expression was profiled in leukemic and normal bone marrow. Bioinformatic analysis was then used to examine the different miRNA expression profiles between normal and leukemic marrow. Our study showed that miRNAs are dysregulated in AML. miRNAs from the miR-17-92 and its paralogous cluster miR-106b-92 were amongst the miRNAs to be found down regulated in AML As had been seen previously at an mRNA level, on an miRNA level the LSC populations more closely resembled more mature progenitor populations than HSC and MPP populations, however the LSC populations did display an aberrant stem cell-like miRNA signature.

Acknowledgments

Firstly I would like to thank Professor Vyas for the opportunity to work in his laboratory and for his supervision throughout. I would also like to thank Dr Porcher and Professor Patient for their help and advice. A huge thank you to all the members past and present of the Vyas and Porcher labs for your help and guidance in the lab and out.

With special thanks to Nicolas for teaching me about the wonderful world of FACS and French sayings my particular favorite being “and my arse is chicken”. Emanuele for leading me through the complex bioinformatics universe and putting up with my endless questions on the topic, Lynn and Gaetan for their priceless tips and Kate for reading my thesis, I could not have got there without you.

I have made some great friends along the way who have helped me through the good times and the bad and made the journey more enjoyable, tea breaks filled with laughter, tasty treats at cake club and maybe a few too many at the White Hart, so big thanks to Angela, Andy, Julia, Dimple, Elisa, John, Inn Inn, Harry, Chrissy and Jenny. Special thanks to Marina for not just being a great friend but for putting up with me as bench neighbour and not batting an eyelid when I threw things at the computer in frustration. Also thanks to Ali for being a much needed distraction and understanding when I needed the time to work

My family for their love and understanding throughout. Rona Scott and John Cornforth, Mum and Dad for putting up with my whinging and always being there to spur me on, I cannot thank you enough for all you have done for me and would not have made it without you. Matthew and Olivia my brother and sister, for your amazing hugs and for telling everyone that I was a mad scientist. Finally I would like to dedicate this work to the memory of my late beloved grandfather Ronald Moffat; I hope I would have made you proud.

Abbreviations

3’	Three Prime end
5’	Five Prime end
% CV	Coefficient of Variation
°C	Degrees Celsius
3D	Three Dimensional
ALL	Acute lymphoblastic Leuekmia
AML	Acute Myeloid Leukemia
APC	Allophycocyanin
BM	Bone Marrow
BSA	Bovine Serum Albumin
CCR8	Chemokine (C-C motif) Receptor 8
CD	Cluster of Differentiation
cDNA	Complementary DNA
CLL	chronic lymphocytic leukaemia
CMP	Common Myeloid Progenitor
CSC	Cancer stem cell
CT	Cycle Threshold
CV	Coefficient of Variation
CXCR4	C-X-C Chemokine Receptor type 4
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
FAB	French American British classification system
FACS	Fluorescence-Activated Cell Sorting
FCS	Foetal Calf Serum

FE	Feature extraction (software)
FITC	Fluorescein Isothiocyanate
FLT3-L	Fms-related Tyrosine Kinase Factor 3 Ligand
FSC-A	Forward Scatter-Area
G-CSF	Granulocyte Colony Stimulating Factor
GSEA	Gene Set Enrichment Analysis
GMP	Granulocyte-Macrophage progenitor
HSC	Hematopoietic Stem Cell
IMDM	Iscove's Modified Dulbecco's Medium
L-IC	Leukemia initiating cell
Lin	Lineage
LMPP	Lymphoid Primed Multipotent Progenitor
LSC	Leukemic stem cell
MDS	Myelodysplastic Syndrome
MEP	Megakaryocyte-Erythroid progenitor
ml	Millilitre
µl	Microliter
µm	Micrometre
Min	Minute
miRNA	microRNA
mRNA	messenger RNA
MNCs	Mononuclear Cells
Mob-PB	Mobilized Peripheral Blood
MPD	Myeloproliferative disorder
MPP	Multipotent Progenitor
ng	Nanogram
NA	Not Available

ND	Not Done
NR	No Result
PBE	PBS/BSA/EDTA
PBS	Phosphate Buffered Saline
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
PE	Phycoerythrin
PMNCs	Peripheral Blood Mononuclear Cells
qPCR	Quantitative real time PCR
RIN	RNA Integrity Number
RBC	Red Blood Cell
rpm	Rotations Per Minute
RNA	Ribonucleic Acid
RMA	Robust Multi-array average
RPMI	Roswell park memorial institute medium
RT	Reverse transcription
SCID	Severe combined immunodeficiency
SCF	Stem Cell Factor
snRNA	Small Nuclear RNA
snoRNA	Small Nucleolar RNA
SORP	Special Order Research Product
SSC-A	Side Scatter-Area
TPO	Thrombopoietin
WHO	World Health Organization
VC	Vacuum Concentrate

Table of Contents

Chapter 1: Introduction

1.1 Hematopoiesis

Haematopoiesis is the process of blood cell development; it generates all blood cells needed throughout the life span of an individual from hematopoietic stem cells (HSC). The general model of hematopoiesis is a hierarchy with the self-renewing HSC at the apex, and mature blood cells at the bottom. Mature blood cells are segregated into two cell types lymphoid and myeloid cells. Cells of the adaptive immune system B, T and Natural killer cells are classified as lymphoid cells. Red blood cells, megakaryocytes and cells of the innate immune system; monocytes and granulocytes (neutrophils, basophils and eosinophils) are classified as myeloid cells. In between stem cells and mature cells, are short lived progenitor cells that can give rise to multiple cell types, these cells become more lineage restricted as they progress down the hierarchy. The exact nature and progression of progenitor cells is still not fully understood.

1.1.1 Hematopoietic tree

Many studies have been conducted in mice to unravel the blood system hierarchy. This has led to multiple models of hematopoiesis being proposed. The first model proposed was from Akashi et al, they proposed that the first lineage commitment step for progenitor cells was to become either a lymphoid or myeloid primed [1] (Figure 1.1A i).

Morrison et al then identified a multipotent progenitor cell that fell between the HSC and the lymphoid/myeloid progenitors that could give rise to all cell types, like the HSC but critically had lost the ability to self-renewal, this was termed the multipotent progenitor (MPP) [2]. The classic model was then re-proposed with the addition of the MPP[3] (Figure 1.1A ii).

The classic model however was turned on its head with the discovery of a Lymphoid primed progenitor that retained myeloid but not erythroid or megakaryocyte potential

[4, 5]. Revealing that the lineage commitment process may not be straight forward as originally conceded. A new model incorporating this progenitor was suggested by Adolfsson et al 2005 [6] (Figure 1.1A iii). Much work has also been done to identify these populations in the human system using in vitro and xenograft assays.

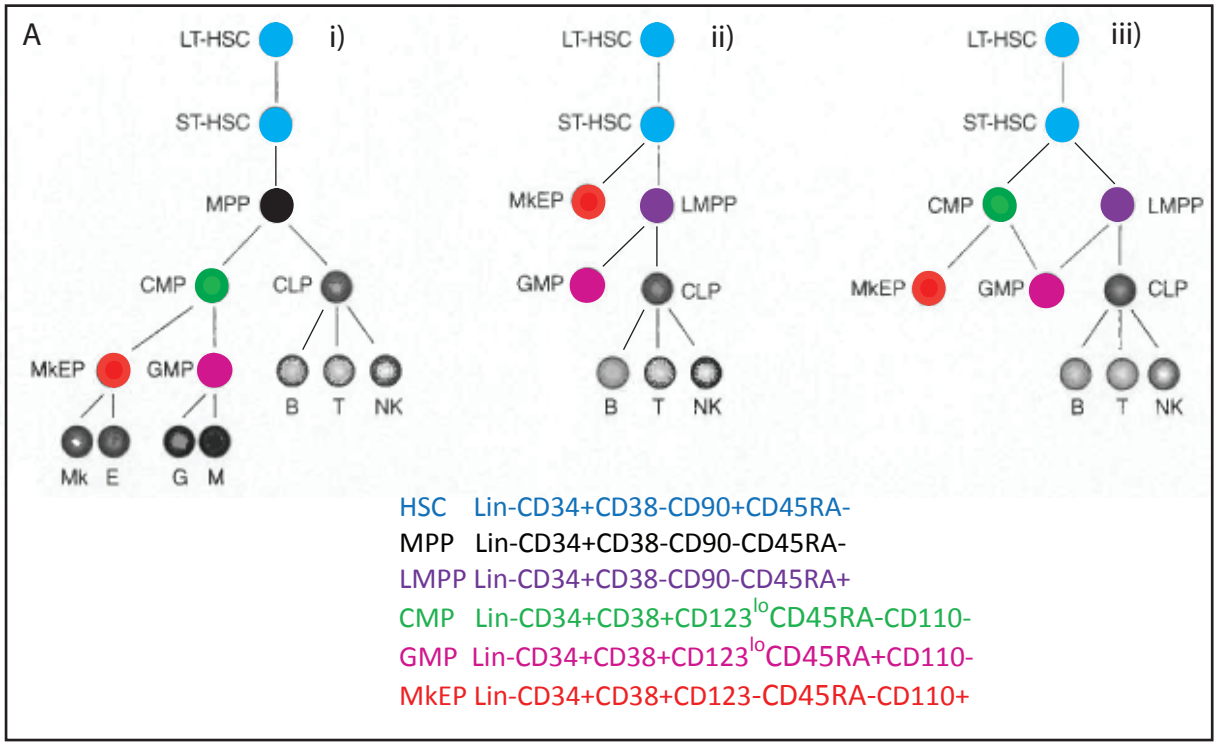
1.1.2 Immunophenotype of human stem and progenitor cells

It was shown by Civin et al that <5% of total human bone marrow was positive for the CD34 antigen and that this population was enriched for HSC and progenitor cells [7]. This was the first human HSC marker to be identified. Various groups then demonstrated that the human HSC could be found in a fraction of cells defined as lineage negative (lin-) CD34+, CD90+ (thy1) [8-10]. Whilst other groups reported that the human HSC could be found within a fraction of cells defined as lin-, CD34+, CD38- cells [11-13].

Majeti et al divided then took the human CD34+, CD38- population and further divided it based on the expression of CD90 and CD45RA into three further subpopulations. (Figure 4.2). They demonstrated that the human HSC fell within in the lin-, CD34+, CD38-, CD90+ and CD45RA- fraction. They also identified a human equivalent of the multipotent progenitor that fell within lin-, CD34+, CD38-, CD90- and CD45- population. However they failed to uncover the role of the final fraction, lin-, CD34+, CD38-, CD90-, CD45RA+ [14].

Goardon et al however demonstrated that this population most closely resembles the murine LMPP population, and was able to give rise to monocytes and granulocytes as well as cells of a lymphoid lineage [15]. Doulatov et al also isolated a human lymphoid progenitor with retained myeloid potential the MLP, which retained the ability to give rise to monocytes but had lost granulocyte potential [16]. On-going work in our laboratory suggests this may be a downstream progenitor of the LMPP in humans.

Figure 1



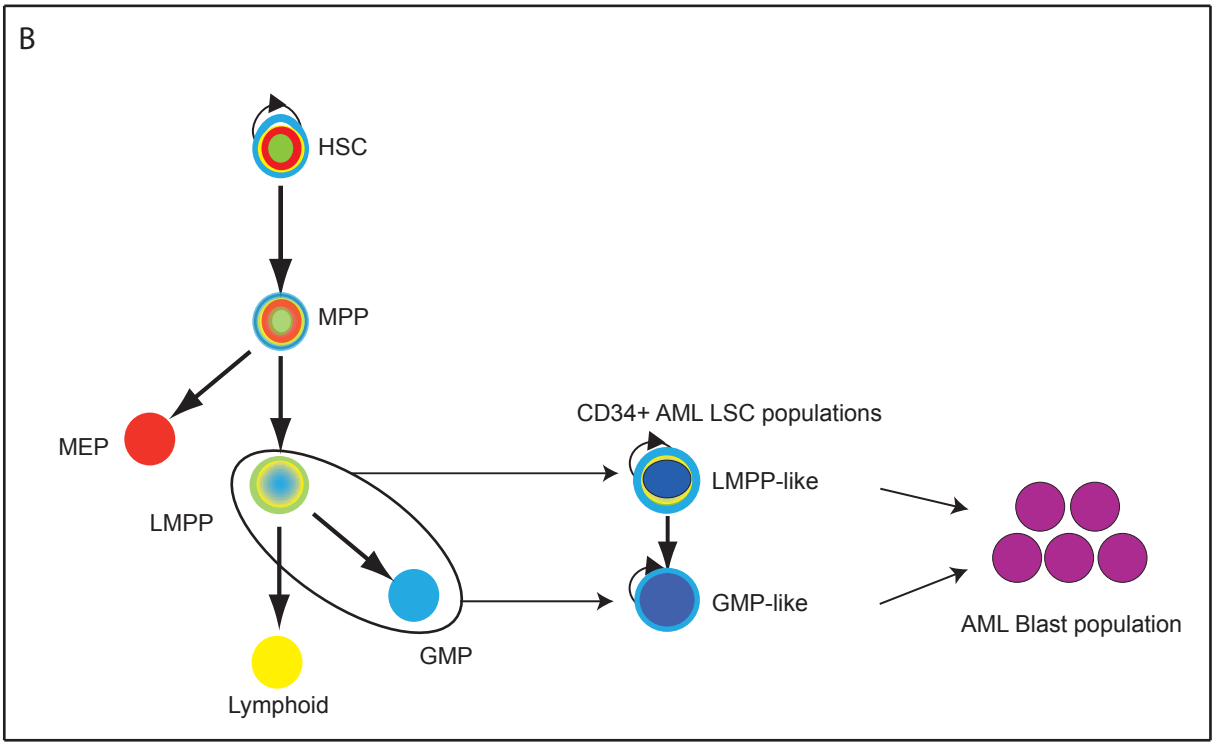


Figure 1: Models of haematopoiesis

A) Different proposed models of normal haematopoiesis with human immunophenotypes listed adapted from color atlas of clinical haematology[53]. i) Model proposed by Akashi et al [54]. ii) model proposed by Adolfsson et al [19] iii) two previous models combined. B) Schematic diagram o normal haematopoiesis combined with hierarchy in acute myeloid leukemia, AML and the cascade of lineage-affiliated gene signatures. HSC- and LMPP-affiliated gene signature (green); myeloid lineage-affiliated gene signature (blue); lymphoid lineage-affiliated gene signature (yellow); megakaryocyte-erythroid lineage-affiliated gene signature (red), taken from Goardon et al [20].

Studies were also carried investigating cells found within the CD38+ fraction of the lin- CD34+ cells, in human marrow. Manz et al showed that the CD38+ population in humans could be separated into three myeloid progenitor populations based on the expression of CD45RA and CD123. The common myeloid progenitor was defined as CD45RA- and CD123+, its downstream progenitors; the granulocyte/monocyte progenitor was defined as CD45RA+ and CD123+, and the megakaryocyte/erythroid progenitor was defined as CD45RA- and CD123- [17]. However the separation of the CMP and the MEP based on CD123 expression was difficult to achieve.

Edvardsson et al went on to show that the inclusion of CD110 led to better separation of the CMP and MEP populations, and allowed a purer MEP population to be isolated [18]. However it was later shown that the antibody believed to bind CD110 in this paper did not actually bind CD110. BD the company supplying this antibody then removed it from the market. It was not known what this antibody was binding, although the molecule it was recognising allowed for a more accurate isolation of the MEP fraction. The immunophenotype panel described above was shown by our lab to have the expected cellular output and gene expression signatures hence this was the panel I took forward to begin my study [15].

1.2 AML

Hematopoiesis is a tightly regulated process in normal individuals. However if this process goes awry it can lead to many different disorders of the marrow from stem cell exhaustion to cancer. The focus of my study was one of the disorders affecting the myeloid branch of haematopoiesis, a terminal cancer defined as acute myeloid leukemia.

1.2.1 Outline of AML

AML is one of the most aggressive haematological malignancies and the most common acute adult leukemia. AML is classified by the accumulation of immature

myeloid cells within the bone marrow. To diagnose AML the blast count must be 20% or higher in the blood or marrow. AML can arise either de novo; from a pre-leukemic conditions known as myelodysplastic syndrome, MDS or myeloproliferative disease (MPD) (this is Secondary AML); and finally as a consequence of chemotherapy treatment for another cancer (therapy-related AML) [19]. The most common cause of death in AML is bone marrow failure due to presence of the leukemic cells in the bone marrow preventing normal haematopoiesis [20]. AML is generally regarded as a disease of the elderly with the average age of presentation being around 70 years. This makes treatment more difficult as cytotoxic chemotherapy is often too toxic for older patients [20] [21]. Relapse in AML is very common and thought to be due to the persistence of chemotherapy resistant leukemic stem cells (LSCs) [22] [23].

1.2.2 AML classification

AML is a highly heterogeneous disease and the different aspects of the disease are often used to classify the subtype. There are two common systems used for a classifying AML. The first system the French American British (FAB) classification system proposed in 1976 defines AML subtype by cell morphology and cytochemistry [24] (Table1.1).

Type	Name	Percentage of adult AML patients
M0	Minimally differentiated acute myeloblastic leukemia	5%
M1	acute myeloblastic leukemia, without maturation	15%
M2	acute myeloblastic leukemia, with granulocytic maturation	25%
M3	promyelocytic or acute promyelocytic leukemia (APL)	10%
M4	acute myelomonocytic leukemia	20%
M4eo	myelomonocytic together with bone marrow eosinophilia	5%
M5	acute monoblastic leukemia (M5a) or acute monocytic leukemia (M5b)	10%
M6	acute erythroid leukemias, including erythroleukemia (M6a) and very rare pure erythroid leukemia (M6b)	5%
M7	acute megakaryoblastic leukemia	5%

Table 1.1 French American British classification system [24, 25]

There are 8 different types identified as M0-M7, they are defined by the extent and nature of the maturation of the leukemic blasts.

The second system is the world health organization (WHO) classification of myeloid neoplasms and acute leukemias, the most recent review of this system was in 2008 [26]. The WHO classification uses cell morphology and cytochemistry like the FAB system but also integrates immunophenotype, genetics and clinical phenotype into defining disease subtype (Table 1.2).

Acute myeloid leukemia with recurrent genetic abnormalities: • AML with t(8;21)(q22;q22); RUNX1-RUNX1T1 • AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11 • APL with t(15;17)(q22;q12); PML-RARA • AML with t(9;11)(p22;q23); MLLT3-MLL • AML with t(6;9)(p23;q34); DEK-NUP214 • AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); RPN1-EVI1 • AML (megakaryoblastic) with t(1;22)(p13;q13); RBM15-MKL1 • Provisional entity: AML with mutated NPM1 • Provisional entity: AML with mutated CEBPA	Acute myeloid leukemia, not otherwise specified: • AML with minimal differentiation • AML without maturation • AML with maturation • Acute myelomonocytic leukemia • Acute monoblastic/monocytic leukemia • Acute erythroid leukemia ○ Pure erythroid leukemia ○ Erythroleukemia, erythroid/myeloid
Acute myeloid leukemia with myelodysplasia-related changes	Myeloid proliferations related to Down syndrome: • Transient abnormal myelopoiesis • Myeloid leukemia associated with Down syndrome
Therapy-related myeloid neoplasms	Myeloid sarcoma
Blastic plasmacytoid dendritic cell neoplasm	

Table 1.2 Acute myeloid leukemia and related neoplasm section of the WHO classification of myeloid neoplasms and acute leukemias [26].

1.2.3 Genetics of AML

With advances in technology such as whole genome sequencing more and more is being discovered about the cytogenetic and the molecular mutational status of AML. This information can be used in classifying the disease as demonstrated by the WHO classification system, as well as being useful in the initial diagnosis of AML [27, 28]. The cytogenetic and molecular mutational status of AML is also predictive of clinical outcome [29-33]. As our knowledge of the mutations in AML increases the more we begin to understand of the molecular biology of the disease and this knowledge can be used to develop of targeted treatments. Gene profiling studies looking at mRNA expression or miRNA expression have shown that AML samples cluster according to their chromosomal or gene abnormalities [34].

In the pathogenesis of AML, the mutations that occur can be classified into two complementation groups; 1. Mutations that impair differentiation and 2. Mutations that infer a survival or proliferative advantage. This is described as the two hit model [35]. However as more mutations have been uncovered, a three hit model has now be proposed. Which proposes to include a third complementation group; of mutations of the epigenome regulators [36].

1.2.3A Cytogenetics

In AML around 55% of patients have chromosomal abnormalities at diagnosis [20]. Many of the chromosomal abnormalities lead to fusion proteins, such as AML-ETO, which is generated by the translocation of chromosome 8 and 21. According to the cytogenetics status patients can be grouped into 3 different risk groups; favourable,

intermediate and adverse. Table 1.3 below displays common chromosomal abnormalities, with the resulting fusion proteins, risk group and the frequency of the abnormality.

A blast count of 20% or more in the blood or bone marrow is normally needed to diagnose AML. However if the blast count is below 20% and following cytogenetics rearrangements are found t(8;21)(q22;q22), inv(16)(p13.1q22), t(16;16)(p13.1;q22), or t(15;17)(q22;q12) WHO guidelines recommend a diagnosis of AML [27, 28] .

chromosomal abnormality	Resulting fusion protein	Risk group	frequency
CN-Normal karyotype	NA	Intermediate	45.1%
t(8;21)	AML-ETO	Favourable	5.5%
t(15;17)	PML-RARA	Favourable	7.6%
inv(16)/(16;16)	CBFβ-MYH11	Favourable	4.7%
inv(3)/t(3:3)	RPN1-EVI1	Adverse	2%
(-)5/5q(-)	NA	Adverse	7.2%
t(6;9)(p23;q34)	DEK/NUP214	Adverse	0.7%
(-)7/7q(-)	NA	Intermediate/Adverse	8.4%
(+)8	NA	Intermediate	9.1%
t(9;11)(p22;q23)	MLLT3-MLL	Intermediate/Adverse	2.1%
t(9;22)(q34;q11)	BCR-ABL1	Adverse	0.8%
del 11q23	MLL-fusion	Intermediate/Adverse	0.9%
del17p/t17p	loss of p53	Adverse	2.2%
(+)21	NA	Intermediate	2.2%
-(Y)	NA	Intermediate	4.1%
complex >3 abnormalities	NA	Adverse	10.7%
complex >5 abnormalities	NA	Adverse	8.8%

Table 1.3 AML cytogenetics [28-32]

1.2.3B Gene mutational status

There is a still variability in clinical outcome of patients assessed as intermediate and favourable risk categorizes by cytogenetics. However the identification of gene mutations has further enhanced prognosis prediction. The outcome of CN-AML was particularly variable, the advances in identifying gene mutations has “revolutionized the

prognosis stratification" of this subtype [27] (Table 1.4). The most commonly assessed genes are NPM1, CEBPα and FLT3, however more and more genes are being identified as mutated (such as IDH1,2, kit, KRAS ect) or deregulated (such as BAALC, ERG, EVI1 ect) and shown to be indicative of prognosis [27, 33, 37]. The discovery of mutated genes not only enhances the prediction of outcome but also enhances our understanding of the molecular basis of the disease. Certain gene mutations are also related to other aspects of disease, such as the NPM1 mutation being associated with the CD34- form of the disease [38].

Favourable	**Unfavourable**	**No contribution to risk/still to be determined**
CEBPA mut	FLT3-ITD mut	NRAS
NPM1 mut	KIT mut	TP53
	MLL tandem duplication	TET2
	BAALC overexpression	ASLX1
	IDH1 or 2 mut	
	WT1 mut	
	Runx1 mut	

Table 1.4 Risk significance of gene mutations in AML [27, 33]

1.3 Cancer stem cells (CSC)

Cancer stem cells are defined as "a cell within a tumour that possesses the capacity to self-renew and to cause the heterogeneous lineages of cancer cells that comprise the tumour" [39]. Although cancer stems cells were originally identified in leukemia, many examples have also now been identified in solid tumours, such as breast, colon and brain tumours [40-43]. However there remains a degree of controversy over whether there are CSCs present in all tumours types [44, 45] and it may be the case that cancer stem cells do not exist for every type of cancer.

1.3.1 Leukemic stem cell (LSC)

The study of the LSC in AML has led the way for research in the field of cancer stem cells [46, 47]. AML is thought to resemble normal haematopoiesis, with a leukemic stem cell at the apex and more differentiated blasts at the bottom (Figure 1.1B) [47] [48]. The LSC has also been shown to share several properties with normal HSC such as the ability to self-renew, apoptosis resistance and a more quiescent cell state [49-52]. The LSC is able to reproduce the original leukemia when transplanted into immunodeficient mice. The LSC is also thought to be more resistant to chemotherapy than the rapidly proliferating blast cells[23].

1.3.2 History of the LSC in AML

The first important paper on the road to discovering the LSC in AML was Lapidot's paper in 1994 [53]. They were able to show that only certain cells within a leukemic population could initiate AML when transplanted into SCID mice, the leukemia initiating cells L-ICs were CD34+CD38- cells. Bonnet and Dick were then able to demonstrate that these L-ICs were actually leukemic stem cells. They were able to do this by proving that the L-ICs have self-renewal potential, by carrying out secondary transplantation assays in NOD/SCID mice [54] . Both these papers suggested that the LSC was only found within the CD34+CD38- compartment of AML.

As LSCs were initially CD34+CD38- and were able to self-renew the Dick laboratory suggested the LSC may have originated from HSCs. This is now thought to be incorrect. Also many groups working with mouse models have been able transformed GMPs into LSC, by transfecting GMPs with common AML oncogenes [55-57]. This

brought into question the origin cell of the LSC and where the LSC lies within the human disease.

Taussig et al highlighted an important problem with our models systems for reading out human leukaemia. In the models previously used there is CD38 antibody mediated cell clearance. When this effect is eliminated CD34+CD38+ leukemic cells can engraft primary and secondary recipients and generate leukaemia [58]. Moreover, in ~5% of primary human AML samples where CD34+ account for < 0.05% of cells, the Bonnet lab has shown that LSC activity can also reside in the CD34- compartment [38]. Studies have now shown that LSCs are found within both the CD34+CD38- fraction and CD34+CD38+ fraction. These studies have also shown through limit dilution assays that in these fractions that LSC is rare and the frequency is varied, with more LSCs found within the CD34+CD38- fraction [15, 59].

Previous studies had not clarified the relationship between LSCs and normal stem/progenitor populations in human AML, until the paper from our group by Goardon. For example, not only does the normal CD34+CD38- compartment contain HSCs, but it also contains the MPP and the LMPP [14]. Clarifying the cellular and molecular relationships between different AML LSC populations within a patient, between different patients, and finally, between AML LSC and normal HSC/progenitor populations was required to help define the molecular pathways important in the stepwise transformation of normal HSCs/progenitors [15].

1.3.3 Previous work on AML LSCs in our laboratory

Our group has studied CD34+ AML to more precisely map LSC populations in an individual patient and between patients to better understand the cellular hierarchy in

AML. In addition, our laboratory has studied the relationship between LSCs and normal hematopoietic stem/progenitor cells.

Immunophenotyping of 82 primary human CD34+ AMLs was carried out. In the AML CD34+ populations there were two patterns of immunophenotypes. In the first, accounting for ~80% of AML samples, within the CD34+CD38- compartment there was a dominant expanded CD34+CD38-CD90-CD45RA+ population. This leuekmic population had the same immunophenotype as the normal LMPP population. Within this AML group there was also a co-existent expansion of the CD34+CD38+Cd123+/loCD110-CD45RA+ population within the CD34+CD38+ fraction. This leukemic population had the same immunophenotype as normal GMPs.

In the second immunophenotypic group of AML (accounting for the rest of the 20% of samples) there was an expansion of a CD34+CD38-CD90-CD45RA- population (i.e. MPP-like as the normal CD34+CD38-CD90-CD45RA- population has been defined as an MPP) within the CD34+ CD38- compartment. In these samples, there was a coexistent expanded CD34+CD38+CD123+/loCD110-CD45RA- population (i.e. CMP-like as the normal CD34+CD38+CD123+/loCD110-CD45RA-population has been defined as an CMP) in the CD34+CD38+ fraction.

As the LMPP-like/GMP-like expanded AML group is the more prevalent group (~80%) of CD34+ AMLs, further work was focused on this group of AML. Our lab were able to show that both the LMPP-like and GMP-like populations within a patient had LSC activity in primary and secondary NOD/SCID recipient mice treated with anti-CD122 antibody (to eliminate NK cells and CD38 antibody mediated clearance). Both in vitro and in vivo there was a hierarchical relationship between these populations. The LMPP-like LSCs were able to self-renew, give rise to GMP-like LSCs and bulk blast

populations. In contrast, the GMP-like LSC population could self-renew and generate bulk blast populations but not give rise to the LMPP like LSCs.

Global mRNA expression profiles were analysed for LMPP-like and GMP-like populations in AML samples and for HSCs, MPPs, LMPs, GMPs and CMPs in normal bone marrow samples. This analysis revealed that although the GMP-like and LMPP-like populations were molecularly distinct, they most closely resembled the normal population with which they shared an immunophenotype with [15].

1.4 MicroRNAs

MicroRNAs (miRNAs) are small non-coding RNAs that are involved in the regulation of mRNA [60]. They were first described in *C. elegans* in 1993 (lin-4) [61]. It took nearly another 10 years until the second miRNA, let-7, was discovered [62]. But after an initially slow start there has been an explosion of papers relating to miRNAs and they have now become a field in their own right. miRNAs have been shown to be implicated in many diseases and have been shown to be aberrantly expressed in AML.

1.4.1 MicroRNA biogenesis and Function

The majority of miRNAs like mRNAs are transcribed from DNA by RNA polymerase II and have poly A tails and 5'caps [63, 64]. The transcript generated is known as primary miRNAs or pri-miRNAs that form hairpin-like structures [63]. Pri-miRNAs are then processed in the nucleus by a complex containing Drosha (RNase III enzyme) and DGCR8 that cleaves the pri-miRNA at the base of the hairpin [65-67]. This generates what is known as the precursory miRNA, pre-miRNA. The pre-miRNAs are then exported from the nucleus by exportin-5 into the cytoplasm [68]. Where the RNA is

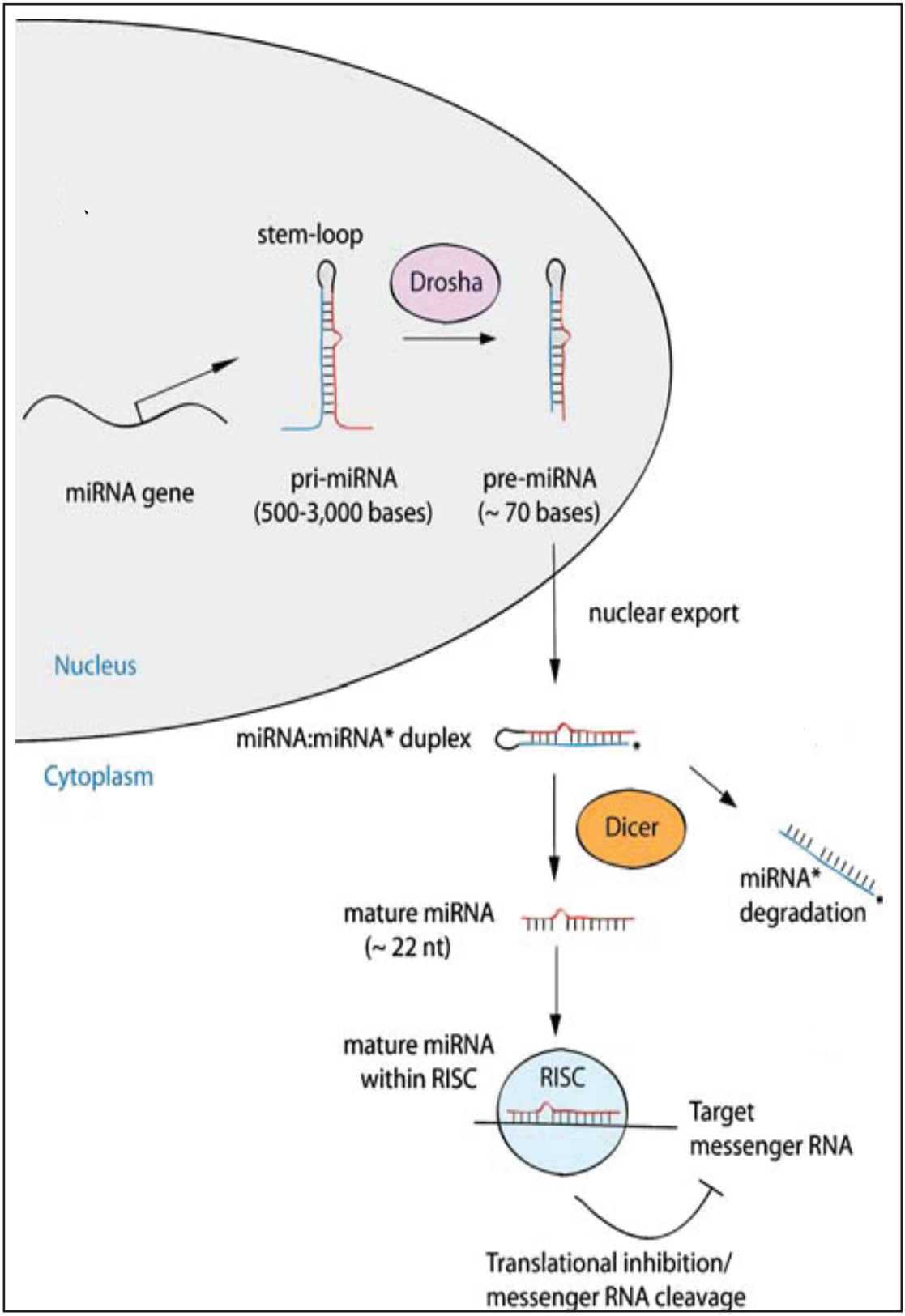


Figure 1.2: schematic representation of miRNA biogenesis and function.
Taken from Sassen et al 08.[99] (permission to reproduce this figure was granted by Springer via rightslink)

processed into 22nt nucleotide double stranded RNA containing the mature miRNA guide and the passenger strand by Dicer (RNase III enzyme) [69, 70]. The guide strand is loaded into the RISC complex, where it binds to its mRNA target and exerts its function and the passenger strand is degraded (Figure 1.2) [71, 72]. miRNAs either cause cleavage of their target mRNA, destabilization of their target mRNA leading to enhanced degradation or a block translation of their target mRNA. The latter methods are thought to occur in mammals [60, 73]. miRNAs do not have to base pair perfectly to target mRNAs to stop protein expression; this means a single miRNA can have multiple targets. Currently, there are 2042 humans miRNAs listed in miRBase [74, 75], and are thought to regulate expression of over 30% of protein coding genes [76, 77].

1.4.2 MicroRNA nomenclature

The first part of the miRNA I.D. is a three to four letter code that indicates which species the miRNA is from, i.e. if the mRNA is *Homo sapiens* it will be HSA. The middle part of the I.D. specifies the maturity of the miRNA; mir represents the precursor hairpin miRNA and miR represents the mature miRNA. The final part of the I.D. is the miRNA unique annotation number. (e.g. HSA-miR-282= Human mature miRNA number 282)

The mature miRNA has two strands, the functional guide strand and the passenger strand. A star at the end of a miRNA I.D. represents the passenger strand (e.g. miR-25*). However an I.D. should not infer relationship as this may vary in context, as in an alternate cell type the passenger strand may become the guide strand. So a newer nomenclature is being phased in, that denotes the different strands as either 3p or 5p (e.g. miR25-3p).

Identical mature miRNAs that originate from different hairpin precursors are given number suffixes (e.g. miR25-1 and miR-25-2). Paralogous sequences where the mature miRNA differs at only one or two positions are given lettered suffixes (e.g. miR-25a and miR-25b) [78].

1.4.3 MicroRNA in hematopoiesis

miRNAs have been shown to play an important role in hematopoiesis. This was firstly demonstrated by deletion or inactivation of important enzymes involved in the biogenesis of miRNAs, which led to severe hematopoietic defects [79-83]. Work targeting specific miRNAs has also shown them to be important in hematopoiesis and begun to unravel the role they play within hematopoiesis [83].

The miR-125 family have been shown to play a role in normal HSCs. miR-125b is highly expressed in HCSs and its expression decreases in more mature progenitors. Overexpression of this miRNA in HSC enhances their function and increases their ability self-renew [84, 85]. miR-125a has been shown to be play a role in controlling HSC number [82]. The overexpression of either miR-125a or miR-125b has been shown to increase the number of HSC in vivo [82, 84, 85].

miRNAs are also involved in myelopoiesis. miR-223 is involved in myeloid differentiation and is controlled by myeloid transcription factors [86]. It is also known to regulate GMP proliferation with its knockdown leading to an increase in GMP numbers [87]. miR-146 is involved in immune responses and is rapidly induced in myeloid cells upon exposure to TRL ligands [88-91]. miR-221/222 are found to inhibit erythropoiesis through binding to kit [92] and mir-155 drives MEPs toward megakaryopoiesis [93, 94].

miRNAs have all been shown to be important in lymphoid development. With the miR-17-92 polycistronic cluster being essential for B cell development [95] and miR-150 controlling B cell development by regulating the expression c-Myb [96, 97].

1.4.4 MicroRNAs in cancer

miRNAs have been implicated in a wide range of disease including cancer [98]. miRNAs are often found at fragile sites in the genome or regions commonly amplified or deleted in human cancers [99]. The first miRNAs shown to be involved in cancer were miR-15/16 in CLL. A well-known 30 kb deletion occurs on chromosome 13 in CLL and is the most frequent abnormality in CLL. Studies previously failed to uncover a causative gene in this region. Calin et al identified two miRNAs within this region miR-15/16. They subsequently showed that miR-15/16 were absent or downregulated compared to normal tissue in 68% of the patients analysed [100].

miRNAs are known to function as tumour suppressors and oncogenes [98]. Let-7 is a miRNA known to act as tumour suppressor and is down regulated in many cancers [101, 102]. The mir-17–92 polycistronic cluster was one of the first groups of miRNAs shown to have oncogenic potential. It is overexpressed in human B-cell lymphomas and accelerates tumorigenesis in a *c-myc* mouse model of lymphoma [99, 103].

1.4.5 MicroRNAs in AML

Many profiling studies have been conducted examining miRNA expression in AML. Mi et al performed miRNA expression profiling on AML and Acute lymphoid leukemia (ALL) samples. They showed that when using hierarchal clustering miRNA expression profiles cluster by leukemia type, with 27 miRNAs being differentially expressed

between the two leukemias. They also demonstrated that using the expression of the four most discriminatory miRNAs (miR-128a, miR-128b, let-7b and miR-223), they could accurately predict whether samples were AML or ALL [104].

miRNA expression studies have also been done looking at AMLs with specific gene mutations or gene deregulations. These studies showed that distinctive miRNA signatures could be correlated with gene mutations such as NPM1 [105], IDH1/2 [106], FLT3-ITD [107] and TET2 [108] and deregulated genes such as BAALC and ERG [109].

Multiple papers have also shown that when using hierarchal clustering, miRNA profiles of AML samples cluster by cytogenetic subtype and molecular mutations. miRNA signatures could also be derived that could be used to predict subtype [105, 110-112].

As well as looking at miRNAs in relation to cytogenetics and gene mutations, miRNA expression has also been found to be correlated with clinical outcome [113]. Dixon-Mciver found that certain miRNAs were associated with different cytogenetic risks groups. Patients in the favourable cytogenetic risk group had low expression of let-7b and miR-9, whereas patients in the adverse or intermediate group had higher expression of these miRNAs [112] . Higher expression of miR-191 and miR-199a has also been linked to lower overall survival in AML [114].

Although to date few functional studies have been conducted on miRNAs in AML, some examples do exist that show that miRNAs are playing a functional role in AML. Han et al found that in mice miR-29a was upregulated in HSCs and MPPs compared to more mature progenitor cells. When miR-29a was overexpressed in mice HSC/progenitor cells they developed a Myeloproliferative disorder (MPD). When the BM of these mice was transplanted into secondary recipients, these mice went on to

develop AML. They also looked at miR-29a expression in primary human samples and found it was overexpressed in AML CD34+CD38- cells as compared to normal CD34+CD38- [115].

miR-196b is a transcriptional target of MLL (mixed lineage leukemia gene) a protein commonly associated with fusion proteins in AML. When Popovic et al transformed mouse BM cells with the MLL-AF9 fusion protein; they saw an increase in the proliferative capacity of the cells, as well as an increase in miR-196b. Treatment of these cells with a miR-196b antagomir abrogated their re-plating potential, demonstrating a central role for miR-196b in this phenotype. They also profiled primary leuekmic samples and found that miR-196b was overexpressed in MLL-associated leukemias compared to non MLL-associated leukemias [116].

miR-155 expression has been reported to be overexpressed in certain subtypes of AML. O'Connell et al found when they overexpressed miR-155 in mouse BM cells this lead to an MPD, but did not progress to acute leukemia. Suggesting that overexpression of this miRNA can lead to a pre-leukemic condition but further mutations are required for its progression to leukemia [117].

However studies comparing AML miRNAs profiles to normal miRNA profiles have shared very few similarities. This is likely because studies comparing AML samples to normal samples have used a wide variety of different cells types as control/reference populations; such as CD34+ BM, whole BM, MNCs or mobilized peripheral blood. This highlights the importance of selecting the most appropriate control population [33, 104, 111, 114, 118-120]. Most of these studies also analysed bulk populations, such as CD34+ cells. The normal CD34+ population contains HSCs and many types of different progenitor cells. With the example of LMPP-like/GMP-like expanded AML the CD34+ population in this subtype of AML is largely made up of LMPP-like and GMP-

like cells [15]. Indicating comparing normal bulk CD34+ cells to bulk leukemic CD34+ cells may not be a representative comparison. Also analysing bulk populations may mask leuekmic alterations. To address this we must more careful consider what cells we use as control/reference populations. Using a variety of normal purified normal populations as references compared to purified leuekmic populations, would allow us to more clearly understand what is happening at a molecular level.

1.5 Project aims

- To better purify the GMP-like and LMPP-like populations in AML, by identifying antigens that would better discriminate these populations from each other, when using flow cytometry.
- To generate miRNA expression profiles of highly purified normal and leukemic populations. (GMP-like and LMPP-like AML populations and normal HSC, MPP, LMPP, CMP and GMP populations.)
- To use bioinformatic analysis to understand how the different normal and leukemic populations related to each other on a miRNA level.
- To investigate which normal population, the leukemic populations most resembled on a miRNA level.
- Finally to identifying miRNAs that are aberrantly expressed in AML.

Chapter 2: MATERIALS AND METHODS

2.1 Patient Samples and cell lines

After washing samples were centrifuged at 1200rpm in a bench top centrifuge unless otherwise stated.

2.1.1 Sample source

All AML bone marrow and peripheral blood samples were obtained from AML patients with informed consent. All Normal Marrow samples were obtained from patients undergoing orthopaedic surgery with normal blood counts/films with informed consent. (Protocols 06/Q1606/110 and 05/MRE07/74 approved by Oxford Ethics Committee B).

2.1.2 Ficoll density gradient separation

Mononuclear cells are obtained from bone marrow and peripheral blood samples by Ficoll density gradient separation. All samples were diluted at least 2x in RPMI medium (Gibco) and then layered on to ficoll (Lymphoprep^{TM}-Axis-shield). Samples were then centrifuged at 800g for 25 minutes. After centrifugation three layers are formed with the mononuclear cells being found in the middle phase. The layer containing mononuclear cells was removed and washed twice with PBE (PBS (Gibco)/0.5% BSA (sigma Aldrich) / 2mM EDTA (Sigma Aldrich)).

2.1.3 CD34 enrichment

The mononuclear cells were enriched for CD34+ cells using MACS magnetic beads and columns (Miltenyi Biotec). Mononuclear cells were resuspended at $1x10^8$ cells/500µl PBE or 500µl if less than 10^8 cells. 100µl of FcR block per 500µl of cells was added to the cells. Cells were then incubated with 100µl of MACs microbeads conjugated to anti-human CD34 antibody (clone: QBEND10) per 500µl of cells for 30mins at 4°C. Cells were then washed with PBE and resuspended at $1x10^8$ cells/500µl PBE or 500µl if less than 10^8 cells. Cells were filtered through a 30µm filter (CellTrics) onto a LS selection column. LS columns were used according to manufacturer's instructions.

CD34 kit comparison: Mononuclear cells were enriched for CD34+ using either the Miltenyi kit as above or using the Stem Cell Technologies' EasySep human CD34 positive selection kit according to manufacturers' instructions or using the Life Technologies' Dynabeads CD34 positive isolation kit according to manufacturer's instructions.

2.1.4 Freezing samples

After CD34 enrichment, enriched and depleted fractions were either used fresh or cryopreserved by freezing and storing in vapour phase liquid nitrogen. Samples that were frozen were resuspended in 1ml of 90% Foetal calf serum (FCS) (PAA)/10% DMSO (VWR) in cryo-vials. Samples were then subject to control rate freezing by placing vials into a Mr.Frosty (Nalgen) then storing this at -80°C overnight before transferring vials to vapour phase liquid nitrogen for long term storage.

2.1.5 Thawing and overnight culture

Before use frozen samples were thawed and then left in culture overnight to allow recovery from the thawing process.

Thawing: vials were flash thawed by swirling in a 37°C water bath for approximately 30 seconds. Cells were then washed with 15ml of FACS buffer (90% IMDM no phenol red (Gilbco)/10%FCS (PAA)/DNase (Roche)). To wash cells after thawing FACS buffer was added slowly drop by drop to cells whilst continuously swirling.

Overnight culture: cells were cultured in stemspan (Stem Cell Technologies) supplemented with 100ng/ml TPO, SCF, and FLT3-L (PeproTech). In 6 well plates (BD Falcon) for 5-10 $x10^6$ cells and 12 well plates (BD Falcon) for <5 $x10^6$ cells.

2.1.6 K562 cells

K562 cells were maintained in RPMI medium (Gibco) supplemented with 10% FCS (PAA) and L-Glutamine (Gibco).

2.1.7 Cell counting

Cell counts were done manually using a haemocytometer and trypan blue (life technologies) to exclude for dead cells.

2.2 Flow cytometry

2.2.1 Cytometers

All analysis was carried out on a Cyan ADP (Dakocytomation). Cell sorting was performed on a FACS Aria II (Becton Dickinson). FACS data was analysed on either summit software (Dakocytomation) or FACS Diva software (BD biosciences).

2.2.2 Cell staining

(After washing cells were centrifuged at 1200rpm in a bench top centrifuge.)

Standard staining: Cells were incubated with antibodies for 20 mins on ice in the dark before being washed with FACS buffer. Hoechst (life technologies) was added to cells just before analysis.

Multistage staining: Cells were first incubated with purified anti-lineage antibodies for 20 mins on ice in the dark, before being washed with FACS buffer. Cells were then incubated with either QDot 605 or QDot 655 goat anti-mouse for 20 mins on ice in the dark, before being washed twice with FACS buffer. Cells were then incubated with antibody for 20 mins on ice in the dark, before being washed with FACS buffer. Finally cells were incubated with streptavidin for 10mins on ice in the dark before being washed with FACS buffer. Hoechst was added to cells just before analysis. (All antibodies were titrated before use).

2.2.3 Analysis

Cyan: Data was analysed with summit software (Dakocytomation). Voltages were set using an unstained control and compensation was determined manually using single

colour compensation controls. Gates were set post compensation; using unstained controls and single colour compensation controls, as well as visually inspecting population by eye.

Aria: Data was analysed with Diva software (BD biosciences). Voltages were set using an unstained control; each voltage was also checked using single colour compensation controls. Compensation was determined automatically by software using single colour compensation controls. Compensation was then manually checked/corrected using FMO controls. Gates were set post compensation; using unstained controls and FMO controls, as well as visually inspecting population by eye. Example of how gates are set (Figure 2.1).

2.2.4 Cell sorting

Before sorting cells were first compensated and analysed, gates were then tightened for sorting (Figure 2.1). Sorting on the Aria was done using a 70µm nozzle and sorting on the Moflow was done using a 100µm nozzle. The flow rate for analysis and sorting on the Aria was set between 1 and 5. A purity mask of 32-32-0 was used for sorting on the Aria Post sort cells were collected in FACS tubes containing ~1-2ml of FACS buffer. Collection tubes were reanalysed through sorter, after the sort had been completed to check final purity of sample.

The Aria and moflow only allow four way sorting, so for the stem and progenitor sort the GMP, CMP and CD38- fractions were sorted first. The CD38- fraction was then re-sorted into the HSC, MPP and LMPP fractions.

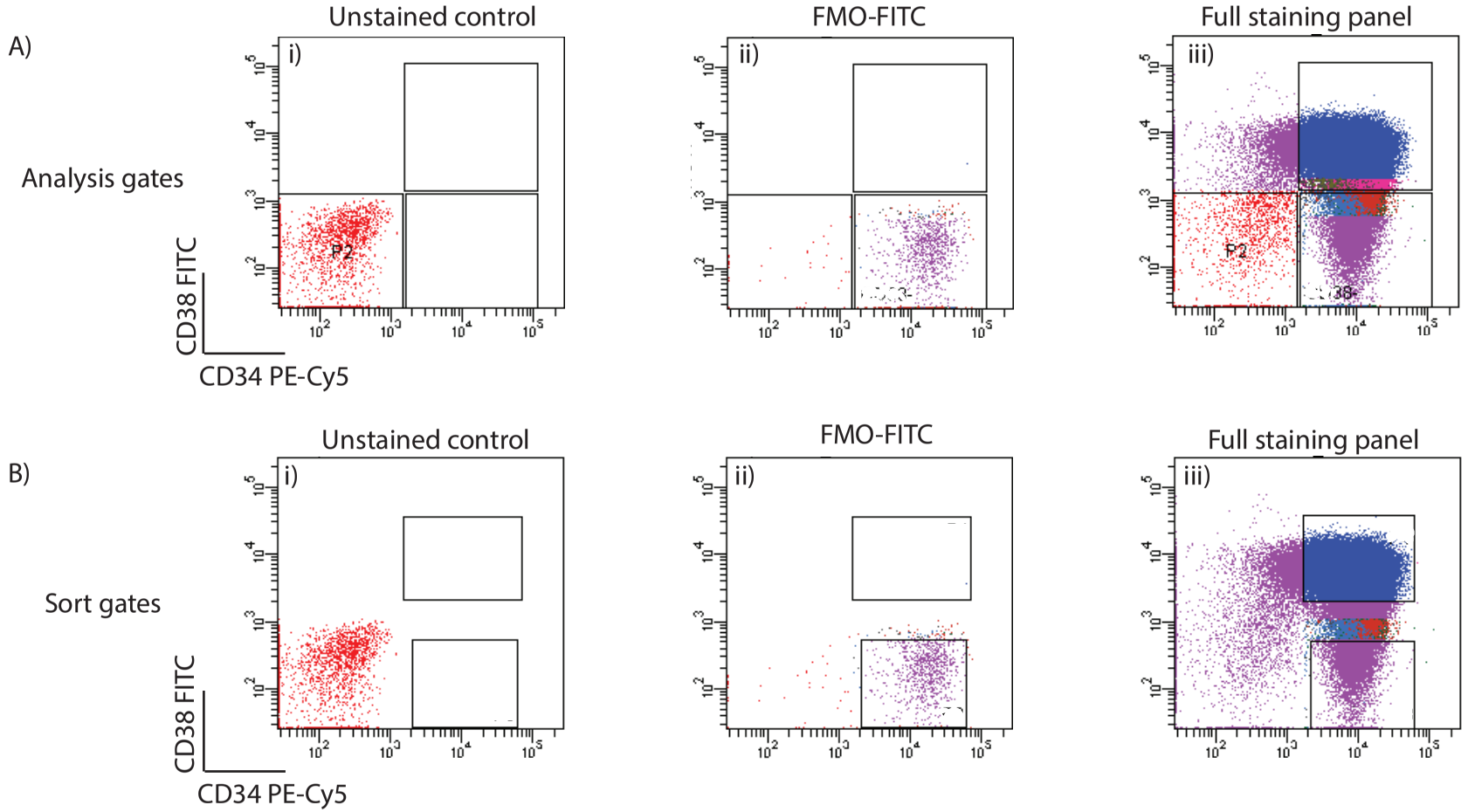


Figure 2.1 Setting FACS gates.
An example of how FACS analysis and sort gates were set for analysis and sorting. Example shown is how CD38 FITC gates were set. To set gates the unstained control and the flurochrome FMO were used, as well as visually inspecting the populations, however this is more difficult with a continuum such as the CD38 populations. A) Analysis gates. B) Sort gates; analysis gates were set first then tightened for sorting.

2.2.5 Optimization of cell sorting procedure

Due to there being problems obtaining a high enough yield of the LMPP cells required for the microarray, a programme of optimization was set about to see if the any of the cell sorting steps could be improved to increase yield.

2.2.5.1 Optimizing cell staining

The first step was to try and reduce the number of cells lost during staining, via altering tubes the cells were stained in and the centrifugation speed used. My standard method was to use FACS tubes and a centrifugation speed of 200g. An alternate method via other groups in the department was to use Eppendorf tubes. A comparison of different tube types and centrifugation speeds was compared on cell loss (Table 2.1).

	Starting cell count	Post staining with FACS tube		Post staining with Eppendorf tubes	
centrifugation speed		400g	200g	400g	200g
OX541					
Cell numbers	2.6	NA	2.03	NA	1.96
% retained		NA	78%	NA	75%
OX564					
Cell numbers	0.6	NA	0.525	NA	0.378
% retained		NA	88%	NA	63%
OX484					
Cell numbers	4	2.72	3.08	2.24	NA
% retained		68%	77%	56%	NA
OX454					
Cell numbers	2.97	1.31	1.27	0.954	0.93
% retained		44%	43%	32%	31%

Table 2.1 Comparing different tubes for FACS staining.

Tests were done with several different samples to test if one type of tube retained cells better than the other, during the staining procedure. Cells were split equally between FACS tubes and Eppendorfs, then a dummy staining experiment that had multiple centrifugation steps was carried out. Cells were counted with trypan blue so cell counts only reflect viable cells. Different centrifugation speeds were also tested. Cells used were MNCs that had previously been depleted for CD34.

Increasing centrifugation speed and using Eppendorfs tubes actually increased cell loses, possibly due to harsher centrifugation forces felt by the cells. Therefore the standard method was maintained.

2.2.5.2 Analysing sorter efficiency

To ensure that the FACS sorter I was using was the most efficient available to me, I tested the efficiency of the different FACS sorters in the institute. The sorter I was using was a BD FACS Aria II Sorter, also available to me was a BD Aria special order research product (SORP) and a Beckman Coulter Moflo Cell Sorter. To investigate the efficiencies of the different sorters I did head to head comparisons of the different sorters to the Aria II, using the same sample on the same day.

A normal CD34+ enriched bone marrow sample was analysed and sorted, on the Aria II and Aria SORP. The FACS plots from the different Arias are very similar (Figure 2.2 and 2.3). Table 2.2A shows the percentage each population made up of the CD34+ fraction, as determined by analysis on the different Arias, the values were comparable between the two Arias. The efficiency of each Aria was calculated based on yields obtained compared to the theoretical yield (Table 2.2). The result of this was that the Aria II was more efficient.

A second normal CD34+ enriched bone marrow sample was analysed and sorted, on the Aria II and Moflo. The FACS plots revealed that signal was lower and population definition poorer on the moflow (Figure 2.4 and 2.5). Table 2.3A shows the percentage each population made up of the CD34+ fraction, as determined by analysis on the different sorters. The population percentages differed a lot between the Aria and the Moflo; this is most likely to reduced resolution on the Moflo. The efficiency of each

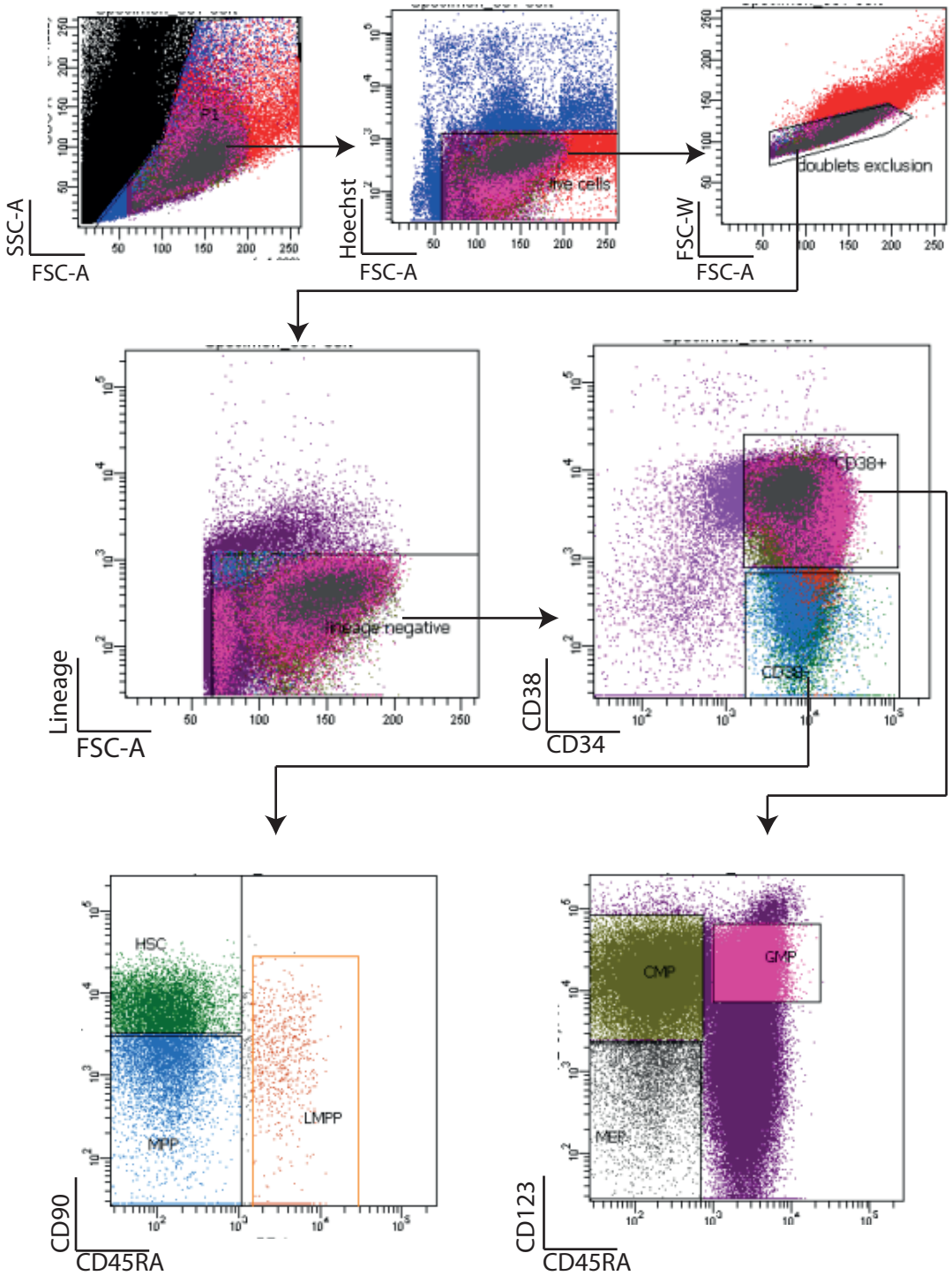


Fig 2.2 Comparing FACS sorters, ARIA II sortert versus Aria SORP sorter. (Aria II).
The same sample, on the same day was sorted on two different FACS sorters to test sorter efficiency, Normal bone marrow sample: OX555 was used. FACS plots from sort and analysis on Aria II.

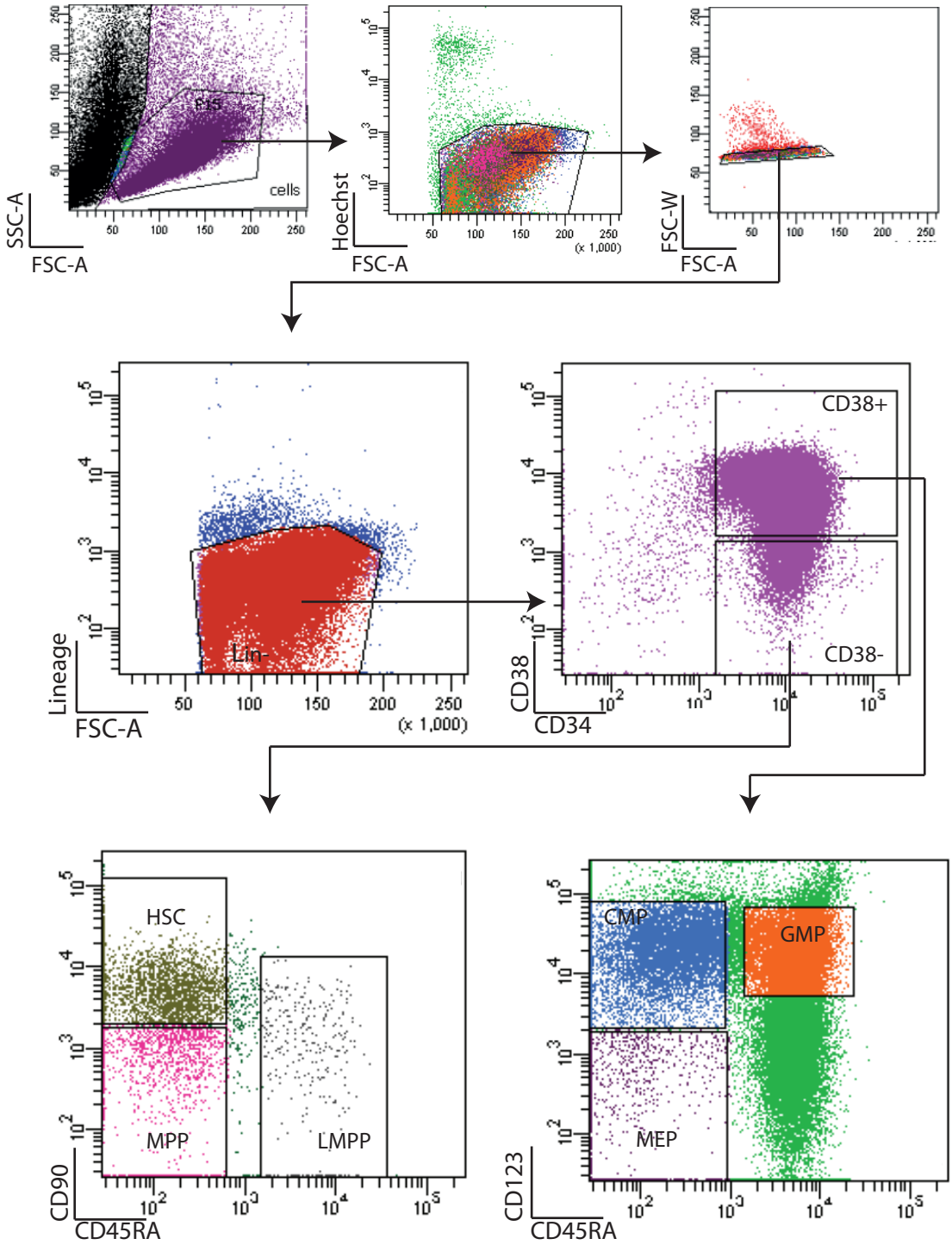


Fig 2.3 Comparing FACS sorters, Aria II sorter versus Aria SORP sorter. (Aria SORP).
The same sample, on the same day was sorted on two different FACS sorters to test sorter efficiency, Normal bone marrow sample: OX555 was used. FACS plots from sort and analysis on Aria SORP.

A)

Sorter	ARIA SORP	ARIA II
	population % of CD34+ cells	
CD38-	9.9	9
CD38+	90.1	91
HSC	6.2	5.9
MPP	2.6	2.1
LMPP	0.6	0.5
CMP	18.2	15.4

B)

		Sort Sample	Sample population		Yield			
Population	**Sorter**	**Starting cell count x10^6**	**Frequency %**	**Post sort purity %**	**Theoretical**	**Actual**	**Adjusted**	**% of theoretical yield sorted**
LMPP	ARIA SORP	0.36	0.277	94	997	550	517	**52**
CMP	ARIA SORP	0.36	6.2	96	22,320	12,037	11,556	**52**
MPP	ARIA SORP	0.36	1.1	90	3,960	2,579	2,321	**59**
HSC	ARIA SORP	0.36	3.3	97	11,880	6,419	6,226	**52**
LMPP	ARIA II	0.36	0.132	100	475	367	367	**77**
CMP	ARIA II	0.36	5.7	99	20,520	15,987	15,827	**77**
MPP	ARIA II	0.36	1.56	96	5,616	4,448	4,270	**76**
HSC	ARIA II	0.36	3.1	100	11,160	8,552	8,552	**77**

Table 2.2 Comparing FACS sorters, details of Aria II sort versus Aria SORP sort.A) Percentage each population is of CD34+ cells; after analysis on either Aria II or Aria SORP. Percentages were calculated using analysis gates. B) Details of sort yields, for each population when sorted on either Aria II or Aria SORP. (Frequency percentages were calculated using sort gates and purity percentages were calculated using analysis gates.)

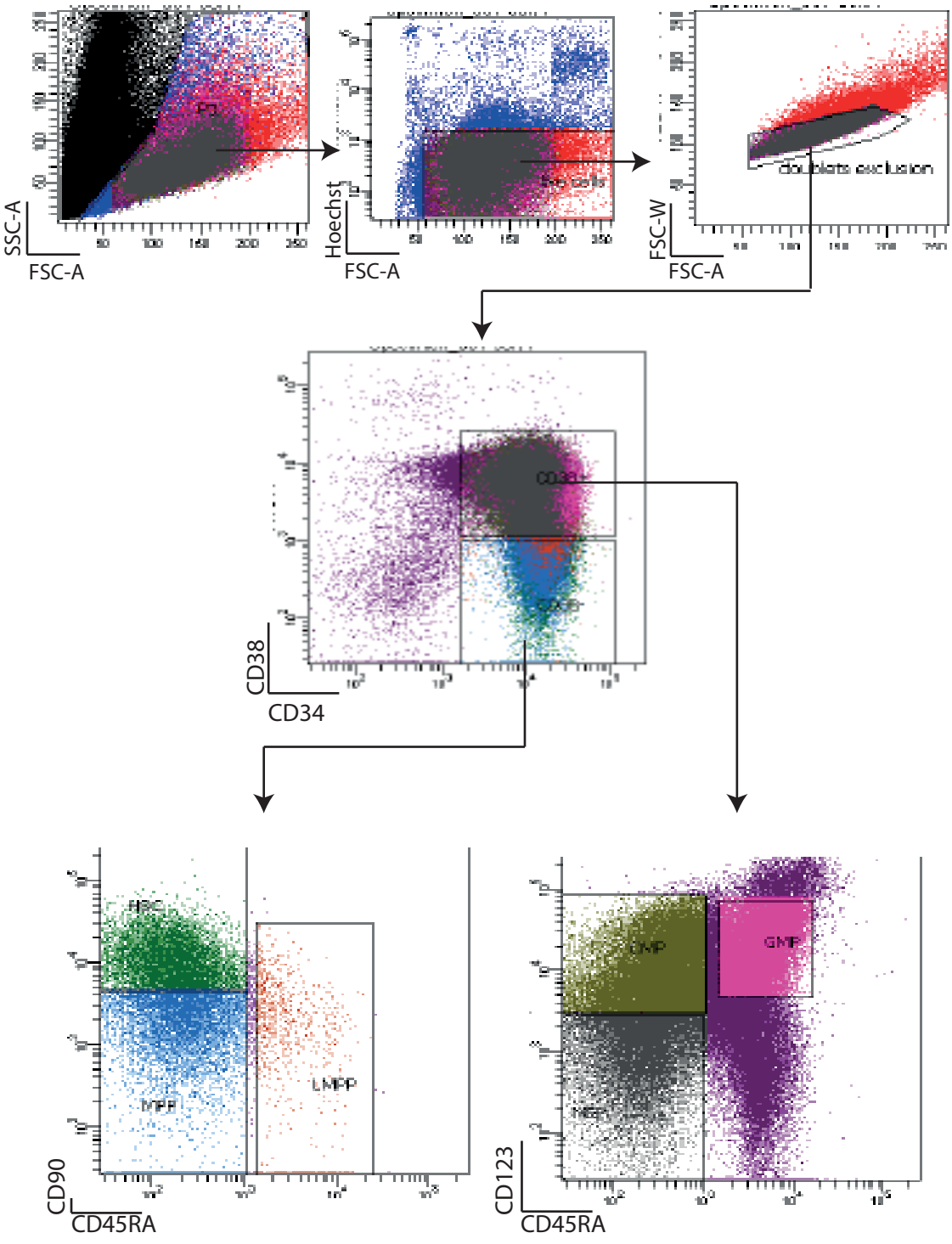


Fig 2.4 Comparing FACS sorters, Aria II sorter versus Moflo sorter. (Aria II).
The same sample, on the same day was sorted on two different FACS sorters to test sorter efficiency, Normal bone marrow sample: OX529 was used. FACS plots from sort and analysis on Aria II.

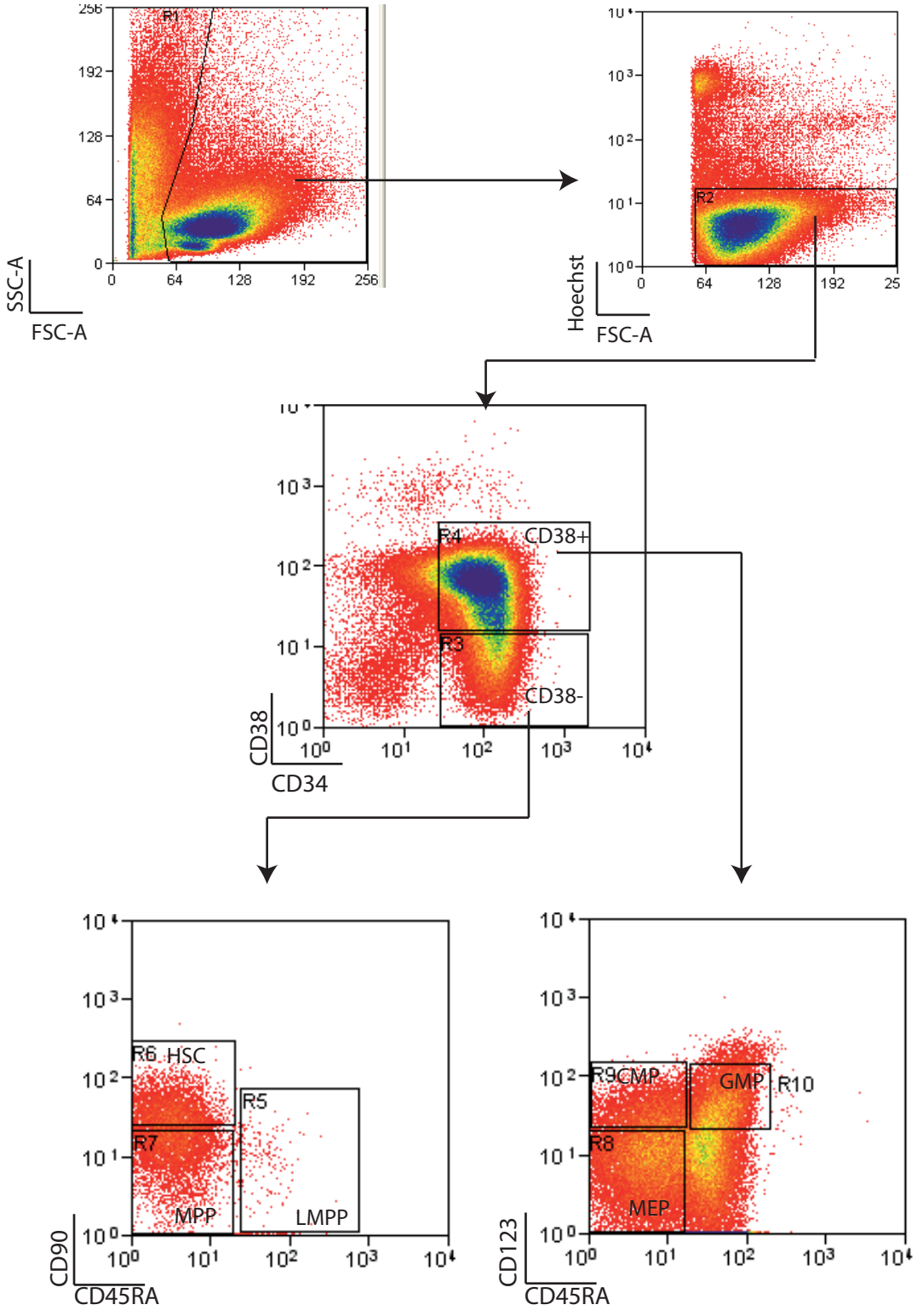


Fig 2.5 Comparing FACS sorters, Aria II sorter versus Moflo sorter. (Moflo).
The same sample, on the same day was sorted on two different FACS sorters to test sorter efficiency, Normal bone marrow sample: OX529 was used. FACS plots from FACS plots from sort and analysis on the Moflo.

A)

Sorter	Moflo	Aria II
	population % of CD34+ cells	
CD38-	14.5	13
CD38+	85.5	87
HSC	5.8	7.8
MPP	7.95	4.4
LMPP	0.2	0.6
CMP	6.57	22

B)

		Sort Sample	Sample population		Yield			
Population	sorter	Starting cell count x10^6	Frequency %	post sort purity %	Theoretical	Actual	Adjusted	% of theoretical yield sorted
LMPP	ARIA II	0.225	0.2	92	450	377	347	**77**
CMP	ARIA II	0.225	9	96	20,250	16,634	15,969	**79**
MPP	ARIA II	0.225	1.4	99	3,150	2,606	2,580	**82**
HSC	ARIA II	0.225	3.7	99	8,325	6,892	6,823	**82**
LMPP	Moflo	0.225	0.13	100	293	219	219	**75**
CMP	Moflo	0.225	4.12	47	9,270	7,826	3,678	**40**
MPP	Moflo	0.225	4.95	65	11,138	7,633	4,961	**45**
HSC	Moflo	0.225	2.2	90	4,950	3,838	3,454	**70**

Table 2.3 Comparing FACS sorters, details of Aria II sort versus Moflo sort.
A) Percentage each population is of CD34+ cells; after analysis on either the Moflow or Aria II. Percentages were calculated using analysis gates. B) Details of sort yields, for each population when sorted on either the Moflo or Aria II. (Frequency percentages were calculated using sort gates and purity percentages were calculated using analysis gates.)

Sorter was calculated based on yields obtained compared to the theoretical yield (Table 2.3). Again the Aria II was the most efficient. Therefore the Aria II was most efficient sorter available and sorting was continued on this machine.

2.2.5.3 Flow rate analysis

The flow rate used during sorting was also analysed to see if it could be optimized. The flow rate on the Aria can be set between 1 and 11, and for sorting and analysis purposes is set between 1 and 5. A CD34+ enriched pathological normal BM sample was used to assess the different flow rates. The sample was divided up into equal amounts, and aliquots were run through the cytometer at different speeds (Table 2.4).

The greatest number of cells lost was at the slowest flow rate 1, with the least cell losses at flow rate 3. Prior to optimization experiments I used a Flow rate 3, so continued using this speed as it was the most optimum.

flow rate	starting cell number	starting volume	flush out volume	cell number through cytometer	cell number through cytometer in flush out	total cell number through cytometer	cell loss
1	100000	100ul	50ul	38860	612	39472	61%
1	100000	100ul	50ul	35638	4354	39992	60%
3	100000	100ul	50ul	53463	4376	57839	43%
5	100000	100ul	50ul	52484	1787	54271	46%
5	100000	100ul	50ul	45815	2186	48001	52%

Table 2.4 Testing different FACS sorter flow rates.
One sample was split into equal cell numbers and volume, then each tube was run through the Aria II at different flow rates. Starting cell number was counted manually on a haemocytometer. Cell number through cytometer was from digital counts made by cytometer software. Sample used was a pathological normal. Samples were run in a random order, position in table does not reflect this.

2.2.5.4 Cell sorting optimization summary

The optimization analysis demonstrated that procedures I was using were the most efficient, and therefore could not be improved upon to increase LMPP yield.

2.3 RNA preparation

2.3.1 Extraction

Total RNA was extracted with either Trizol (Invitrogen) or with miRNeasy extraction kit (Qiagen) according to manufacturer's instructions. To Elute RNA from miRNeasy columns two elutions of 30µl of nuclease free water were used. After extraction if RNA was too dilute for next applications it was concentrated using a vacuum concentrator (SpeedVac) at 37°C.

To quantitate RNA concentration the Qubit fluorometer (Invitrogen) was used with the Qubit RNA assay kit (invitrogen).The kit and fluorometer were used according to manufacturer's instructions.

2.3.2 Checking RNA Quality

To check the RNA integrity the 2100 Bioanalyser (Agilent) was used with either the RNA 6000 Nano kit or the RNA 6000 Pico kit (Agilent), this was carried out according to manufacturer's instructions.

2.4 Microarray

2.4.1 Performing the arrays

Microarrays were performed using the Agilent miRNA microarray system. Agilent Human miRNA microarray release 15, 8X15K microarray slides were used. All RNA used was extracted using the miRNeasy kit. RNA was labelled and hybridized to the arrays using the Agilent miRNA complete labelling and hyb kit, according to manufacturer's instructions. Between labelling and hybridization samples were purified using the Micro-bio-Spin 6 (Bio-Rad) chromatography columns to remove unbound dye. After hybridization reaction was prepared it was pipetted on to a SureHyb gasket slide (Agilent), microarray slide was then carefully placed on top of gasket slide and slides were held together in a hybrizdation chamber (Agilent). Chamber was placed in a hybridization oven at 55°C rotating at 20rpm for 20 hours. After hybridization slides were washed using gene expression wash buffer kits according to manufacturer's instructions. Separation of gasket slide and array slide was done whilst both slides were totally submerged in wash buffer 1. Clean designated staining dishes were used

for wash steps. Directly after washing, array slides where scanned by the Agilent high resolution C scanner.

2.4.2 Array data processing

Feature extraction (FE) software version 10.7.3.1 (Agilent) was used to extract data from scanner produced image of microarray. To extract data grid template; 029297_D_F_20110707, FE protocol; miRNA_107_sep09 and QC metric set; miRNA_QCMT_sep09 were used. Data was extracted in full format to an Excel spread sheet (Microsoft).

2.4.3 Bioinformatics

Normalization: was done in R [108] with package; AgiMicroRNA, Robust Multi-array average (RMA) normalization was implemented for internal array normalization and quantile normalization was implemented to normalize between arrays. Normalization was done as in [6].

Array data analysis: multiple array viewer MeV version 4.5.0 software package [107] was used for array data analysis.

Statistical tests: were performed either in MeV or GraphPad prism software version 5.04

Gene set enrichment analysis (GSEA): was performed using gene set enrichment analysis software [121] [122].

Principle component analysis (PCA): Was performed in MeV software [107] the coordinates from the analysis were then imported into R and plots were generated in R [108].

2.5 RT-qPCR

RT-qPCR protocol was adapted from Jang et al [85]. All reagents used in RT-qPCR came from Life Technologies unless otherwise specified.

2.5.1 Reverse transcription

cDNA was prepared from 10ng of RNA extracted using the miRNeasy kit. cDNA was prepared using the TaqMan MicroRNA reverse transcriptase kit. Multiplex RT reactions were carried out in 15µl volume containing: 2µl of total RNA, 0.2µl of 100mM dNTPs, 1.5µl of Multiscribe reverse transcriptase (50 U/µl), 1.5µl 10 RT buffer, 0.2µl of RNase inhibitor (20 U/µl), 1.6 µl of nuclease free water and 8µl of RT primer mix (0.05X of each). For primer details see table 2.4. The reaction mixture was mixed and incubated on a thermocycler as follows; 16°C for 30 min, 42°C for 30 min and then 85°C for 5 min. cDNA was either used immediately or stored at -20°C.

2.5.2 Pre-Amplification

After reverse transcription cDNA was pre-amplified in a multiplex reaction. Pre-amplification reaction was carried out in a 10ul volume containing: 5µl of Taqman PreAmp master mix, 2.5µl of cDNA and 2.5µl pooled assay mix (0.2X of each). For assay details see table 2.4. The reaction mixture was mixed and incubated on a thermocycler as follows; 95°C for 10 mins, then 12 cycles of; 95°C for 15 secs and 60°C for 4 mins. Pre-amplified cDNA was either used immediately or stored at -20°C.

2.5.3 qPCR on the ABI 7500 Fast real time PCR system

Pre-amplified cDNA was diluted 1/20 with nuclease free water. qPCR reaction was carried out in a 20µl volume containing; 1µl of 20X assay, 5µl of diluted cDNA, 10µl of Taqman universal PCR master mix no AmpErase UNG and 4µl of water. qPCR reactions were done in a 96 well plate. Following conditions were used for qPCR 95°C for 10 mins, then 40 cycles of; 95°C for 15 secs and 60°C for 1 min.

2.5.4 qPCR on the Fluidigm biomark system

Pre-amplified cDNA was diluted 1/5 with nuclease free water. A 96 well plate was set up for assays, one well per assay was used. 3µl of 2x assay loading reagent (Fluidigm) was added to 3µl of each 20x assay, in the plate. A 96 well plate was set up for cDNA, one well per sample was used. 3µl of Taqman universal PCR master mix no AmpErase UNG and 0.3µl of GE sample loading reagent (Fluidigm) was added to 2.7µl of each diluted cDNA sample, in the plate. 300µl of control line fluid was injected into the accumulator on a 48.48 gene expression dynamic array (Fluidigm). The array was then loaded on to the IFC (integrated fluid circuit) controller MX (Fluidigm) to prime the array. After priming a multichannel pipette was used to pipette the assay and sample mixes into designated wells on the array. After samples and assays mixes had been pipetted on to the array, array was loaded back on to the IFC controller MX and the array was primed with samples and assays mixes. The array was then loaded into the biomark HD reader (Fluidigm) and the qPCR reaction was run using the following conditions: 95°C for 10 mins then 40 cycles of; 95°C for 15 secs and 60°C for 1 min.

2.5.5 qPCR analysis

Analysis of standard real time data was done using ABI 7500 software version 2.05. Analysis of Fluidigm data was done using Fluidigm Real-time PCR analysis software version 3.0.2. Threshold was set manually in the exponential phase of the amplification curve. Statistical analysis of data was carried out in GraphPad Prism software version 5.04.

Table 2.5 FACS antibodies

Name	Clone	Conjugate	Company
CD34	581	PE-Cy5	Beckman coulter
CD34	581	PE-Cy7	BD pharmingen
CD38	HIT2	FITC	Ebioscience
CD38	HIT2	PE-Texas red	BD
CD38	HIT2	Alexa Fluro 700	Ebioscience
CD45RA	HI100	PE	Ebioscience
CD45RA	HI100	FITC	BD pharmingen
CD90	ebioSE10	Biotin	Ebioscience
CD123	6H	PE-Cy7	Ebioscience
SAV		APC-efluro 780	Ebioscience
anti-mouse goat f(ab') 2		QDot605	Invitrogen
anti-mouse goat f(ab') 2		QDot655	Invitrogen
CD80	2D10.4	PE	Ebioscience
CD95	DX2	PE	Ebioscience
CXCR4	1295	PE	Ebioscience
CCR8	191704	PE	R & D Sytems
CCR8	191704	APC	R & D Sytems
CD2	RPA-2.10	Purified	Ebioscience
CD3	HIT3a	Purified	Ebioscience
CD4	RPA-T4	Purified	Ebioscience
CD7	ebio124-IDI	Purified	Ebioscience
CD8a	RPA-T8	Purified	Ebioscience
CD11b	ebiocb-CALLA	Purified	Ebioscience
CD14	ICRF44	Purified	Ebioscience
CD19	6ID3	Purified	Ebioscience
CD20	2H7	Purified	Ebioscience
CD56	MEM188	Purified	Ebioscience
GPA	HIR2	Purified	Ebioscience

Table 2.6 lineage antibody cocktails

Lineage cocktail	purified antibodies
AML	CD2, CD3, CD4, CD8a, CD19, CD20, and GPA
Normal	CD2, CD3, CD4, CD7, CD8a, CD11b, CD14, CD19, CD20, CD56 and GPA

Table 2.7 FACS antibodies used in each figure

Figure	antibodies
3.2-3.4	CD80 PE, CD95 PE, CXCR4 PE, CCR8 PE, CD34 PE-Cy5, CD38 PE-Texas Red, CD45RA FITC, CD123 PE-Cy7, CD90 Biotin, SAV APC-efluro 780, AML Lineage cocktail, anti-mouse goat f(ab') 2 Qdot655 + Hoechst
3.6	CCR8 APC, CD34 PE-Cy7, CD38 FITC, CD45RA PE + Hoechst
4.3, 4.4, 4.9-4.12, 4.17, 4.17, 4.5 (normal)	CD34 PE-Cy5, CD38 FITC, CD45RA PE, CD123 PE-Cy7, CD90 Biotin, SAV APC-efluro 780, Normal Lineage cocktail, anti-mouse goat f(ab') 2 Qdot605 + Hoechst
Normal sample sorts	CD34 PE-Cy5, CD38 FITC, CD45RA PE, CD123 PE-Cy7, CD90 Biotin, SAV APC-efluro 780, Normal Lineage cocktail, anti-mouse goat f(ab') 2 Qdot605 + Hoechst
4.5 (AMLs)	CD34 PE-Cy5, CD38 FITC, CD45RA PE, CD123 PE-Cy7, CD90 Biotin, SAV APC-efluro 780, AML Lineage cocktail, anti-mouse goat f(ab') 2 Qdot605 + Hoechst
AML sample sorts	CD34 PE-Cy5, CD38 FITC or Alexa Fluro 700, CD45RA PE, CD123 PE-Cy7, CD90 Biotin, SAV APC-efluro 780, AML Lineage cocktail, anti-mouse goat f(ab') 2 Qdot605 + Hoechst
4.13	CD34 PE-Cy5

Assay ID	Assay Name	miRBase Accession Number	miRBase Alias	Target Sequence
001001	RNU24	NR_002447	U24 snoRNA /U24 small nucleolar RNA	AUUUGCUAUCUGAGAGAUGGUGAUGA CAUUUUAAACCACCAAGAUCGCUGAUG CA
001002	RNU66	NR_002444	U66	GUAACUGUGGUGAUGGAAAUGUGUUA GCCUCAGACACUACUGAGGUGGUUCUU UCUAUCCUAGUACAGUC
001006	RNU48	NR_002745	U48	GAUGACCCCAGGUAACUCUGAGUGUGU CGCUGAUGCCAUCACCGCAGCGCUCUG ACC
001094	RNU44	NR_002750	U44	CCUGGAUGAUGAUAGCAAAUGCUGACU GAACAUGAAGGUCUUAAUUAGCUCUAA CUGACU
001093	RNU6B	NR_002752	U6	CGCAAGGAUGACACGCAAAUUCGUGAA GCGUUCCAUAUUUUU
001973	U6 snRNA	NR_004394	U6 snRNA	GUGCUCGCUUCGGCAGCACAUAUACUA AAAUUGGAACGAUACAGAGAAGAUUAG CAUGGCCCCUGCGCAAGGAUGACACGC AAAUUCGUGAAGCGUUCCAUAUUUU
000393	hsa-miR-17-5p	MIMAT0000070	hsa-miR-17-5p	CAAAGUGCUUACAGUGCAGGUAGU
000395	hsa-miR-19a	MIMAT0000073	hsa-miR-19a	UGUGCAAAUCUAUGCAAAACUGA
001014	hsa-miR-20b	MIMAT0001413	hsa-miR-20b	CAAAGUGCUCAUAGUGCAGGUAG
002438	hsa-miR-21#	MIMAT0004494	hsa-miR-21*	CAACACCAGUCGAUGGGCUGU
001271	hsa-miR-363	MIMAT0000707	hsa-miR-363	AAUUGCACGGUAUCCAUCUGUA
001535	hsa-miR-551b	MIMAT0003233	hsa-miR-551b	GCGACCCAUACUUGGUUUCAG
002895	hsa-miR-720	MIMAT0005954	hsa-miR-720	UCUCGCUGGGGCCUCCA
002194	hsa-miR-886-3p		hsa-miR-886-3p	CGCGGGUGCUUACUGACCCUU
000431	hsa-miR-92a	MIMAT0000092	gga-miR-92a	UAUUGCACUUGUCCCGGCCUGU
000465	hsa-miR-142-5p	MIMAT0000433	gga-miR-142-5p	CAUAAAGUAGAAAGCACUAC
001097	hsa-miR-146b	MIMAT0002809	hsa-miR-146b-5p	UGAGAACUGAAUUCCAUAGGCU
002642	hsa-miR-151-5P	MIMAT0004697	hsa-miR-151-5p	UCGAGGAGCUCACAGUCUAGU
000482	hsa-miR-181c	MIMAT0000258	hsa-miR-181c	AACAUUCAACCUGUCGGUGAGU
000486	hsa-miR-186	MIMAT0000456	hsa-miR-186(	CAAAGAAUUCUCCUUUUGGGCUU
000494	hsa-miR-195	MIMAT0000461	hsa-miR-195	UAGCAGCACAGAAAUAUUGGC
000512	hsa-miR-210	MIMAT0000267	mmu-miR-210	CUGUGCGUGUGACAGCGGCUGA
000525	hsa-miR-222	MIMAT0000279	gga-miR-222a	AGCUACAUCUGGCUACUGGGUCUC
000399	hsa-miR-23a	MIMAT0000078	hsa-miR-23a	AUCACAUUGCCAGGGAUUUCC
000404	hsa-miR-26a	MIMAT0000082	hsa-miR-26a	UUCAAGUAAUCCAGGAUAGGC
000407	hsa-miR-26b	MIMAT0000083	hsa-miR-26b	UUCAAGUAAUUCAGGAUAGGU
000602	hsa-miR-30b	MIMAT0000420	bta-miR-30b	UGUAAACAUCCUACACUCAGCU
000420	hsa-miR-30d	MIMAT0000245	hsa-miR-30d	UGUAAACAUCCCCGACUGGAAG
002147	hsa-miR-342-5p	MIMAT0004694	hsa-miR-342-5p	AGGGGUGCUAUCUGUGAUUGA
000425	hsa-miR-34a	MIMAT0000255	has-miR-34a	UGGCAGUGUCUUAGCUGGUUGUU
002116	hsa-miR-361-3p	MIMAT0004682	hsa-miR-361-3p	UCCCCCAGGUGUGAUUCUGAUUU

Table 2.8 RT-qPCR Taqman assays

Chapter 3: RESULTS I

NEW ANTIGENS TO ENHANCE AML IMMUNOPHENOTYPE

The aim of this section of work was to investigate whether additional cell surface antigens could be identified to better purify the GMP-like and LMPP-like cell populations within AML.

Previously in the lab GMP-like and LMPP-like populations in AML had been purified by flow cytometry [15]. mRNA expression profiles from the purified GMP-like and LMPP-like populations were generated and analysed. This revealed that GMP-like and LMPP-like cell populations in AML were molecularly distinct. However there was an overlap in their gene expression profiles, which suggested that these populations could be further purified.

Multiple antigens are used to define the GMP-like and LMPP-like populations when purified by flow cytometry (LMPP-like population: lin-, CD34+, CD38-, CD90-, CD45RA+ and GMP-like population: lin-, CD34+, CD38+, CD123+, CD45RA+) [15]. However the only antigen whose expression differs between the two populations is CD38. The GMP-like population is CD38+ and the LMPP population is CD38- (Figure 4.2). When CD38 expression is examined by flow cytometry, in both normal and leukemic marrow it is observed as a continuous population rather than distinct populations. Therefore it is more difficult to define the positive and negative fractions, and harder to sort populations based on its expression. As CD38 is the only antigen currently used to discriminate between the GMP-like and LMPP-like populations we may not be wholly purifying these populations from each other.

The CD38 mRNA expression level is variable within these populations (Figure 3.1). In addition the cDNA difference in mRNA expression between the GMP-like and LMPP-like populations is not consistent, with no difference in CD38 mRNA expression between the two populations in some samples (Figure 3.1). However there was difference at the protein level between these populations as this is what these populations were originally purified on, but highlights why this antigen may be not the best to purify these two populations.

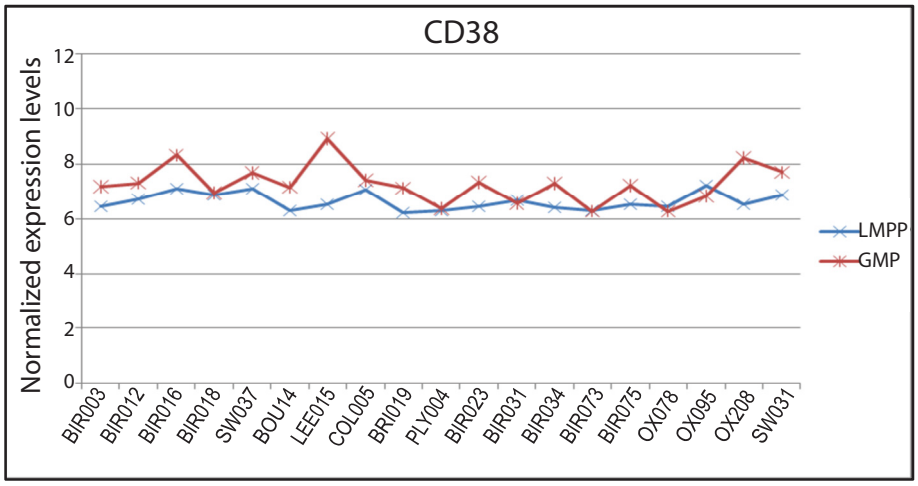


Figure 3.1 CD38 mRNA expression in GMP-like and LMPP-like cell populations in AML.
Graph showing the levels of CD38 mRNA expression from array data in patient matched GMP-like and LMPP-like cell populations in AML.

I set out to establish if these populations could be better purified. The question underlying this was if these populations could be better purified, would there be less overlap in their gene expression profiles. If further antigens were found to better purify the GMP-like and LMPP-like populations, further mRNA profiles would need to be generated to answer this question.

3.1 Cell surface antigens are differentially expressed at an mRNA level in GMP-like and LMPP-like populations

To better purify the GMP-like and LMPP-like cell populations in AML by FACS, I needed to identify antigens that would be expressed differently on the cell surface of the GMP-like and LMPP-like populations. Gene expression profiles for GMP-like and LMPP-like populations in AML, generated from work previously published in the lab [15] were used to identify potential antigens of interest. A standard t-test comparing the gene expression profiles of GMP-like and LMPP-like populations was performed. From the t-test a list of differentially expressed genes was identified, and from this list cell surface antigens were selected (Table 3.1). CD38 was amongst the most significantly differentially expressed cell surface antigen genes. To reduce the list of genes of interest I identified which antigens had commercially available antibodies suitable for flow cytometry, which can be seen highlighted in red (Table 3.1). This gave a list of four potential antigens: CXCR4, CD95, CCR8 and CD80. CXCR4 was the most promising candidate as it had a fold change in mRNA expression between the two populations of 1.8, this was greater than the fold change seen with CD38, 1.5. The other antigens had much lower fold changes in mRNA expression between the two populations, between 1.1-1.2. However as there were only four potential antigens with commercially available antibodies suitable for flow cytometry, all were selected for further investigation. There are three caveats with using the mRNA expression profiles to select potential antigens. First, the assumption is being made that the sample group used to generate the mRNA expression profiles is representative of all CD34+ GMP-like/LMPP-like expanded AML. Secondly we assume that the mRNA expression levels reflect total protein levels. Thirdly that mRNA levels reflect expression of protein at the cell surface.

Fold change in expression between GMP-like and LMPP-like	Gene Symbol	Gene Name	P Value	Expected to be more expressed on:
2.2	HMMR	Hyaluronan-mediated motility receptor	5.22E-05	GMP-like
2.2	HMMR	Hyaluronan-mediated motility receptor	1.89E-05	GMP-like
1.8	CXCR4	C-X-C chemokine receptor type 4	4.98E-03	GMP-like
1.5	CD38	cluster of differentiation 38	7.17E-04	GMP-like
1.3	CNIH4	cornichon homolog 4	9.16E-03	GMP-like
1.3	ERLIN1	ER lipid raft associated 1	9.34E-03	GMP-like
1.3	SPA17	sperm autoantigenic protein 17	5.22E-03	GMP-like
1.2	FAS (CD95)	TNF receptor superfamily, member 6	1.04E-04	GMP-like
0.9	PMEPA1	rostate transmembrane protein, androgen induced 1	6.74E-04	GMP-like
0.9	KCNH6	potassium voltage-gated channel, subfamily H (eag-related), member 6	8.21E-03	GMP-like
0.9	CACNA1G	alcium channel, voltage-dependent, T type, alpha 1G subunit	6.43E-03	GMP-like
0.9	OR56A4	olfactory receptor, family 56, subfamily A, member 4	3.77E-03	LMPP-like
0.9	CCR8	chemokine (C-C motif) receptor 8	9.47E-03	LMPP-like
0.9	PGLYRP1	peptidoglycan recognition protein 1	3.71E-03	LMPP-like
0.9	DSP	desmoplakin	9.61E-03	LMPP-like
0.9	EMR4P	egf-like module containing, mucin-like, hormone receptor-like 4 pseudogene	3.96E-03	LMPP-like
0.9	SLC28A1	solute carrier family 28 (sodium-coupled nucleoside transporter), member 1	5.74E-03	LMPP-like
0.9	KCNJ14	potassium inwardly-rectifying channel, subfamily J, member 14	9.75E-03	LMPP-like
0.9	TAS2R41	taste receptor, type 2, member 41	5.00E-03	LMPP-like
0.9	CHRNA4	cholinergic receptor, nicotinic, alpha 4	2.21E-03	LMPP-like
0.9	KCNH1	potassium voltage-gated channel, subfamily H (eag-related), member 1	5.24E-03	LMPP-like
0.9	CDH26	cadherin 26	1.69E-03	LMPP-like
0.8	CD80	CD80 molecule	3.08E-03	LMPP-like

Table 3.1 Cell surface antigens differentially expressed between GMP-like and LMPP-like cell populations in AML.

Gene list generated using a t-test comparing gene expression of GMP-like and LMPP-like cell populations in AML, array data from [15] was used. Table shows differentially expressed mRNAs encoding for cell-surface expressed antigens. A P-value cut off of ≤0.01 was used. Genes highlighted in red are antigens with commercially available antibodies suitable for FACS.

3.2 Antigen expression on the cell surface of GMP-like and LMPP-like cell populations within AML

Antibodies against the four potential antigens were added to the current antibody panel used to define GMP-like and LMPP-like populations, to explore whether they were expressed differently between the two populations.

Before the new antibodies were used to stain AML samples, they were first titrated to ensure the optimum amount of antibody was used during staining. Titration of antibody amount is critical to obtain the most accurate flow cytometry data. The standard method used to titrate antibodies in the laboratory, was to titrate antibodies using PBMCs from normal patient samples. As the antigens were identified in AML gene expression profiles, it was not guaranteed that the antigens would be expressed on normal BM cells. CXCR4, CD95 and CCR8 showed high levels of staining on normal PBMCs and were titrated using these cells. CD80 did not show any staining on normal PBMCs, so instead a cell line was used to titrate this antibody.

Within CD34+ AML two types of AML have been observed GMP-like/LMPP-like expanded AMLs and CMP-like/MPP-like expanded AMLs (ref). As the aim was to increase the purity of GMP-like and LMPP-like populations, GMP-like/LMPP-like expanded AMLs samples were needed to examine the expression of potential antigens. Four CD34+ AML samples were selected from the tissue bank to investigate the expression of the potential antigens. However information on whether the samples in storage were GMP-like/LMPP-like or CMP-like/MPP-like expanded AMLs was not known, so samples were selected blind. The CD34+ enriched fractions of samples OX290, OX295, BRI010 and BIR089 were chosen to investigate antigen expression.

Sample OX295 was a GMP-like/LMPP-like expanded AML (Figure. 3.2A). CXCR4 and CCR8 were expressed on cells within both the GMP-like and the LMPP-like populations. However, CXCR4 and CCR8 were both expressed on more cells in the GMP-like

population than the LMPP-like population (Figure 3.2B). There was no data for CD95 and CD80 expression available for this sample.

Sample OX290 was a GMP-like/LMPP like expanded AML (Figure. 3.2A). There was no CD80 expression on cells in either the GMP-like or LMPP-like populations in this sample (Figure. 3.2Bi). There was a small amount of cells expressing CXCR4 and CD95 in the GMP-like population, but no cells were expressing either antigen in the LMPP-like population. The level of CXCR4 and CD95 staining on GMP-like cells was very low, indicating this may have been background rather than genuine staining. (Figure. 3.2Bii/iii). There was no data for CCR8 expression available for this sample.

Sample BRI010 was difficult to classify as either a GMP-like/LMPP-like expanded AML or a CMP-like/MPP-like expanded AML. There were more GMP-like cells than CMP-like cells but more MPP-like cells than LMPP-like cells. However both the GMP-like and LMPP-like populations were expanded as compared to normal BM, it was therefore still informative to examine the expression of the potential antigens in the GMP-like and LMPP-like populations of this sample (Figure. 3.4A). There was no CD80 expression on cells in either the GMP-like or LMPP-like populations in this sample (Figure. 3.4Bi). There was a small amount of cells expressing CD95 but the level of staining was very low, and is difficult to distinguish from background (Figure. 3.4Bii). CXCR4 and CCR8 were expressed on cells within both the GMP-like and the LMPP-like populations. However CXCR4 and CCR8 were expressed on many more cells in the GMP-like population than the LMPP-like population (Figure. 3.4 Biii/iv).

Sample BIR089 was not a GMP-like/LMPP-like expanded AML it was a CMP-like/MPP-like expanded AML (Figure. 3.5A). Nonetheless expression of the potential antigens on the GMP-like and LMPP-like populations was still studied in this sample, as small GMP-like and LMPP-like populations were still present. None of the antigens were expressed

on cells in either the GMP-like population or LMPP-like population within this sample (Figure. 3.5B).

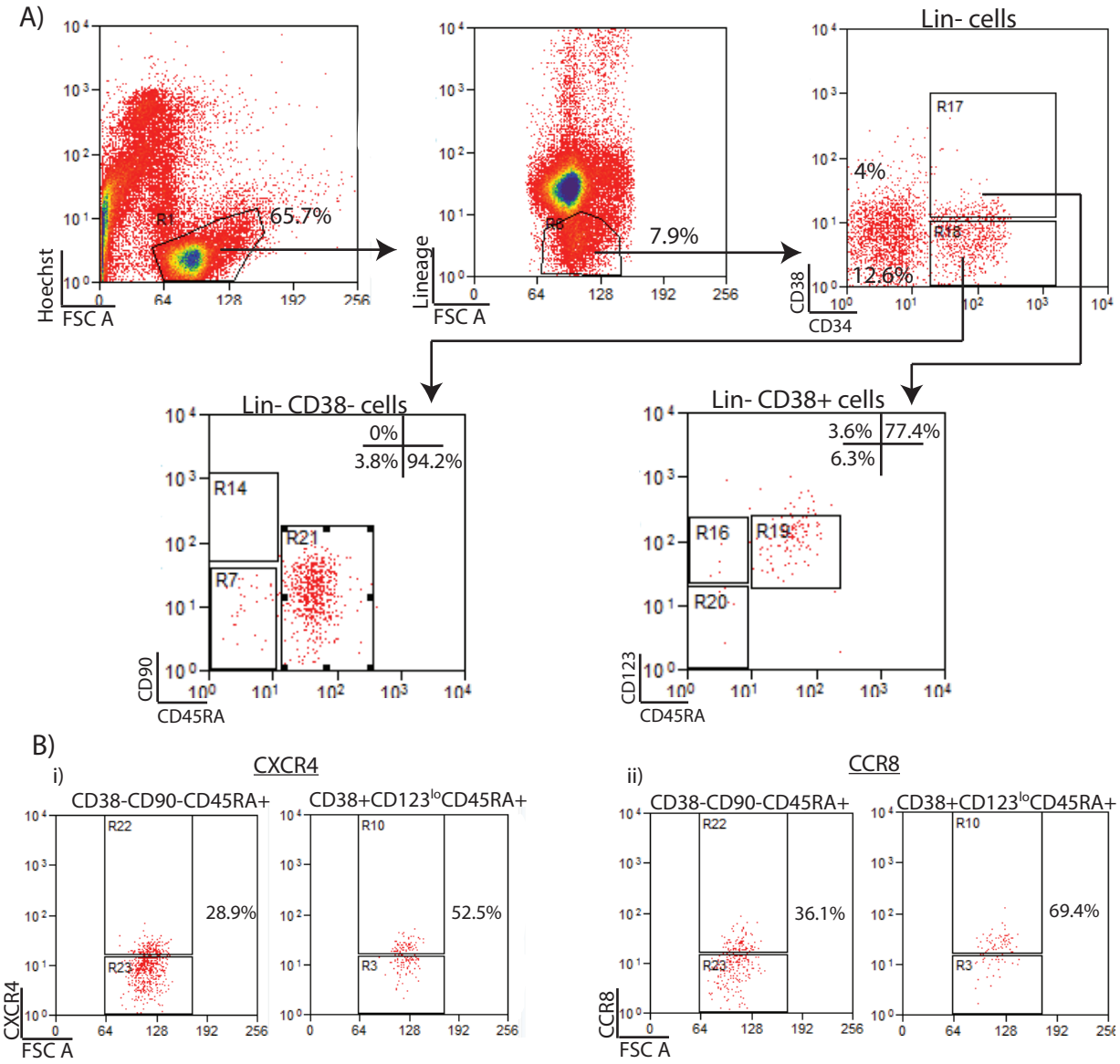


Figure 3.2 Cell surface antigen expression examined by FACS in AML. (AML sample: OX295)
Four antigens from Table 3.1 were selected and their expression on the surface of different cell populations in AML were examined by FACS. A) Immunophenotype of the sample. B) Expression of the four antigens, the left-hand panel shows the antigen expression in the LMPP-like population. The right-hand panel shows the antigen expression in the GMP-like population.

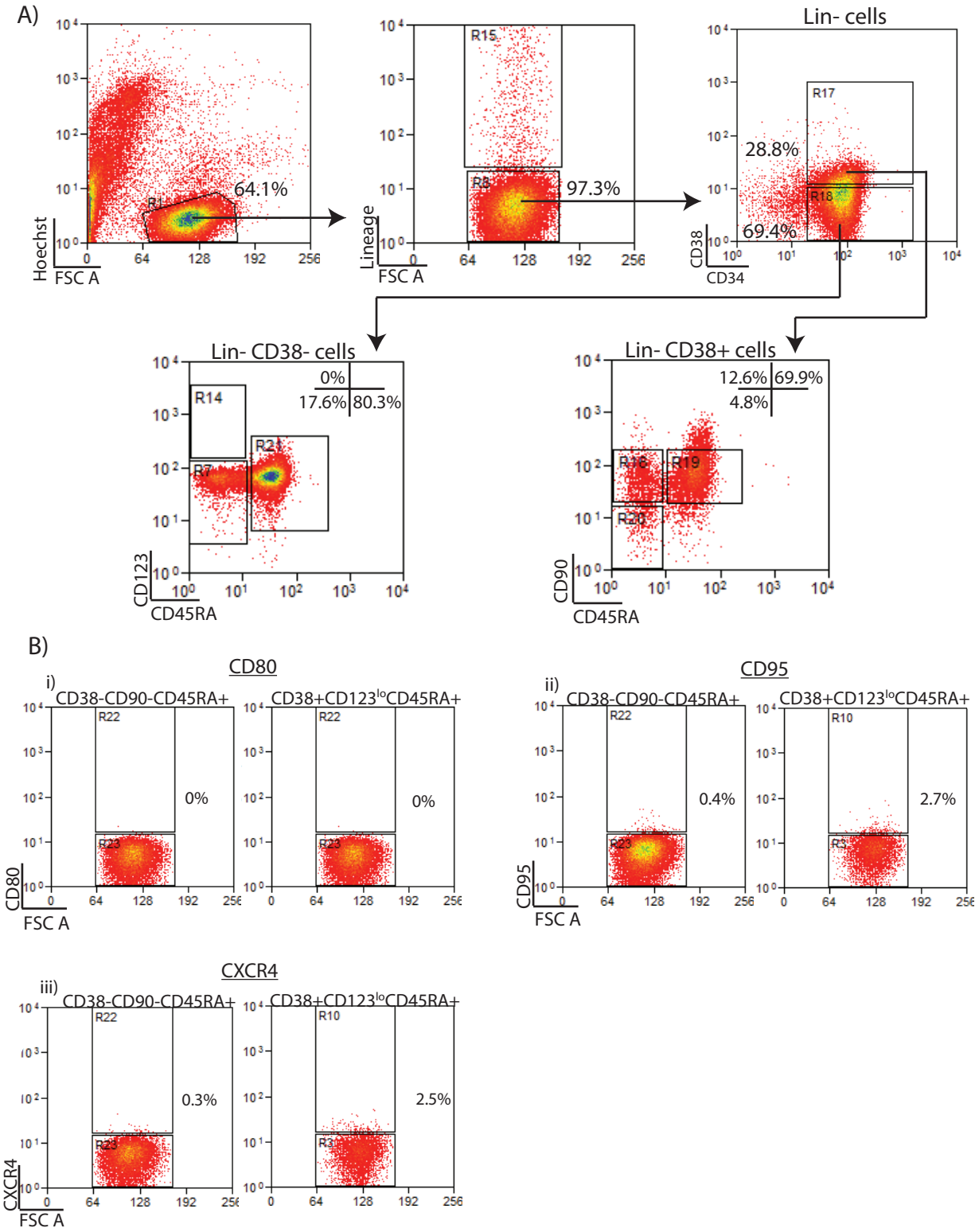


Figure 3.3 Cell surface antigen expression examined by FACS in AML. (AML sample: OX290)
Four antigens from Table 3.1 were selected and their expression on the surface of different cell populations in AML were examined by FACS. A) Immunophenotype of the sample. B) Expression of the four antigens, the left-hand panel shows the antigen expression in the LMPP-like population. The right-hand panel shows the antigen expression in the GMP-like population.

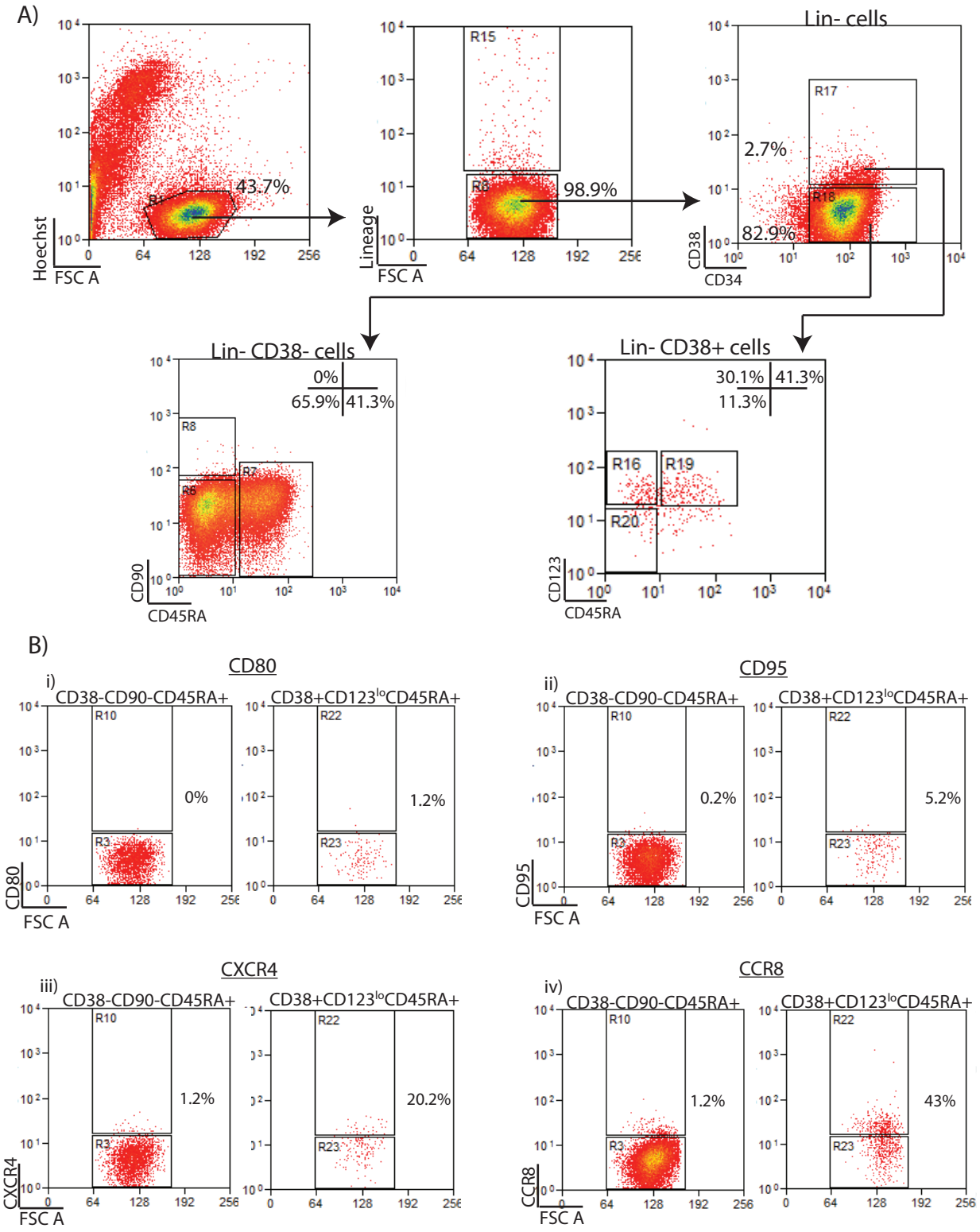


Figure 3.4 Cell surface antigen expression examined by FACS in AML. (AML sample: BRI010)
Four antigens from Table 3.1 were selected and their expression on the surface of different cell populations in AML eres examined by FACS, A) Immunophenotype of the sample. B) Expression of the four antigens, the left-hand panel shows the antigen expression in the LMPP-like population. The right-hand panel shows the antigen expression in the GMP-like population.

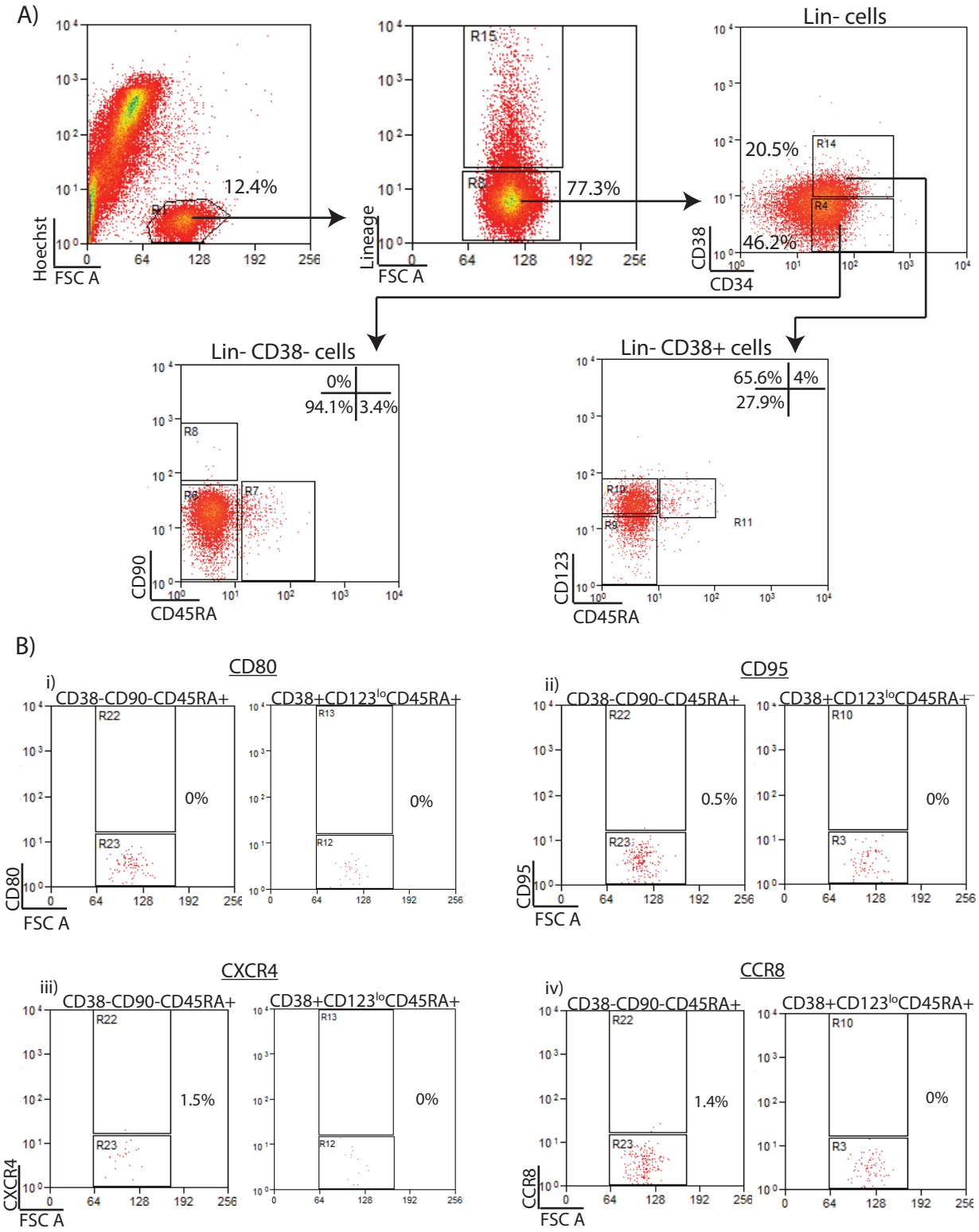


Figure 3.5 Cell surface antigen expression examined by FACS in AML. (AML sample: BIR089)
Four antigens from Table 3.1 were selected and their expression on the surface of different cell populations in AML were examined by FACS. A) Immunophenotype of the sample. B) Expression of the four antigens, the left-hand panel shows the antigen expression in the LMPP-like population. The right-hand panel shows the antigen expression in the GMP-like population.

3.3 Summary of antigens expression in AML

Flow cytometry expression data of the potential antigens in all four AML samples was compiled (Table 3.2). CD80 expression was not detected on cells in either the GMP-like or LMPP-like populations in any of the samples tested. CD95 expression was low on cells in both populations in all samples tested; and was difficult to differentiate from background staining. In the samples where CXCR4 and CCR8 were expressed, both antigens were expressed on more cells in the GMP-like population than the LMPP-like population. From the mRNA expression data CXCR4 was expected to be expressed on more cells in the GMP-like population, whereas CCR8 was expected to be expressed on more cells in the LMPP-like population. The CCR8 mRNA data and flow cytometry data were conflicting, however the mRNA expression may not accurately reflect protein levels as post-transcriptional and post-translation modifications occur in cells.

sample + antigen	% of LMPP-like cells	% of GMP-like cells	fold difference in % between GMP-like and LMPP-like cells
BIR089 + CD80	0	0	N/A
BIR089 + CD95	0.5	0	N/A
BIR089 + CXCR4	1.5	0	N/A
BIR089 + CCR8	1.4	0	N/A
OX290 + CD80	0.1	0.1	N/A
OX290 + CD95	0.4	2.7	6.8
OX290 + CXCR4	0.3	2.5	8.3
OX290 + CCR8	N/A	N/A	N/A
BRI010 + CD80	0	1.2	N/A
BRI010 + CD95	0.2	5.2	N/A
BRI010 + CXCR4	1.2	20.2	16.8
BRI010 + CCR8	1.2	43	35.8
OX295 + CD80	N/A	N/A	N/A
OX295 + CD95	N/A	N/A	N/A
OX295 + CXCR4	28.9	52.5	1.8
OX295 + CCR8	36.1	69.4	1.9

Table 3.2 Summary of selected antigens, cell surface expression in AML.
Table contains percentage of LMPP-like and GMP-like cells expressing antigen as determined by FACS in Figures 3.2-3.5.

3.4 CCR8 expression in a larger cohort of AML samples

Introduction of a programme to immunophenotype AMLs samples prior to storage allowed better classification of samples. This involved a flow cytometry analysis on a small aliquot of each sample before CD34 enrichment and storage. The flow cytometry analysis used a panel of antigens that could define whether a sample was a GMP-like/LMPP-like expanded AML or a CMP-like/MPP-like expanded AML. The immunophenotyping programme provided an opportunity to investigate the expression of one of the potential antigen in a larger cohort of AML samples. Only one antigen could be added to the panel due to restrictions on amount of sample available for immunophenotyping. CD80 and CD95 were not expressed on the first four AML samples tested and so were not considered. CCR8 and CXCR4 had been expressed on at least two of the first four AML samples; both were expressed on more cells in the GMP-like population than the LMPP-like population. However as CCR8 had a bigger fold difference in the number of cells it was expressed on between the GMP-like and LMPP-like populations than CXCR4, it was selected for further investigation.

CCR8 expression was analysed in the eight samples from the immunophenotyping programme that were CD34+ GMP-like/LMPP-like expanded AMLs (Figure. 3.6). It was predicted from the original AML samples that CCR8 would be expressed on more cells in the GMP-like population. In samples OX487, BIR262 and BIR264 there were more cells expressing CCR8 in the GMP-like population than in the LMPP-like population. However, in samples OX472 and BRI070, more cells expressed CCR8 in the LMPP-like population than the GMP-like population. In samples BRI069 (Figure 3.6vi) and PLY004 (Figure 3.6vii) there appeared to be no CCR8 expression, but it was difficult to be sure as there were very few cells to analyse due to high levels of cell death in these samples. In samples OX476 (Figure 3.6ix) and OX491 (Figure 3.6viii) there was no CCR8 expression in either of the two populations. Table 3.3 summarizes the number of cells expressing CCR8 in all eight samples.

Overall analysis of CCR8 in eight further samples, showed that CCR8 was not always more highly expressed on GMP-like cells than LMPP-like cells; the opposite could also be found. Additionally CCR8 was sometimes not expressed on either population. CCR8 therefore would not be a useful antigen to improve the purification of GMP-like and LMPP-like cell populations in AML.

Sample	% of GMP-like cells	% of LMPP-like cells	fold difference in % between GMP-like and LMPP-like cells
OX487	5	1.4	3.6
BRI069	5	6.3	1.3
OX476	0.7	0.1	N/A
OX472	8.7	19.2	0.5
BRI070	27.1	54.3	0.5
BIR262	70.3	49.8	1.4
BIR264	38.8	17.9	2.2
OX491	5.5	3	1.8
PLY004	0	2	NA

Table 3.3 Summary of CCR8 cell-surface expression in AML.
Table shows percentage of LMPP-like and GMP-like cells expressing CCR8 as determined by FACS in Figure 3.6.

3.5 Cell surface antigens differentially expressed at an mRNA level using an alternative t-test

From the above data and other work in our lab it was clear patient variation had an impact, when comparing the differences between GMP-like and LMPP-like populations in AML.

A paired t-test is a derivative of the standard t-test that is dependent on paired samples; it may reduce the effect of patient variability on data analysis. For this reason a paired t-test was used to compare the gene expression profiles of GMP-like and LMPP-like populations in AML. The paired t-test only used patient samples that had data available for both the GMP-like and the LMPP-like populations, and as part of the test the GMP-like and LMPP-like populations were paired by patient source. From the paired t-test a

Figure 3.6 i

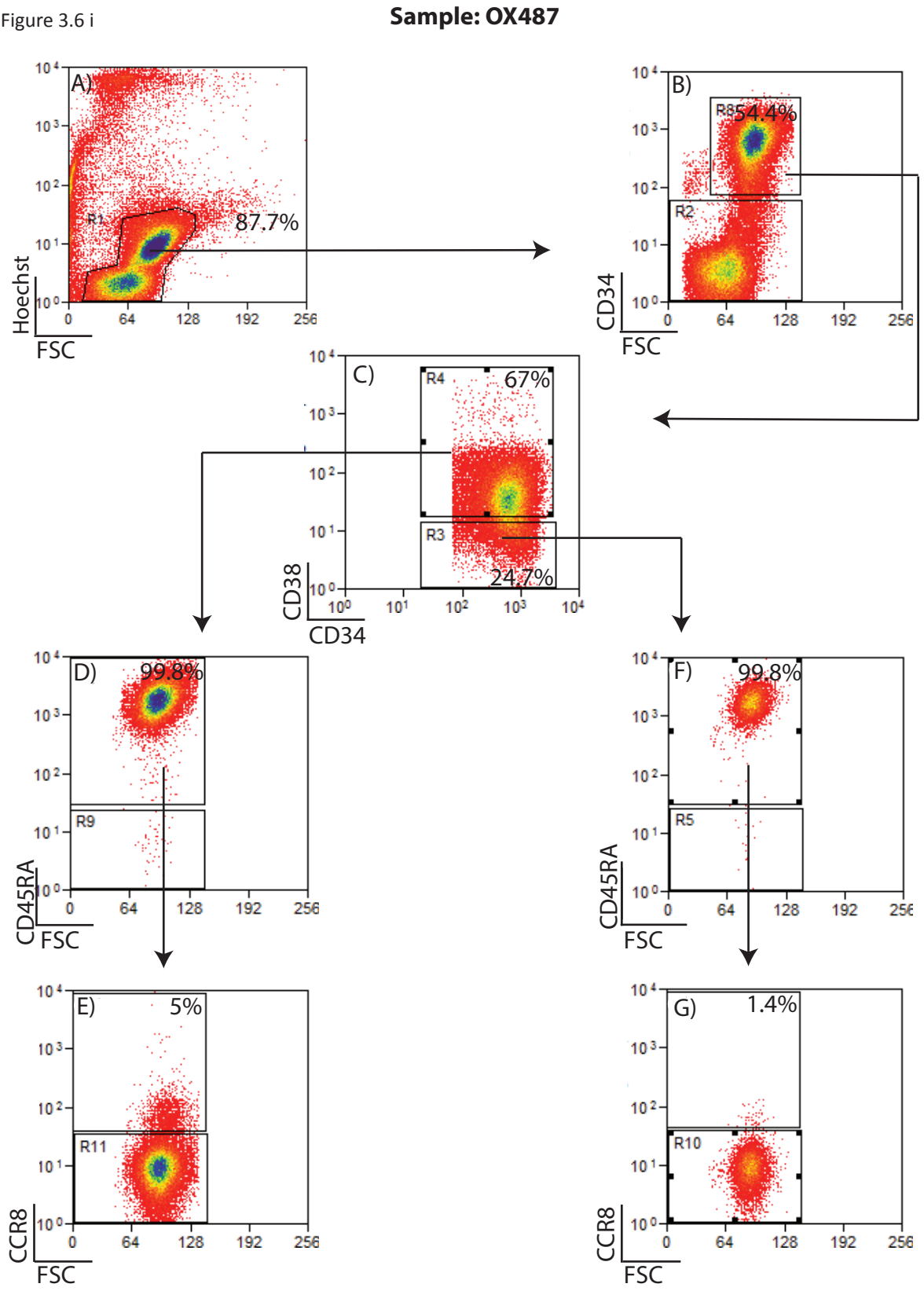


Figure 3.6 CCR8 cell-surface expression in a larger cohort of AML samples.
Before AML samples are CD34 separated an immunophenotypic profile is generated by FACS to enable better identification of the stored samples. CCR8 was added to this panel to allow further characterisation of its expression in AML. The FACS panel displaying the immunophenotype of each sample is shown. Plot E shows CCR8 expression in the GMP-like population, plot G shows CCR8 expression in the LMPP-like population.

Figure 3.6 ii

Sample: BIR262

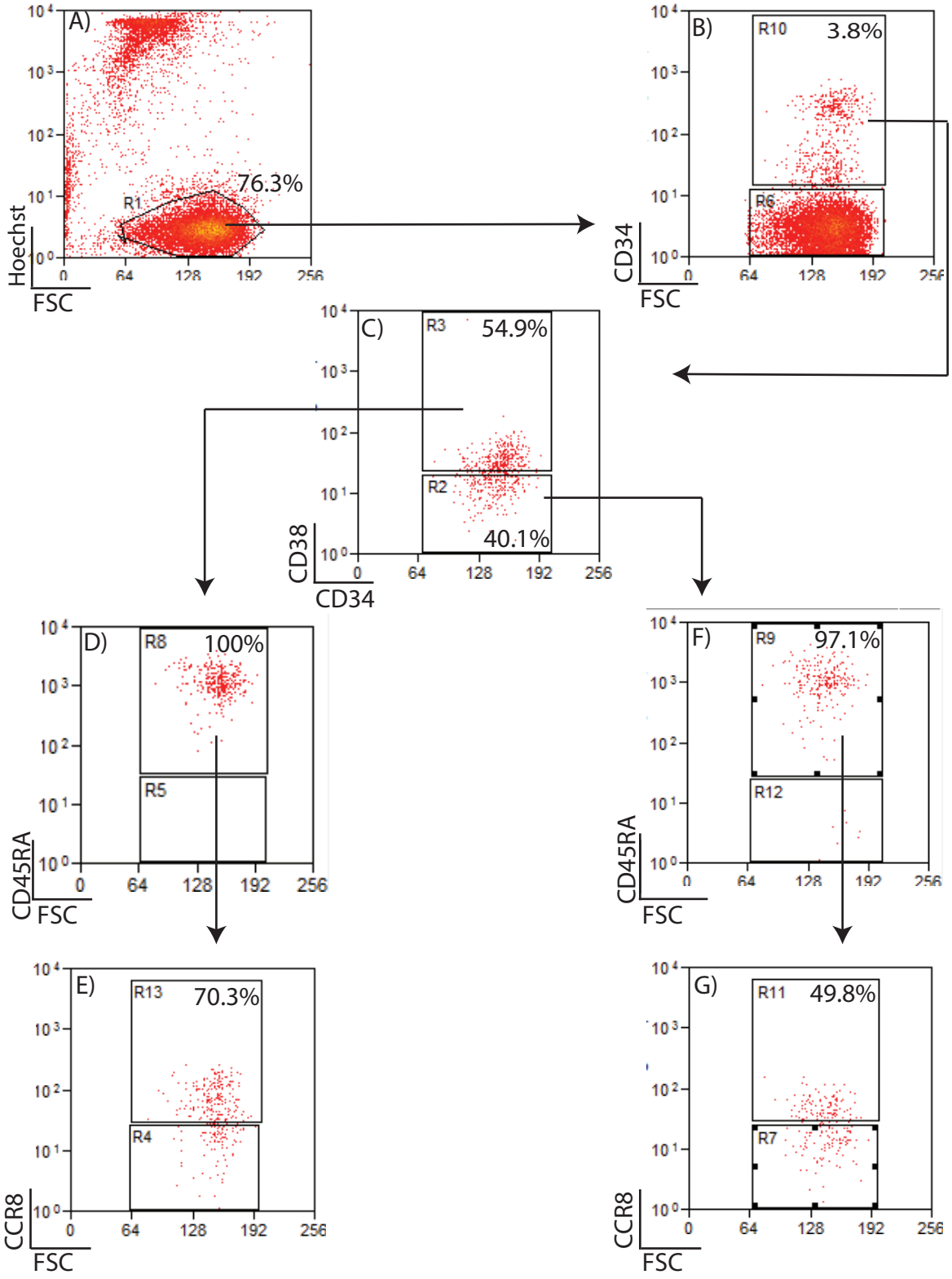


Figure 3.6 iii

Sample: BIR264

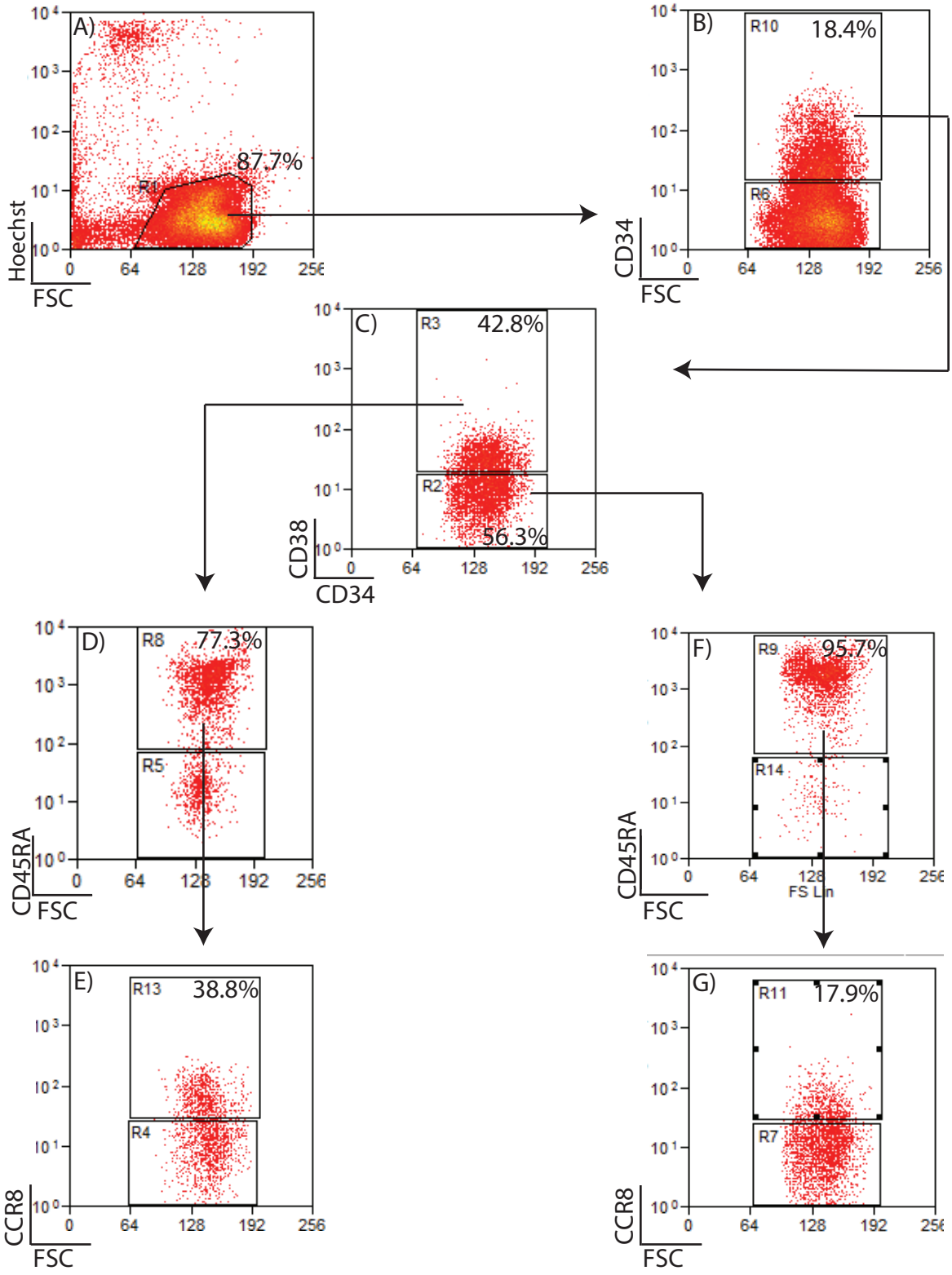


Figure 3.6 iv

Sample: OX472

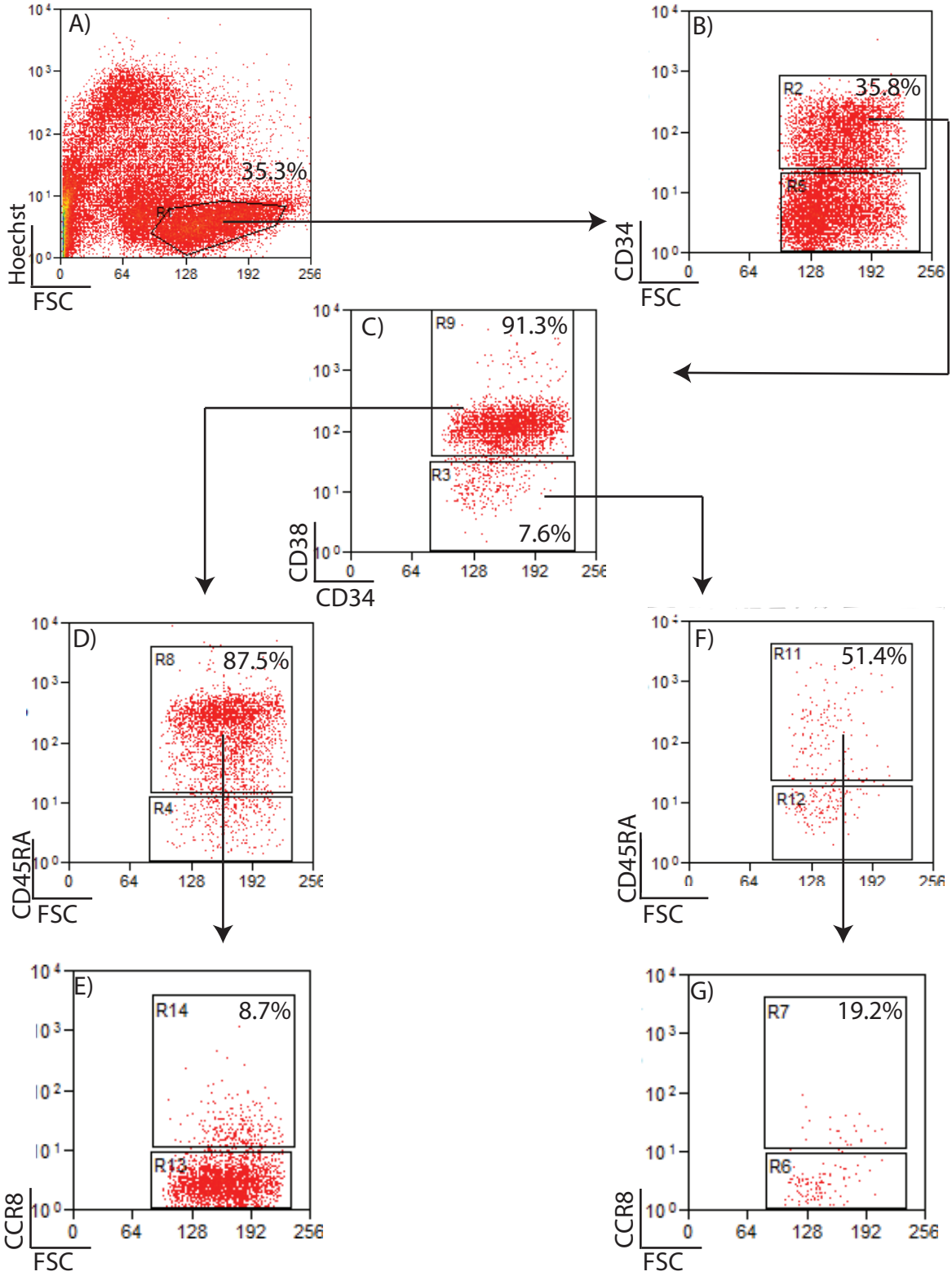


Figure3.6 v

Sample: BRI070

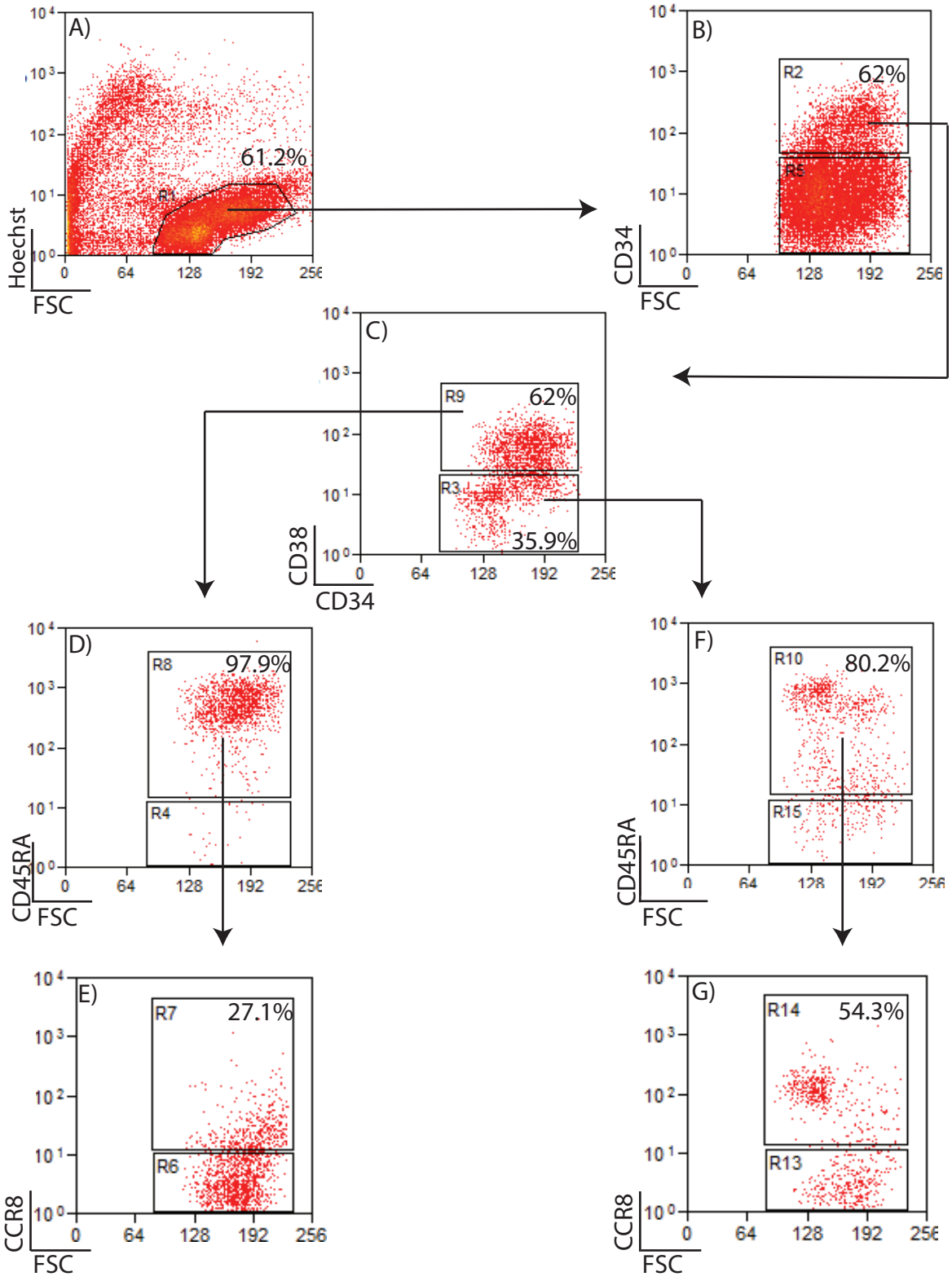


Figure 3.6 vi

Sample: BRI069

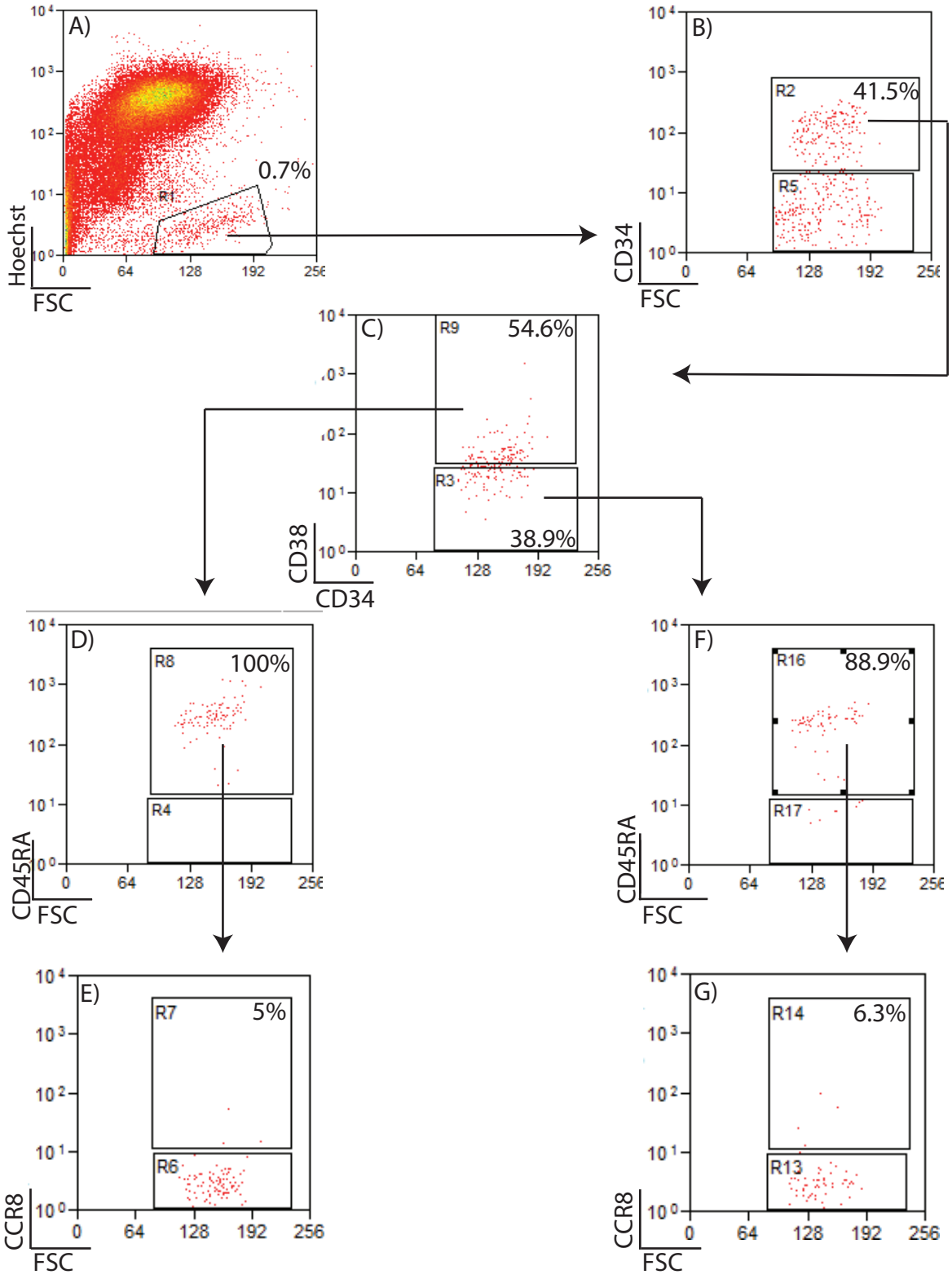


Figure 3.6 vii

Sample: PLY004

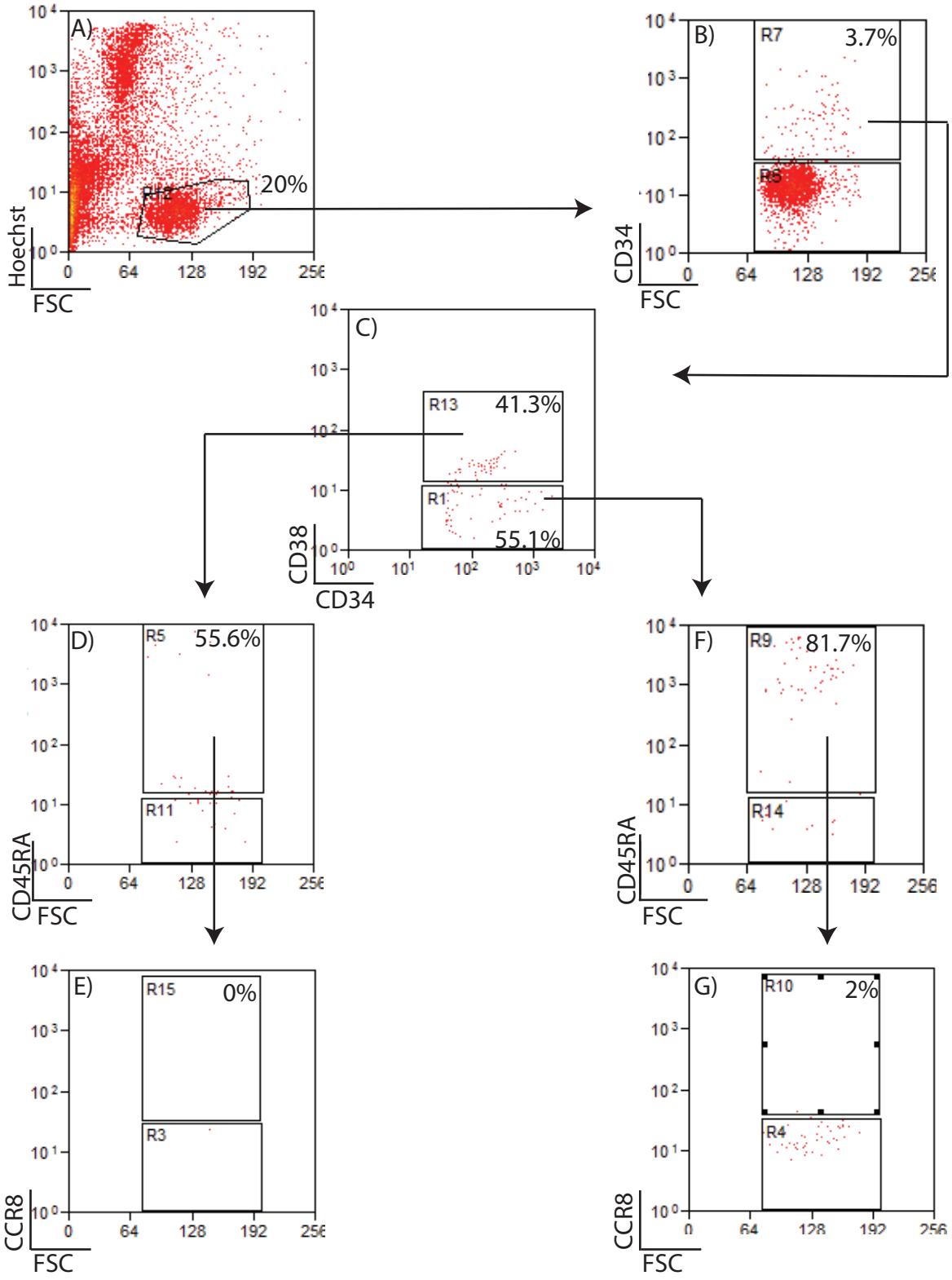


Figure 3.6 viii

Sample OX491

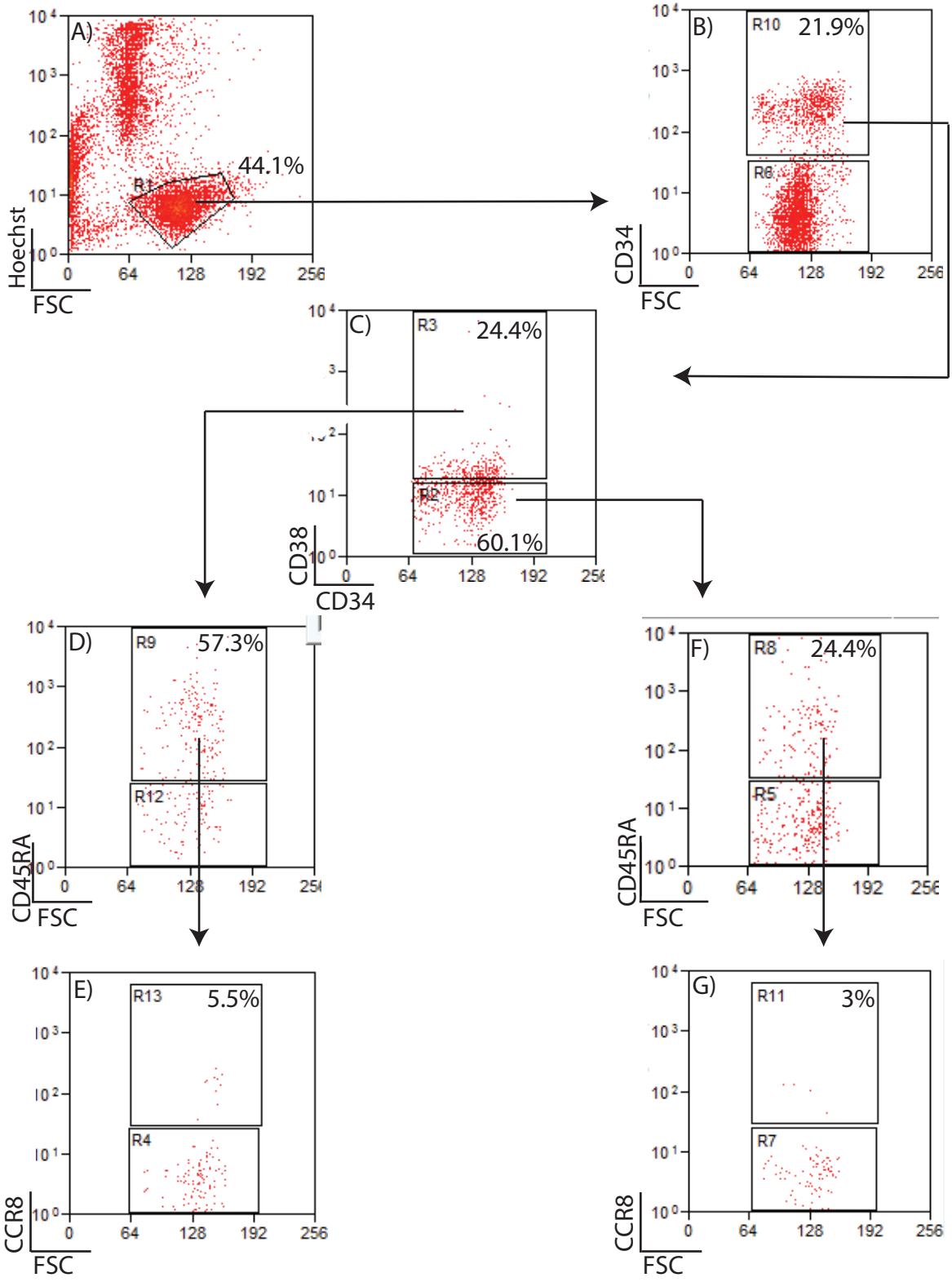
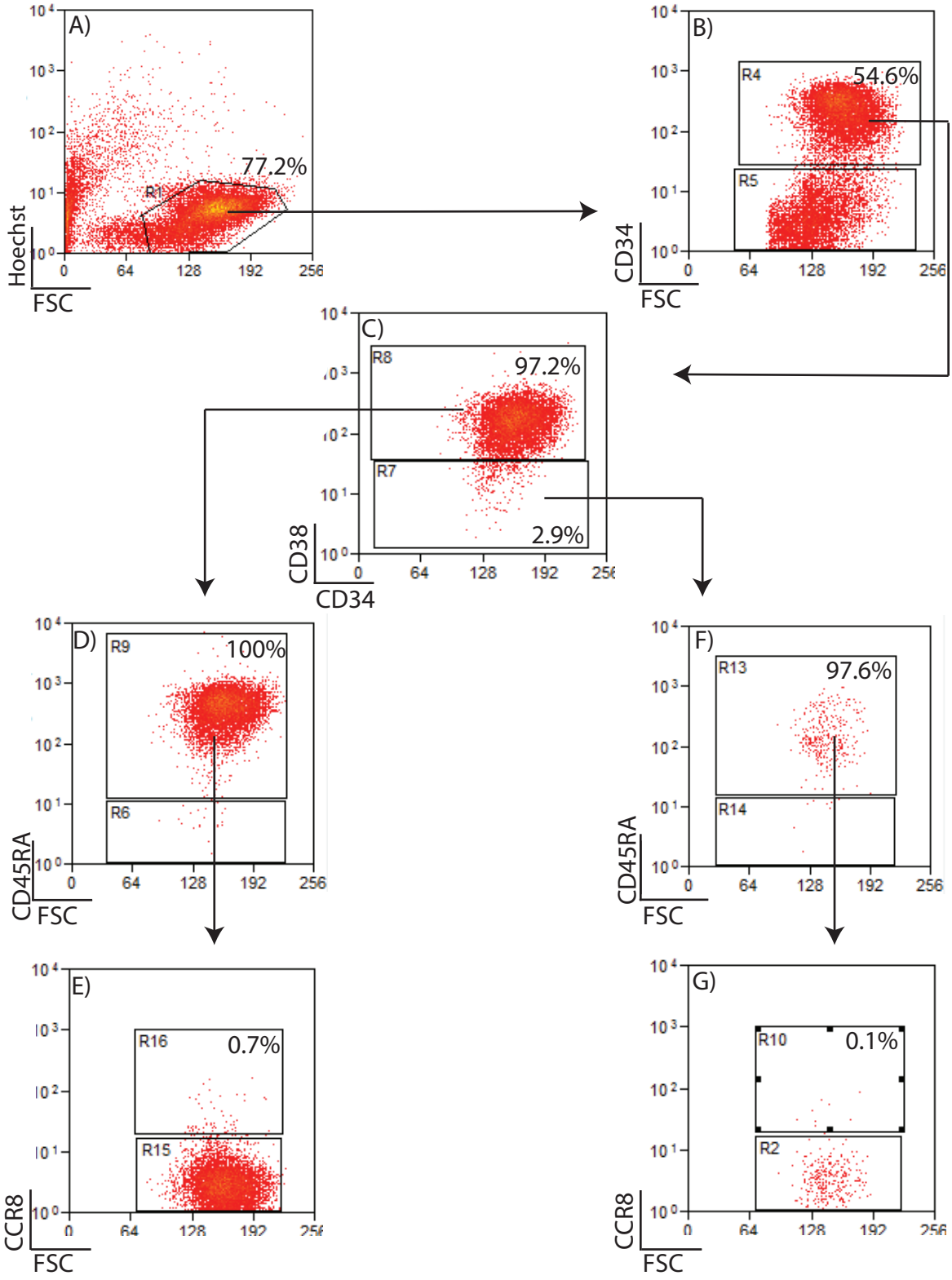


Figure 3.6 ix

Sample: OX476



list of differentially expressed genes was identified and from this list cell surface antigens were selected (Table 3.4). More cell surface antigen genes were identified as differentially expressed using the paired T-test: 72, than by standard T-test: 23. 17/23 of the genes identified as differentially expressed by the standard t-test were also identified by the paired t-test.

The majority of genes identified by the standard T-test were not present due to patient variability. However patient variability did mask significantly different genes, as more genes were identified as significant by the paired t-test. CD38 was identified as one of the most differentially expressed cell surface antigens by the both the paired and standard t-test. Although more genes were identified by the paired T-test than by the standard T-test, there were not many identified with a fold change in mRNA expression between GMP-like and LMPP-like populations of 1.5 or above.

The standard and paired T-test both identified Hyaluronan-mediated motility receptor (HMMR) as the most significantly differentially expressed antigen mRNA between the GMP-like and LMPP-like populations. HMMR also had the biggest fold difference in mRNA expression between the two populations in both tests. HMMR did not have a commercially available antibody suitable for flow cytometry. In order to investigate HMMR's cells surface protein expression by flow cytometry, I would need to produce an antibody and conjugate it to a flurochrome. As this would have taken a large amount of time and would not have been guaranteed to be successful. I decided it was too much of a high risk venture for a DPhil project. I therefore used the rest of my DPhil to investigate another aspect of AML.

Expected to be more expressed on GMP-like cells			
Fold change in expression between GMP-like and LMPP-like	Gene symbol	Gene name	P Value
2	HMMR	Hyaluronan-mediated motility receptor	5.96E-05
1.9	HMMR	Hyaluronan-mediated motility receptor	2.32E-05
1.5	CD93	CD93 molecule	7.21E-04
1.5	CD38	CD38 molecule	4.79E-04
1.5	NKG7	natural killer cell group 7 sequence	1.41E-03
1.5	PMP22	peripheral myelin protein 22	2.88E-03
1.4	FCGR1B	Fc fragment of IgG, high affinity Ib, receptor (CD64)	3.57E-03
1.4	SNTB1	syntrophin, beta 1	5.93E-04
1.4	CSF3R	colony stimulating factor 3 receptor	1.23E-03
1.4	CNIH4	cornichon homolog 4	1.61E-03
1.3	LMBR1	limb region 1 homolog	3.67E-03
1.3	KIAA0152	malectin	9.11E-03
1.3	FCGR1A	Fc fragment of IgG, high affinity Ia, receptor (CD64)	4.87E-04
1.3	OPN3	opsin 3	4.51E-04
1.3	ERLIN1	ER lipid raft associated 1	5.02E-03
1.3	IGSF6	immunoglobulin superfamily, member 6	8.43E-04
1.3	CD151	CD151 molecule	5.07E-03
1.3	NEXN	nexilin	5.97E-04
1.3	CELSR3	cadherin, EGF LAG seven-pass G-type receptor 3	2.47E-03
1.3	MUC1	mucin 1, cell surface associated	3.25E-03
1.3	AGTRAP	angiotensin II, type I receptor-associated protein	6.63E-03
1.3	SPA17	sperm autoantigenic protein 17	4.82E-03
1.2	FCGR1B	Fc fragment of IgG, high affinity Ib, receptor (CD64)	9.98E-04
1.2	ADORA2B	adenosine A2b receptor	9.77E-03
1.2	ADAM15	ADAM metallopeptidase domain 15	2.60E-03
1.2	KCNE1L	KCNE1-like	4.47E-03
1.2	SLC25A1	solute carrier family 25	3.31E-03
1.2	HFE	hemochromatosis	7.76E-03
1.2	CD276	CD276 molecule	7.24E-03
1.2	PMEPA1	prostate transmembrane protein, androgen induced 1	5.18E-05
1.1	PTPRU	protein tyrosine phosphatase, receptor type, U	4.81E-03
1.1	FAS	TNF receptor superfamily member 6	1.70E-04
1.1	KCNH6	potassium voltage-gated channel, subfamily H (eag-related), member 6	5.69E-04
1.1	GPR124	G protein-coupled receptor 124	4.33E-03
1.1	IL1RAP	interleukin 1 receptor accessory protein	6.19E-03
1.1	LRP8	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	3.04E-03
1.1	SDC2	syndecan 2	5.87E-03
1.1	CACNA1G	calcium channel, voltage-dependent, T type, alpha 1G subunit	9.93E-03
1.1	KCNAB1	potassium voltage-gated channel, shaker-related subfamily, beta member 1	9.55E-03

Expected to be more expressed on LMPP-like cells			
Fold change in expression between LMPP-like and GMP-like	Gene symbol	Gene name	P Value
1.1	PCDHB14	protocadherin beta 14	2.13E-03
1.1	FGFR4	fibroblast growth factor receptor 4	8.58E-03
1.1	PLXNB3	plexin B3	2.61E-04
1.1	OR56A4	olfactory receptor, family 56, subfamily A, member 4	4.77E-03
1.1	FZD1	frizzled family receptor 1	3.21E-03
1.1	GRB7	growth factor receptor-bound protein 7	4.35E-03
1.1	PSD2	pleckstrin and Sec7 domain containing 2	9.19E-03
1.1	RGS5	regulator of G-protein signaling 5	3.85E-03
1.1	SYNGR3	synaptogyrin 3	9.59E-03
1.1	ABCD1	ATP-binding cassette, sub-family D (ALD), member 1	6.97E-03
1.1	CHRNA4	cholinergic receptor, nicotinic, alpha 4	9.43E-03
1.1	GPR151	G protein-coupled receptor 151	8.61E-03
1.1	SLC22A18	solute carrier family 22, member 18	6.01E-03
1.1	ITGB1BP3	integrin beta 1 binding protein 3	2.43E-03
1.1	CGN	cingulin	9.08E-03
1.1	CLDN15	claudin 15	9.90E-03
1.1	KCNJ14	potassium inwardly-rectifying channel, subfamily J, member 14	6.17E-03
1.1	DSP	desmoplakin	5.00E-03
1.1	PGLYRP1	peptidoglycan recognition protein 1	1.23E-03
1.1	TAS2R41	taste receptor, type 2, member 41	2.06E-03
1.1	SLC28A1	solute carrier family 28	6.76E-03
1.1	EMR4	EGF-like module containing, mucin-like, hormone receptor-like sequence 4	7.46E-03
1.1	OR13C8	olfactory receptor, family 13, subfamily C, member 8	1.85E-03
1.1	NLGN4X	neuroligin 4, X-linked	6.53E-03
1.1	ACVR2A	activin A receptor, type IIA	2.85E-03
1.2	BTLA	B and T lymphocyte associated	9.41E-03
1.2	NLGN4X	NLGN4X	8.92E-03
1.2	TLR1	toll-like receptor 1	2.43E-03
1.2	GPR81	G protein-coupled receptor 81	5.16E-03
1.3	F2RL1	coagulation factor II (thrombin) receptor-like 1	3.09E-03
1.3	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	6.27E-03
1.5	GPR56	G protein-coupled receptor 56	4.07E-03
1.5	EFNA1	ephrin-A1	5.78E-03

Table 3.4 Cell surface antigens differentially expressed between GMP-like and LMPP-like cells populations in AML, when using an alternative type of t-test.

Gene list was generated using a paired t-test comparing gene expression of GMP-like and LMPP-like cell populations in AML, array data from [15] was used. Only patient samples with data for both cell populations were used. LMPP-like and GMP-like populations were matched by patient source. Table shows differentially expressed mRNAs encoding for cell-surface expressed antigens. A P-value cut off of ≤0.01 was used. Genes highlighted in red are antigens with commercially available antibodies suitable for FACS.

3.6 Discussion I

The aim of this part of my project was to better purify the GMP-like and LMPP-like populations. However for a number of reasons this was not possible.

Four antigens were identified using a t-test on gene expression profiles, which may have been able to increase purification of the two populations. Of these four staining for CD80 and CD95 was not seen in any of the first four AMLs samples used to explore antigen expression. However staining looked promising for CXCR4 and CCR8; CCR8 was taken forward to be looked at in a larger cohort of AMLs. Unfortunately, CCR8 expression varied between AML patient samples.

Many groups have conducted work to find antigens that are differentially expressed between AML LSCs and normal HSCs. The number of antigens found to be differentially expressed between these populations is growing. However a common problem exists, the variability of antigen expression amongst different patients [49]. CD32 and CD25 were both found to be more expressed on AML LSCs than HSCs by Saito et al but CD32 was only expressed in 34.4% of patients and CD25 in 24.6% of patients [52]. CD96 is another antigen found to be more highly expressed on AML LSCs than HSCs but was also only found to be expressed on 68% of patients. However TIM3 was found to be expressed in 90% patients, as well as being more expressed on AML LSC and bulk cells than HSCs [127]. The current literature highlights how difficult it is to find antigens that are consistently expressed across different AML patients however antigens such as TIM3 demonstrate although this is a difficult task it is not an impossible task.

When comparing gene expression profiles of populations from different sample donors, individual genetic alterations may affect results. In normal cell populations this is less of a problem, as there will be fewer of these changes between normal individuals. However, in diseases such as cancer, where the genome is highly unstable, many

different genetic alterations occur. This is known as the patient effect and can make comparing populations trickier. A standard t-test will be sensitive to the patient effect. A paired t-test can overcome this as it uses only patient samples with both a GMP-like and LMPP-like population, which are also paired by patient source as part of the test.

I became aware that the patient effect may have skewed data in our initial comparison. To investigate if this was the case, we performed a paired t-test. Most of the genes from the first t-test were found present in the paired t-test list. This makes it unlikely that the majority of the genes classified as significantly different in the standard t-test were present due to the patient effect. However, three of the genes I tested – CD80, CXCR4 and CCR8 – were only found to be significant by the standard t-test. These antigens may have been identified as significantly different due to the patient effect. This may explain why they were not able to successful further purify the two different populations.

A caveat for this work was that mRNA levels may not accurately reflect protein levels as many post-transcriptional and post-translational regulation events occur in cells. In addition, when we use FACS, we are specifically using antibodies that bind proteins on the cell surface, protein expression level on the cell surface, may also not be accurately reflected by the cells' total protein levels. A better way to identify differentially expressed antigens would be by directly examining antigen expression on the cell surface. One method to do this would be a large scale FACS screening of different cell surface antigens. However this would require availability or generation of a larger number of reagents. Another method would be an antibody phage display library [128]. Using either of these methods would identify antigens expressed on the surface of our cells of interest, as well as which antigens were more highly expressed on one cell type than another. However, both these methods would be expensive and require a larger amount of samples and time than were available during this project.

Another way to address my original aim would be to identify antigens that are differentially expressed between GMP and LMPP populations in normal samples. A much simpler functional read-out experiment could also be done to assess these populations, such as in vitro cell differentiation assays. Using normal samples would reduce the patient effect. However a potential problem with this approach is that while an antigen may separate these populations in normal BM, it may not do the same in AML BM.

Due to the issues surrounding availability of reagents and patient samples and the amount of time it would take I decided not to take this part of my project further.

Future work

All though I was not able to further purify the GMP-like and LMPP-like populations. If in future these populations are further purified, an important question is, will the functionality of these populations change. To address this serial xenograft transplantation assays could be used to read out LSC frequency. As using the current antigen panel, the GMP-like population had a lower LSC frequency than the LMPP-like population.

HMMR was identified by both t-tests as the most differentially expressed cell surface antigen. HMMR has been shown by multiple groups to be expressed in AML and has been considered as a target of immunotherapy [129-132]. Recent work has come to light that this antigen may be expressed on CD34+CD38+ cells in AML but not expressed on CD34+CD38- cells [133]. There is also now a commercially available antibody against this antigen. Therefore if this project was to continue HMMR would be an interesting candidate to take forward to further separate the GMP-like and LMPP-like AML cells.

Chapter 4: Results II Sorting and preparing samples for Microarray

The aim this part of my project was to investigate whether microRNAs are aberrantly expressed in leukemic cells compared to normals cells. To do this, the Agilent miRNA microarray system was chosen to profile microRNAs in our cells of interest. As getting large amounts of RNA from primary cells is difficult, the Agilent system was chosen because it required low RNA amounts compared to other microarrays systems. Alternative options to microarrays that were available when beginning this study were also considered. qPCR arrays could have been used to profile microRNA expression, but were excluded, as they were a more costly option than microarrays and would have taken a much greater amount of time to run. RNA sequencing had also been used to profile microRNA expression, but this was ruled out due to the large amounts of RNA needed.

The cell populations to be profiled were the GMP-like and LMPP-like populations in AML, as well as the normal stem and progenitor populations HSC, MPP, LMPP, CMP and GMP. The GMP-like and LMPP-like AML populations were chosen because GMP-like/LMPP-like expanded AMLs occur more frequently (~80% of CD34+ AMLs), than CMP-like/MPP-like expanded AMLs. Therefore the sample availability for the GMP-like/LMPP-like expanded AMLs was much higher. Also these populations had been analysed using xenograft and in vitro culture experiments previously in the lab, therefore more is known about the GMP-like/LMPP-like populations [15]. The normal populations were chosen as these were the populations most closely resembling the leukemic ones on an immunophenotypic and molecular level.

In order to generate microRNA profiles the different cell populations had to be purified and their RNA extracted, the following chapter details this process.

4.1 RNA extraction method

The first stage in preparing for the array was to choose the best method to extract RNA from our cells of interest. Filter based spin column kits are commonly used and are the routine method for RNA extraction in the lab. However these kits are not suitable for extracting RNA when investigating microRNA expression as they do not retain small RNAs. An alternative method of extracting total RNA, which retained small RNAs was therefore needed. Two methods were available to extract total RNA including small RNAs that would be compatible with the array system: Trizol extraction and kits specifically designed for the retention of small RNAs, such as the Qiagen miRNeasy kit.

Trizol is a solution containing phenol and guanidine salts. It is used as a method of organic extraction. When cells are lysed in the presence of phenol and guanidine salts, the risk of RNA being degraded is reduced, as they inhibit RNase. The sample is first lysed in Trizol. It is then centrifuged, which results in the sample separating into three phases, the upper aqueous phase containing the RNA is removed. RNA is then recovered by alcohol precipitation, the RNA is pelleted by centrifugation and rehydrated. The miRNeasy kit combines the organic and spin column extraction methods. Cells are also lysed with a phenol containing solution, but RNA is recovered using alcohol precipitation and a spin column. To avoid unnecessary technical bias it was important that one method of RNA extraction should be selected and consistently used. The literature suggested that using kits produced better quality RNA and that the best kit for this was the Qiagen miRNeasy kit [134].

To test if this was the case in my hands, I extracted RNA from K562 cells, using trizol and the Qiagen miRNeasy kit, and compared the results (figure 4.1). Both methods consistently extracted similar amounts of RNA. Figure 4.1A shows an example of RNA extracted by both methods run on an agarose gel: clear 28s and 18s bands can be seen for both RNA samples. There is also no evidence of degradation products in either RNA sample. Therefore both methods extracted RNA with good integrity. It was crucial that the extraction method retained small RNAs; this was confirmed for both methods by the presence of small RNAs on the gel.

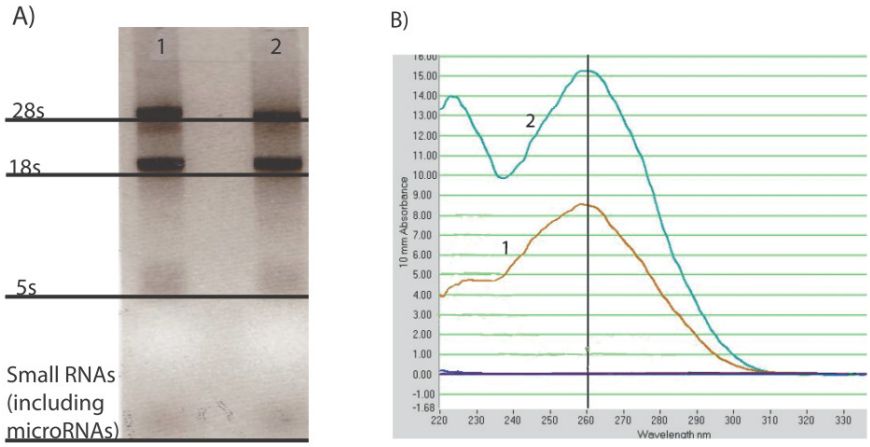


Fig 4.1 RNA quality using two different preparative methods.
Total RNA was extracted from K562 cells using either Trizol or the Qiagen microRNeasy kit, to identify the best method for extracting RNA to be used on the microarray. (Sample 1 is RNA extracted by the Qiagen kit and Sample 2 is RNA extracted by Trizol). A) Extracted RNA was run on an agarose gel to check RNA integrity and to check for the presence of small RNAs. B) RNA was analysed on the nanodrop spectrometer, to look for the presence of contaminants, plot generated is shown.

Another important consideration in choosing an RNA extraction method was the presence of contaminants left over from the extraction process, as these will affect the RNA quantification and bioanalyser analysis (used to determine RNA integrity). Contaminants may also interfere with array reactions. When analysing RNA on a spectrometer, RNA is seen at an absorbance of 260nm, salt contaminants are seen at an absorbance of 230nm and protein contamination is seen at an absorbance of 280nm. Figure 4.1B shows an example of an absorbance spectre from RNA extracted by both methods analysed on a spectrometer. No protein contamination was seen in RNA extracted by either method. However RNA extracted by Trizol consistently showed a much bigger peak at 230nm than RNA extracted by the miRNeasy kit,

indicating that there is a greater carryover of salt contaminants using a Trizol extraction. The miRNeasy kit extracted RNA of an equal amount and integrity to Trizol but less salt contaminants were present in the miRNeasy extracted RNA. Therefore the miRNeasy kit was chosen to extract RNA for the array.

4.2 Sorting samples for microarray

For the microarray analysis, the different populations from normal and leukemic BM had to be purified. To do this, BM samples were enriched for CD34+ cells then FACS sorted. The FACS panel used to separate the different populations can be seen in Figure 4.2. The HSC population is defined as Lin-, CD34+, CD38- , CD90+ and CD45RA-, the MPP as Lin-, CD34+, CD38- , CD90- and CD45RA-, the LMPP as Lin-, CD34+, CD38- , CD90- and CD45RA+, the CMP as Lin-, CD34+, CD38+, CD123+ and CD45RA- and the GMP as Lin-, CD34+, CD38+, CD123+ and CD45RA+. The average age of presentation in AML is around 70 years, so normal marrow samples were obtained from control donors from a similar age group to ensure the most relevant control was used. The example shown in Figure 4.2 is a representative normal age matched BM sample (sample OX709). After sorting each BM sample, the sorted populations were reanalysed to check the sort purity, an example of which can be seen in Figure 4.3 It was important to use samples with high sort purity.

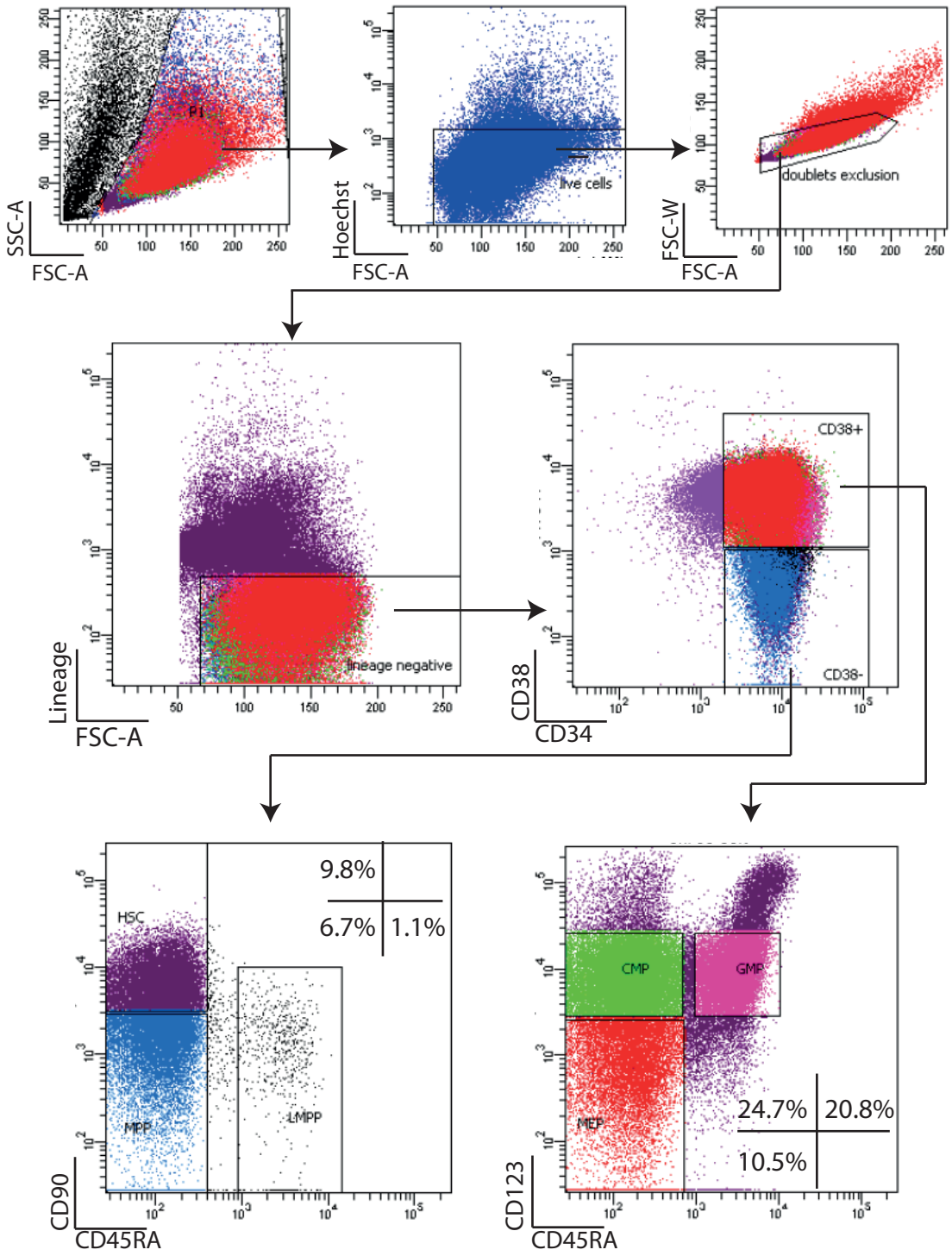


Fig 4.2 An example gating strategy for immunophenotyping and FACS sorting of stem and progenitor compartments. (Normal age matched bone marrow sample: OX709).
Gates shown are analysis gates. Percentages reflect average population percentage of CD34+ cells.

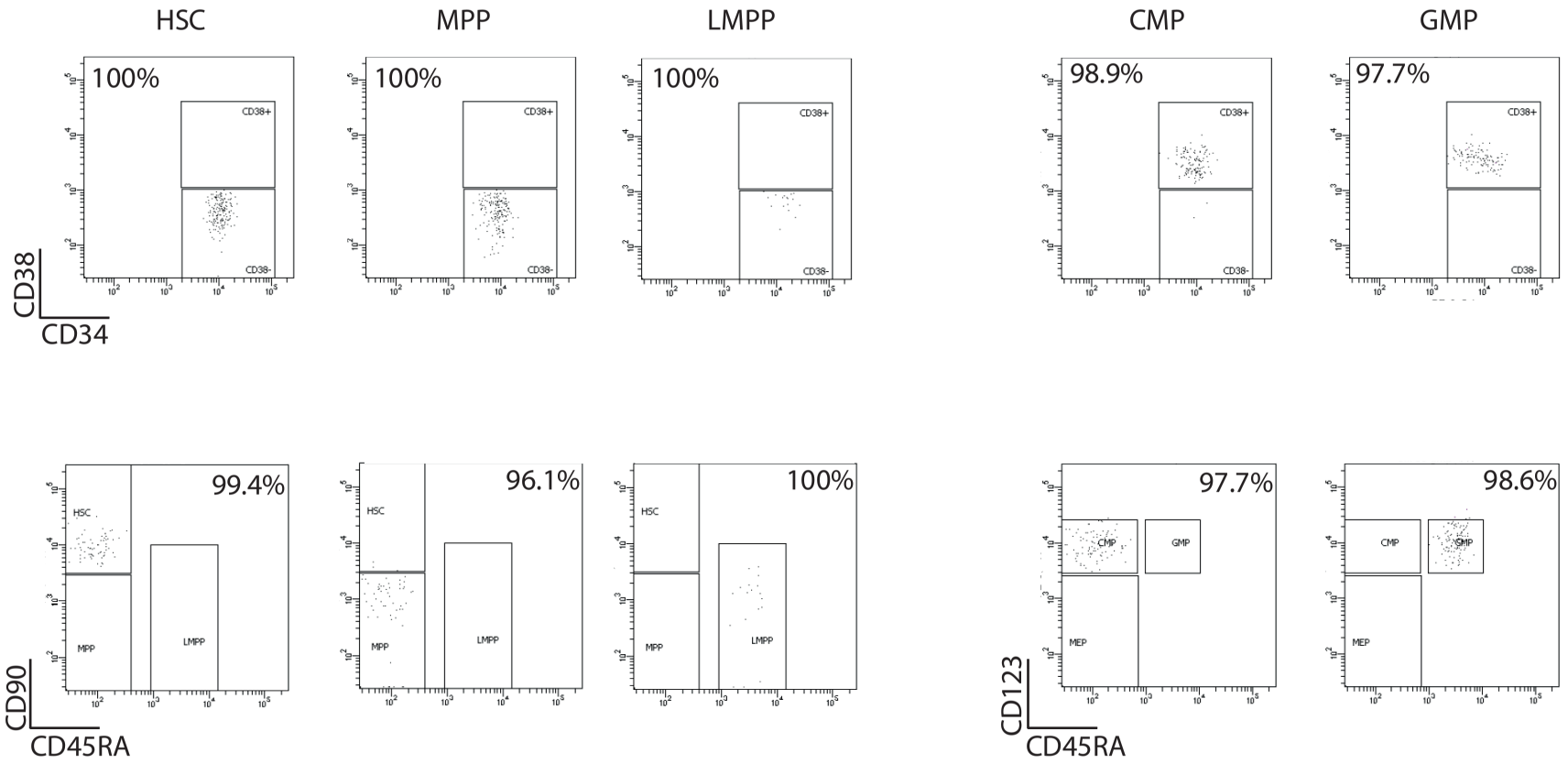


Fig 4.3 An example of FACS sort purity.
Post sort ~100 cells from each sorted population are reanalysed to determine purity of each sample. Shown plots are from sort of sample OX709 (fig4.3).

4.3 Increasing LMPP yield

The LMPP population makes up a very small proportion of normal BM CD34+ cells (figure 4.2). It became apparent, whilst beginning to sort normal BM samples, that obtaining enough normal LMPP cells for the microarray was going to be difficult. The amount of cells originally estimated that would be needed for the array also increased, as methods that extracted total RNA including small RNAs were less efficient than methods being previously used in the laboratory to extract RNA. A considerable time was spent to try and optimize the FACS sorting procedure to increase LMPP yield, details found in materials and methods. However the conclusion from this was the FACS procedures being used were optimal. As this didn't provide a method to improve LMPP yield, we instead looked at increasing the amount of cells available for sorting to increase LMPP numbers.

4.3A An assessment of the different commercially available kits to extract CD34+ cells

To increase CD34+ cell numbers for sorting an assessment of different kits available to extract CD34+ cells was carried out. The kit that was used routinely in the lab was the Miltenyi kit; however there were kits available from Stem Cell Technologies and Invitrogen to extract CD34+ cells, that were being used in the literature. All the kits use the same principle, label CD34+ cells with magnetic particles and use a magnetic field to extract the labelled cells. The magnetic particles bind to CD34+ cells via an antibody raised against CD34. The antibody clone between kits differs and so does the size and type of magnetic particle. The Invitrogen and Stem Cell Technologies kits work by applying a magnetic field to the outside of a sample tube, the CD34- cells are then

poured off or removed by pipetting. The magnetic field is then removed and cells of interest are resuspended in the sample tube. The Miltenyi kit works by putting cells through a column surrounded by a magnetic field, the CD34- cells are washed through the column. The magnetic field is then removed and the CD34+ cells are eluted from the column.

Several different BM samples were used to compare the three kits. MNCs were isolated from BM samples by density gradient separation (see materials and methods), and MNCs were divided equally between the three kits. Flow cytometry and manual cell counts were used to assess the efficacy of the kits. The first sample used was an AML sample: SW144. 6.3% of the starting sample was CD34+ cells (Figure 4.4A). All of the kits left CD34+ cells behind in the depleted fraction (Figure 4.4B). However the CD34+ cells left in the depleted fraction of the Invitrogen, and the Miltenyi kits were cells that expressed low amounts of CD34. The purity of the enriched CD34+ fractions differed between the kits. The Invitrogen kit gave the best purity 95.8% compared to 65.2% from the Miltenyi kit. There were very few cells in the enriched fraction of the Stem Cell Technologies kit and they did not appear to be CD34+ (Figure 4.4 C). The cell numbers and CD34+ % determined by flow cytometry, obtained from the different kits for sample SW144 can be found in table 4.5. The cell count and the purity of the enriched fraction were used to calculate the number of CD34+ cells extracted by each kit. The Miltenyi kit extracted the most CD34+ cells (10.1 $x10^5$) compared to $4.8x10^5$ from the Invitrogen kit, and the Stem Cell Technologies kit did not extract any CD34+ cells.

The next sample used to test the kits was COL096, an MDS sample. The starting amount of CD34+ cells in this sample was 3.1% (Figure 4.5A). Again all the kits left cells behind in the depleted fraction (Figure 4.5B). The purity of the enriched fraction was similar for both the Invitrogen kit 69% and the Miltenyi kit 73% (Figure 4.5C). Once again there were very few cells in the enriched fraction of the Stem Cell Technologies

kit. The cell numbers and CD34+ % determined by flow cytometry, obtained from the different kits for sample COL066 can be found in table 4.6. The Miltenyi kit extracted a much greater number of CD34+ cells ($0.9x10^5$ cells) than the Invitrogen kit ($0.01x10^5$ cells) for this sample.

The final sample the kits were used on was BRI435, an AML sample. The starting percentage of CD34+ cells was 1.9% (Figure 4.6A). As with the previous samples, all the kits left CD34+ cells behind in the depleted fraction (Figure 4.6B). Like sample SW144, the CD34+ cells left in the depleted fraction of the Invitrogen, and the Miltenyi kits were cells that expressed low amounts of CD34. In this sample the purity of the enriched fraction was higher for the Miltenyi kit (84%) than the Invitrogen kit (69%) (Figure 4.6). As previously seen the Stem Cell Technologies kit had very few cells in the enriched fraction. The cell numbers and CD34+ % determined by flow cytometry, obtained from the different kits for sample BRI435 can be found in table 4.7. Again the Miltenyi kit extracted more CD34+ cells ($1.5\ x10^5$ cells) than the Invitrogen kit ($0.3x10^5$ cells).

The Miltenyi kit did not extract all of CD34+ cells present within the MNCs; it extracted 33.7%, 34.6% and 57.7% of the starting CD34+ cells. However this was much more than the Invitrogen kit, which extracted 16%, 0.4% and 11.5% of the starting CD34+ cells. The Invitrogen kit extracted the highest percentage of starting cells from sample SWI144 and this was the sample the Miltenyi kit extracted the least cells from (see Tables 4.1-4.3), suggesting efficacy was sample independent. In the case of samples SW144 and BRI435 it is difficult to say whether the cells staining positive for CD34 within the depleted fractions really were CD34+ cells, as the signal was very low and could be flow cytometry background (see Figures 4.4 and 4.6). The Stem Cell Technologies kit had the poorest performance and did not appear to enrich for CD34+ cells. The Miltenyi kit outperformed the Invitrogen kit in all the samples tested here as well as in other samples tested by colleagues. Therefore to extract the greatest

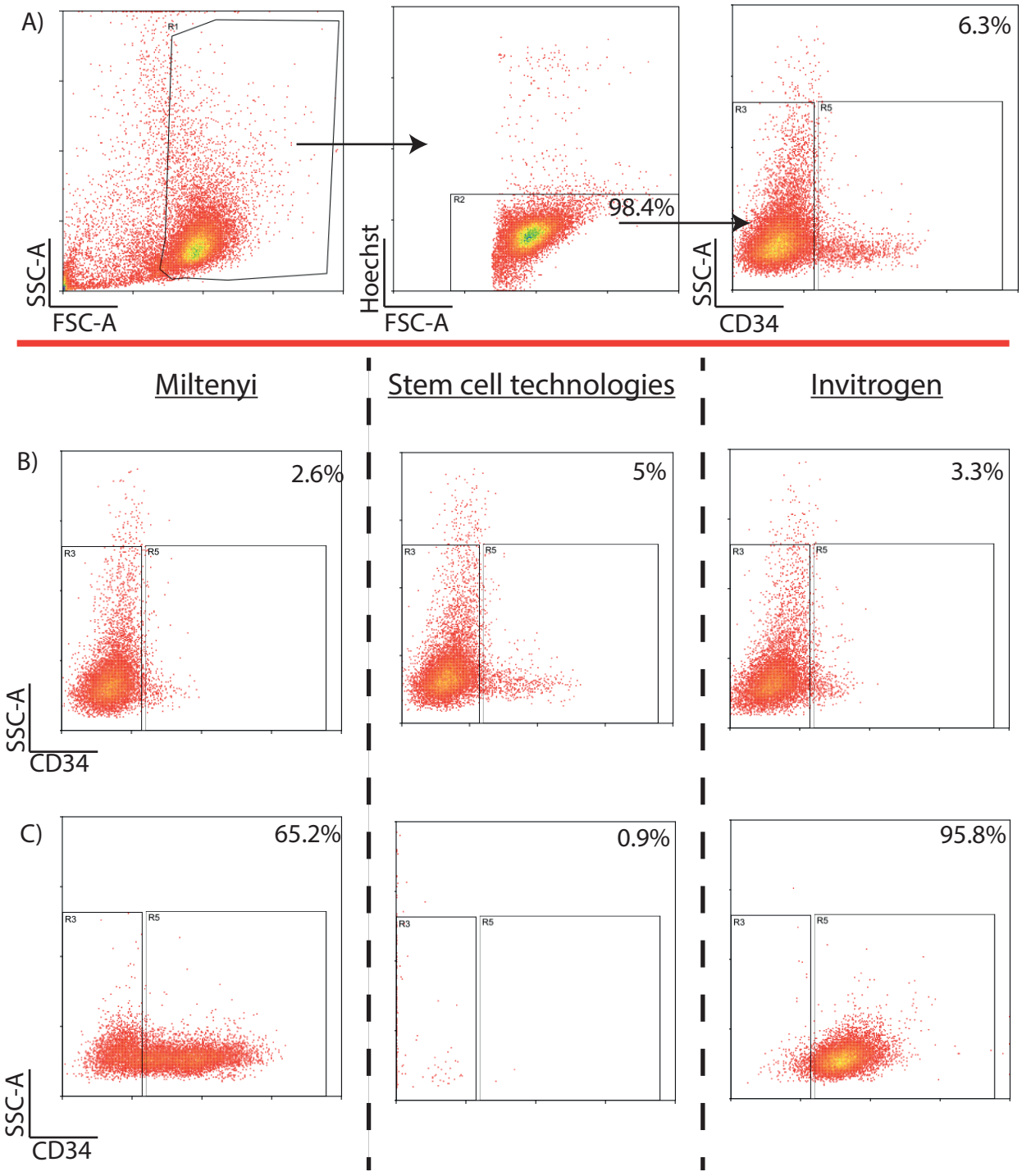


Fig 4.4 Comparing different CD34 enrichment kits (sample: SW144).
Three different kits for enriching CD34+ cells were tested to find the kit that obtained the best yield of CD34+ cells. The same sample was split into three equal parts, and then each kit was used according to manufacturer's instructions. A) The sample before separation, the last panel shows the starting percentage of CD34+ cells. B) Percentage of CD34+ cells remaining in the depleted fraction. C) percentage of CD34+ cells in the enriched fraction.

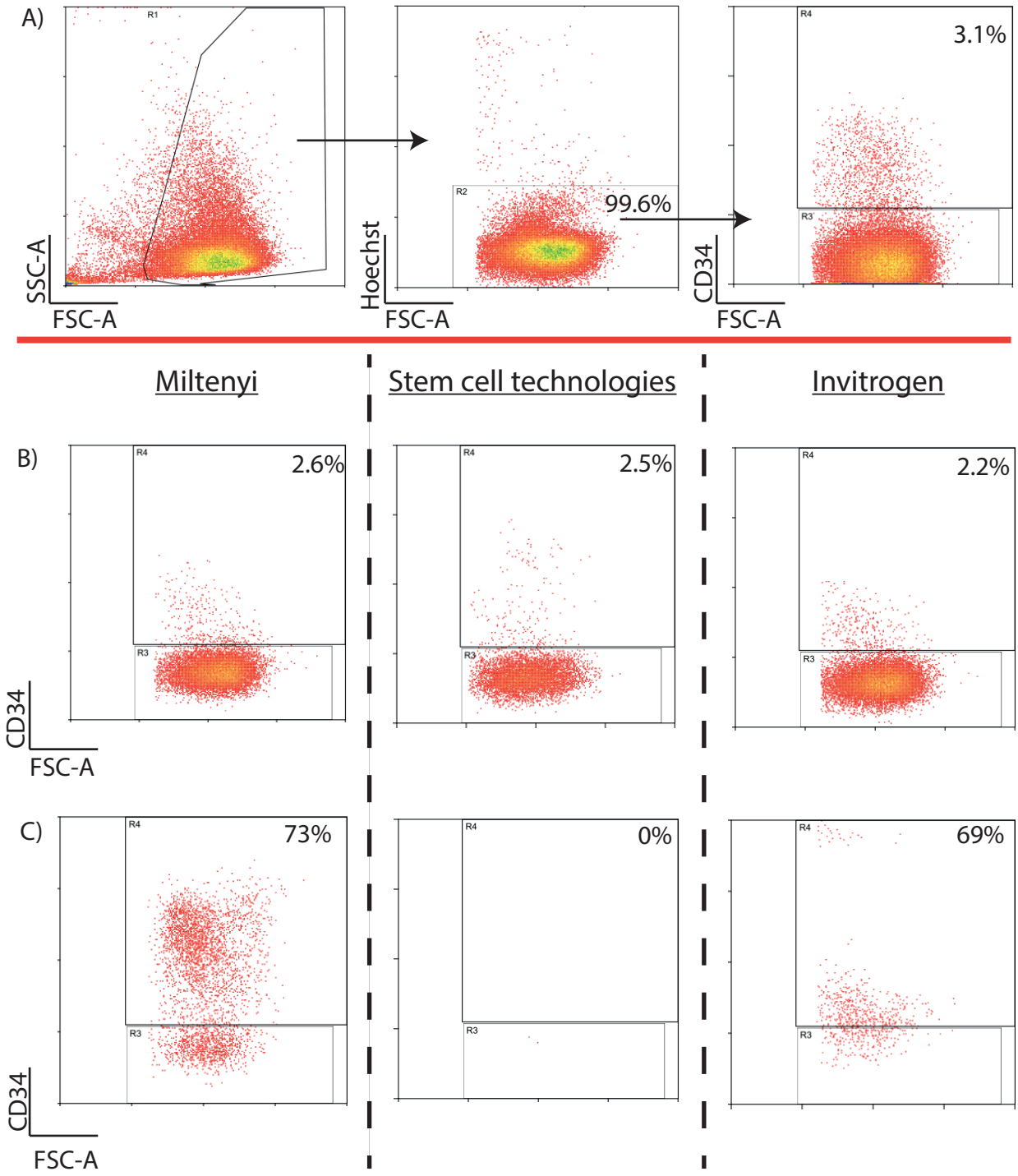


Fig 4.5 Comparing different CD34 enrichment kits (sample: COL096).
Three different kits for enriching CD34+ cells were tested to find the kit that obtained the best yield of CD34+ cells. The same sample was split into three equal parts, and then each kit was used according to manufacturer's instructions. A) The sample before separation, the last panel shows the starting percentage of CD34+ cells. B) Percentage of CD34+ cells remaining in the depleted fraction. C) percentage of CD34+ cells in the enriched fraction.

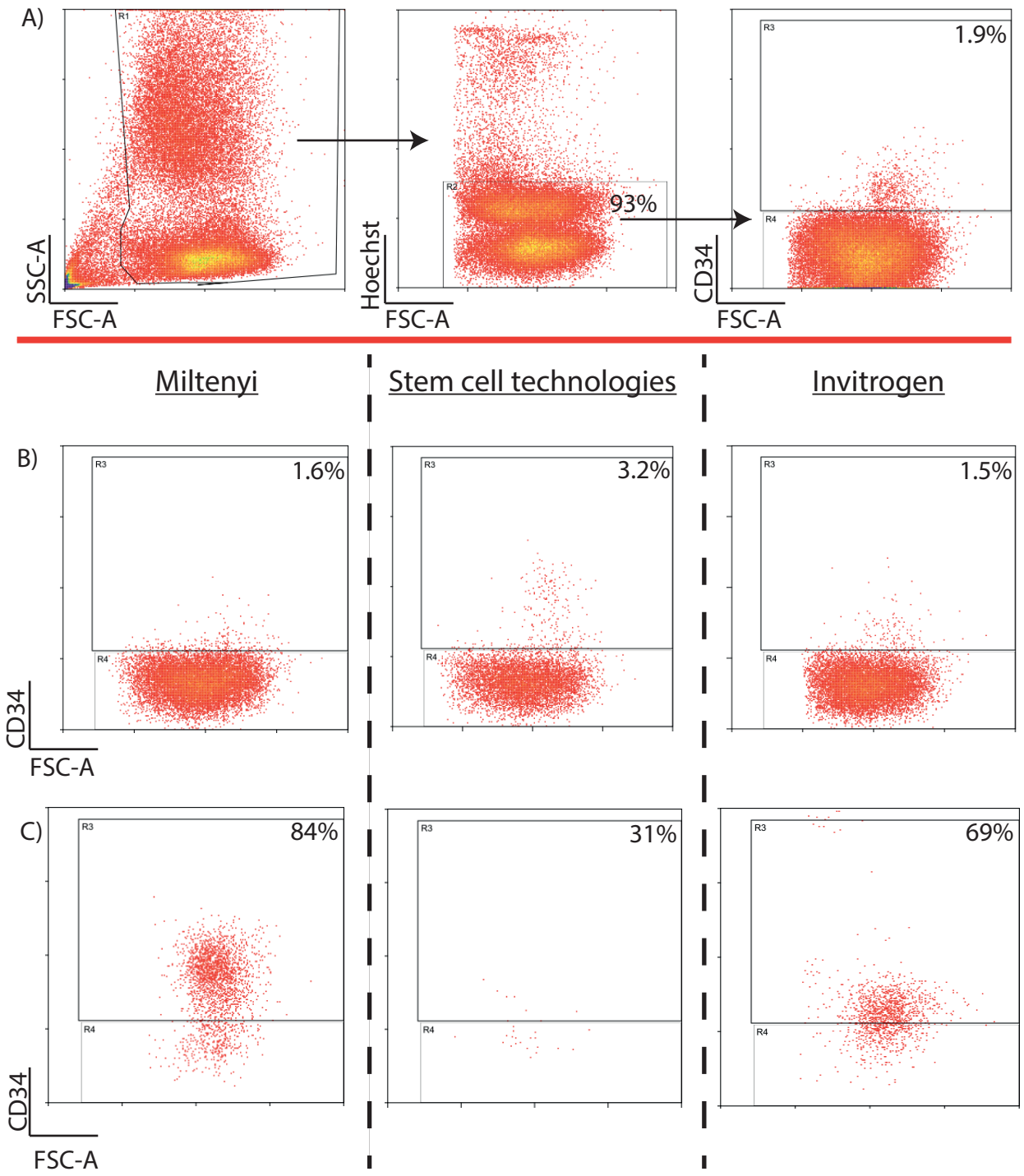


Fig 4.6 Comparing different CD34 enrichment kits (sample: BRI435).
Three different kits for enriching CD34+ cells were tested to find the kit that obtained the best yield of CD34+ cells. The same sample was split into three equal parts, and then each kit was used according to manufacturer's instructions. A) The sample before separation, the last panel shows the starting percentage of CD34+ cells. B) Percentage of CD34+ cells remaining in the depleted fraction. C) percentage of CD34+ cells in the enriched fraction.

Miltenyi	Before separation	Depleted fraction	Enriched fraction
cell count	47 x106	37.5 x106	15.5x105
flow cytometry CD34%	6.3%	2.6%	65.2%
Amount of CD34+ cells calculated from flow cytometry	[30 x105]	[9.8 x105]	[10.1 x105]
percentage of starting CD34+ cell		32.7%	33.7%
Stem cell technologies	Before separation	Depleted fraction	Enriched fraction
cell count	47 $x10^6$	36 $x10^6$	$<4x10^3$
flow cytometry CD34%	6.3%	5 %	0.9%
Amount of CD34+ cells calculated from flow cytometry	[30 $x10^5$]	[18$x10^5$]	[0]
percentage of starting CD34+ cells		60%	0%
Invitrogen	Before separation	Depleted fraction	Enriched fraction
cell count	47 $x10^6$	35 $x10^6$	5 $x10^5$
flow cytometry CD34%	6.3%	3.3%	95.8%
Amount of CD34+ cells calculated from flow cytometry	[30 $x10^5$]	[11.6 $x10^5$]	[4.8 $x10^5$]
percentage of starting CD34+ cells		38.7%	16%

Table 4.1 Comparing different CD34 enrichment kits (sample: SW144).
Cell counts and FACS percentages from CD34 kit experiment using sample SWI044 (Figure 4.13). Numbers in square brackets are theoretical CD34+ cell numbers calculated from CD34 flow cytometry %.

Miltenyi	Before separation	Depleted fraction	Enriched fraction
cell count	8.2 $x10^6$	7.4 $x10^6$	1.2$x10^5$
flow cytometry CD34%	3.1%	2.6%	73%
Amount of CD34+ cells calculated from flow cytometry	[2.6 $x10^5$]	[1.9 $x10^5$]	[0.9 $x10^5$]
percentage of starting CD34+ cells		73%	34.6%
Stem cell technologies	Before separation	Depleted fraction	Enriched fraction
cell count	8.2 $x10^6$	6.16 $x10^6$	0
flow cytometry CD34%	3.1%	2.5%	0%
Amount of CD34+ cells calculated from flow cytometry	[2.6 $x10^5$]	[1.5 $x10^5$]	[0]
percentage of starting CD34+ cells		57.7%	0%
Invitrogen	Before separation	Depleted fraction	Enriched fraction
cell count	8.2 $x10^6$	9.75 $x10^6$	2 $x10^3$
flow cytometry CD34%	31.%	2.2%	69%
Amount of CD34+ cells calculated from flow cytometry	[2.6 $x10^5$]	[2.1 $x10^5$]	[0.01 $x10^5$]
percentage of starting CD34+ cells		80.8%	0.4%

Table 4.2 Comparing different CD34 enrichment kits (sample: COL096).
Cell counts and flow cytometry percentages from CD34 kit experiment using sample COL096 (Figure 4.14). Numbers in square brackets are theoretical CD34+ cell numbers calculated from CD34 flow cytometry %.

Miltenyi	Before separation	Depleted fraction	Enriched fraction
cell count	13.6 x10^6	6.6 x10^6	1.8x10^5
flow cytometry CD34%	1.9%	1.6%	84%
Amount of CD34+ cells calculated from flow cytometry	[2.6 x10^5]	[1.1 x10^5]	[1.5 x10^5]
percentage of starting CD34+ cells		42.3%	57.7%
Stem cell technologies	**Before separation**	**Depleted fraction**	**Enriched fraction**
cell count	13.6 x10^6	11.7 x10^6	0
flow cytometry CD34%	1.9%	3.2%	0%
Amount of CD34+ cells calculated from flow cytometry	[2.6 x10^5]	[3.7 x10^5]	[0]
percentage of starting CD34+ cells		142%	0%
Invitrogen	**Before separation**	**Depleted fraction**	**Enriched fraction**
cell count	13.6 x10^6	10.3 x10^6	4 x10^4
flow cytometry CD34%	1.9%	1.5%	69%
Amount of CD34+ cells calculated from flow cytometry	[2.6 x10^5]	[1.5 x10^5]	[0.3 x10^5]
percentage of starting CD34+ cells		57.7%	11.5%

Table 4.3 Comparing different CD34 enrichment kits (sample: BRI435).
Cell counts and flow cytometry percentages from CD34 kit experiment using sample BRI435 (Figure 4.15). Numbers in square brackets are theoretical CD34+ cell numbers calculated from CD34 flow cytometry %.

number of CD34+ the miltenyi kit should be used. This allowed us to exclude other CD34+ extraction kits as a potential avenue to increase LMPP cell numbers. However it did highlight an area where large amounts of sample were being lost, therefore improvement of these kits could therefore help future studies looking at small cell populations.

4.3B Mobilized peripheral blood as alternate sample source

As large BM samples were rare, Granulocyte colony stimulating factor (GCSF) mobilized peripheral blood was considered as an alternative CD34+ sample source, in an attempt to increase cells available for sorting. Patients are administered with daily

doses of GCSF which induces the CD34+ cells from the bone marrow to mobilize into the blood. The blood is then collected from patients and undergoes leukaphereisis, the red cells are returned to the patient and the white cells are collected and frozen down. Large amounts of G-CSF mobilized peripheral blood samples were available to us from patients that had passed away.

To investigate if GCSF mob-PB could be a viable alternative to BM, I examined how these samples thawed and what their immunophenotypic profiles were. Details on the thawing and processing of the first Mob-PB sample can be seen in Figure 4.7. The viability of the sample post thaw was similar to previous bone marrows – 61% of cells were live cells ($4.8x10^9$) and 17% were CD34+ cells. Post thaw there was a problem with the sample clumping, which arises when there are large amounts of DNA present and occurs when a lot of cells are dying. Post thaw half the sample was used to isolate MNCs by density gradient separation. However cell clumping affected this process and post centrifugation only a very small layer of MNC cells was present. In the bottom layer where red cells and polynuclear cells are usually found a large clump of white cells could be seen. The MNC layer was taken, washed and analysed. Only 4.5 $x10^6$ CD34+ cells remained, indicating ~98.9% of the CD34 cells had been lost. The cells from the MNC layer were not enriched for CD34+ cells, as there was only a small amount of cells and 56% of these cells were CD34+. The cells from the MNC layer were put directly into overnight culture, to allow the cells to recover from the thawing and processing. The cell clump from the bottom of the tube was also analysed. It was vigorously resuspended and filtered to reduce the cell clumping. The cells were then washed and analysed. There were $5.4x10^7$ CD34+ cells present these made up 18% of the total cell fraction. Therefore these cells were enriched for CD34+ cells by magnetic bead enrichment, before being put in overnight culture.

The other half of the sample was not subjected to density gradient separation. This was due to concerns over the efficacy of this process in a sample that contained a

large amount of clumped cells. Instead, CD34+ separation was done on whole blood rather than MNCs. Although the process of leukaphereisis returns the red blood cells back to the patient, many red blood cells still remained in the white blood cell fraction. Red blood cells needed to be removed prior to CD34+ enrichment, so red cell lysis was carried out with ACK lysis buffer on the remaining sample. 67.5% of the total white cells were lost and 75.4% of the CD34+ cells were lost during red cell lysis. After red cell lysis cells were enriched for CD34+ cells and then the enriched cells were put in overnight culture.

There was a great deal of cell loss during overnight culture. For the CD34+ enriched cells, the cell loss was ~70% and for the cells that had not been enriched, the cell loss was 81.3%. After the overnight culture, the cells were stained for the stem and progenitor FACS sort and sorted. Figure 4.8 shows the immunophenotypic profile of the mob-PB after overnight culture. The immunophenotypic profile was similar to what was seen in BM, although there were fewer CD45RA+ cells than was routinely seen in BM. This was not ideal, as the LMPP population was part of the CD45RA+ fraction.

GSCF-mob-PB contains large amounts of granulocytes, which are very fragile and so may not survive the thaw process. When these cells die they release DNA which causes cells to clump together and toxins which can kill other cells. This is likely the reason for the cell clumping and cell lose during the thawing and processing of the sample.

To try and reduce cell clumping and cell lose the method to thaw and process was altered for the second GCSF-mob-PB, details of this can be seen in Figure 4.9. The sample was thawed by the same method as the first sample, but was then washed with a larger volume of buffer to more rapidly reduce the amount of DMSO which was present in the freezing buffer, as DMSO is toxic to cells. The cell viability for this sample was lower: 27% of the cells were alive post thaw as opposed to 61% in the first

sample. FACs buffer contains DNase to reduce cell clumping but this did not overcome the problem of cell clumping in the first sample. To reduce the problem of cell clumping in the second sample, cells were left at 37°C for 30 minutes with regular agitation, to allow the DNase time to work. After this there were still cell clumps but the size and number of clumps appeared to be greatly reduced. However there was cell loss at this stage – 34% of the total cells were lost – but the CD34+ cell loss was much lower, only 1% of the CD34+ cells were lost. Gradient density separation was used to isolate MNCs, as the cell clumping had been reduced. This worked efficiently and good separation was seen. A large MNC layer was present with very little evidence of clumps in the lower phase. The viability of the MNCs was high, 89% were alive. The number of CD34+ cells lost was still high, 61% were lost. MNCs were enriched for CD34+ cells and 77% of the CD34+ cells were extracted. The CD34+ enriched cells were put in overnight culture, 94% of the sample was lost during culture.

Post culture, the sample was stained for the stem and progenitor FACS sort and sorted. The immunophenotypic profile was similar to that seen in BM but there were fewer CD38- cells than what was normally seen in BM (Figure 4.10). This was a problem, as the LMPP population is part of the CD38- fraction.

Table 4.4 shows a summary of the percentage each population made-up of the CD34+ cells, as well as the cell numbers sorted from both GCSF-Mob-PB samples. Both the samples were immunophenotypically similar to BM, however differences in certain populations were observed. The population percentages also differed slightly between the two GSCF-Mob-PB samples. In both samples the LMPP population only made up a very small percentage of the CD34+ cells. As there were problems with cell death and cell clumping, the number of cells that were available for sorting was low. Very few LMPP cells were sorted from both samples, less than had been previously sorted from BM. The Mob-PB was therefore not a useful cell source alternative to BM.

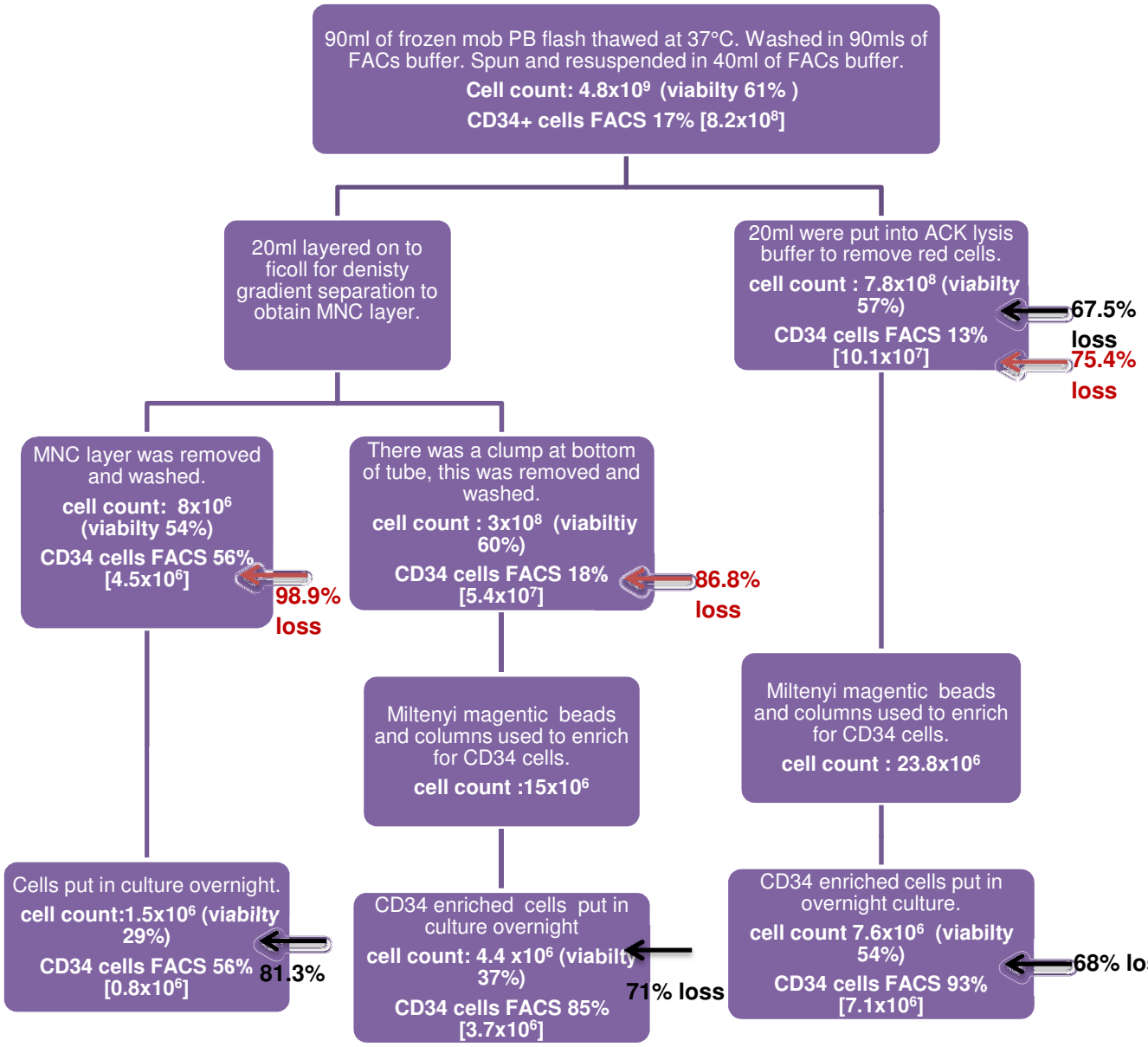


Fig 4.7 G-CSF mobilised peripheral blood as an alternative sample source to bone marrow. (G-CSF mobilized peripheral blood experiment 1.)
As large bone marrow samples were a limiting factor for obtaining LMPP samples, experiments were set up to test whether G-CSF mobilized peripheral blood could be a useful sample alternative to bone marrow.
A schematic showing how sample was thawed and processed, details of cell counts and FACS percentages of CD34+ cells are also shown. Cell counts were done using trypan blue so cell counts only reflect viable cells. Numbers in square brackets are theoretical CD34 cell numbers calculated from CD34 FACS. (MOB PB= mobilized peripheral blood) (MNC = Mononuclear cells)

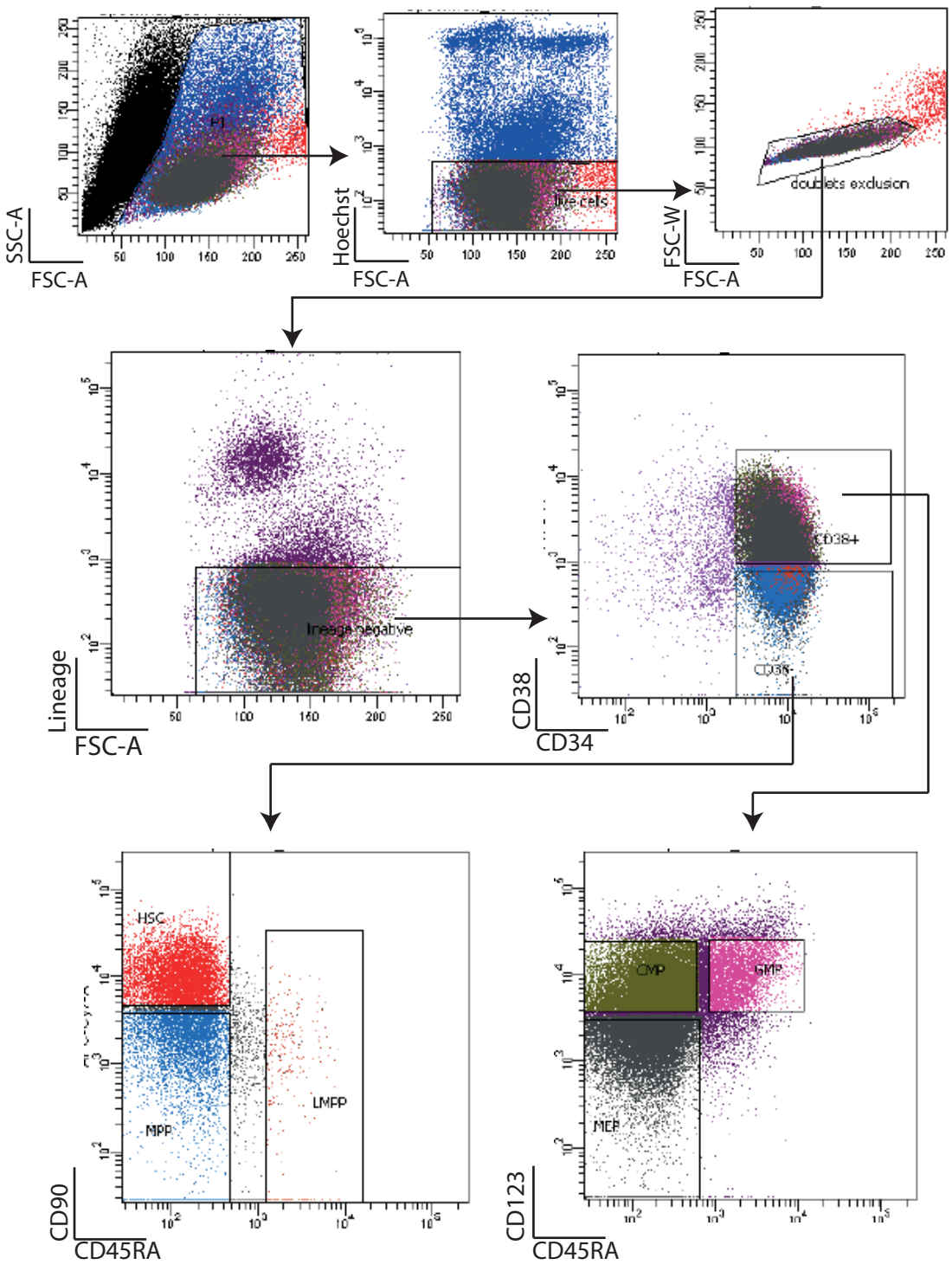


Fig 4.8 G-CSF mobilised peripheral blood as an alternative sample source to bone marrow. (G-CSF mobilized peripheral blood experiment 1.)
Immunophenotype of G-CSF mobilised peripheral blood from experiment 1 (Figure 4.16) after overnight culture.

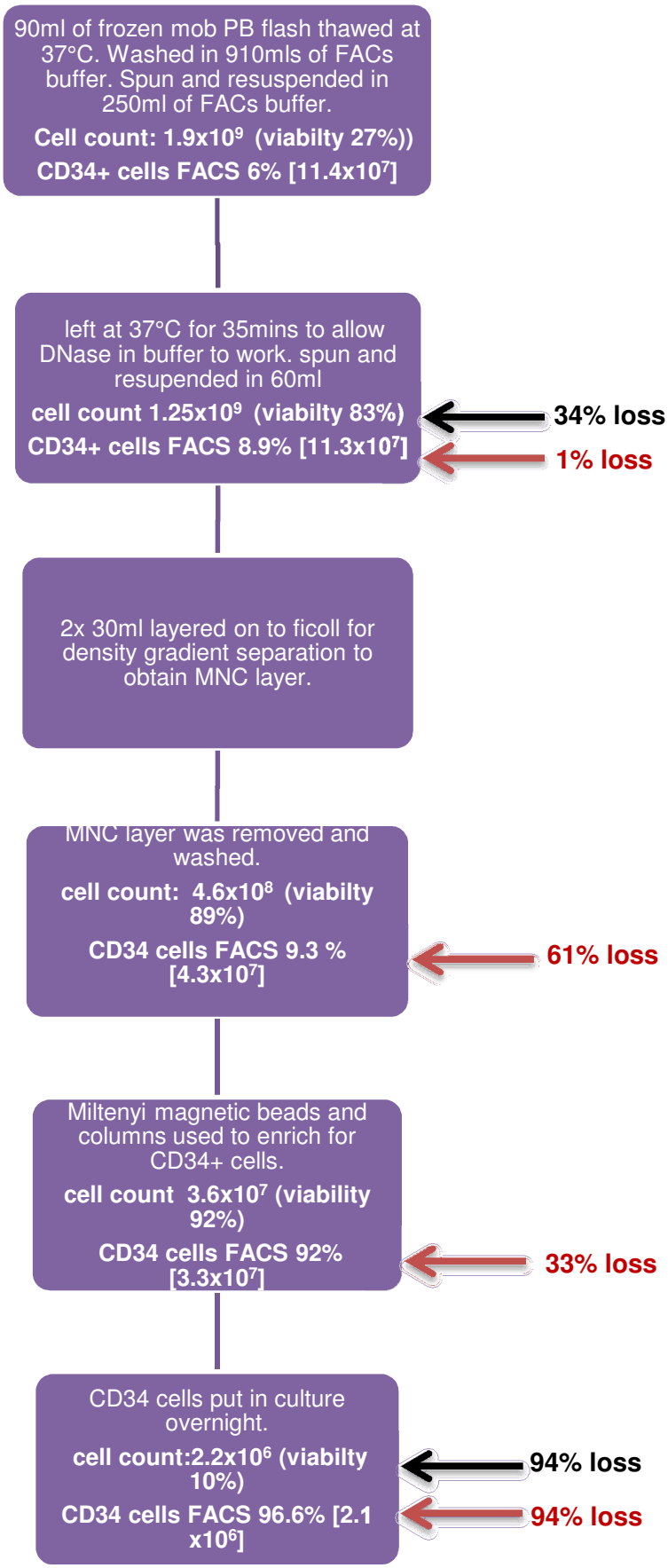


Fig 4.9 G-SCF mobilised peripheral blood as an alternative sample source to bone marrow. (G-CSF mobilized peripheral blood experiment 2.)
As large bone marrow samples were a limiting factor for obtaining LMPP samples, experiments were set up to test whether G-CSF mobilized peripheral blood could be a useful sample alternative to bone marrow. A schematic showing how sample was thawed and processed, details of cell counts and FACS percentages of CD34+ cells are also shown. Cell counts were done using trypan blue so cell counts only reflect viable cells. Numbers in square brackets are theoretical CD34 cell numbers calculated from CD34 FACS. (MOB PB= mobilized peripheral blood) (MNC = Mononuclear cells)

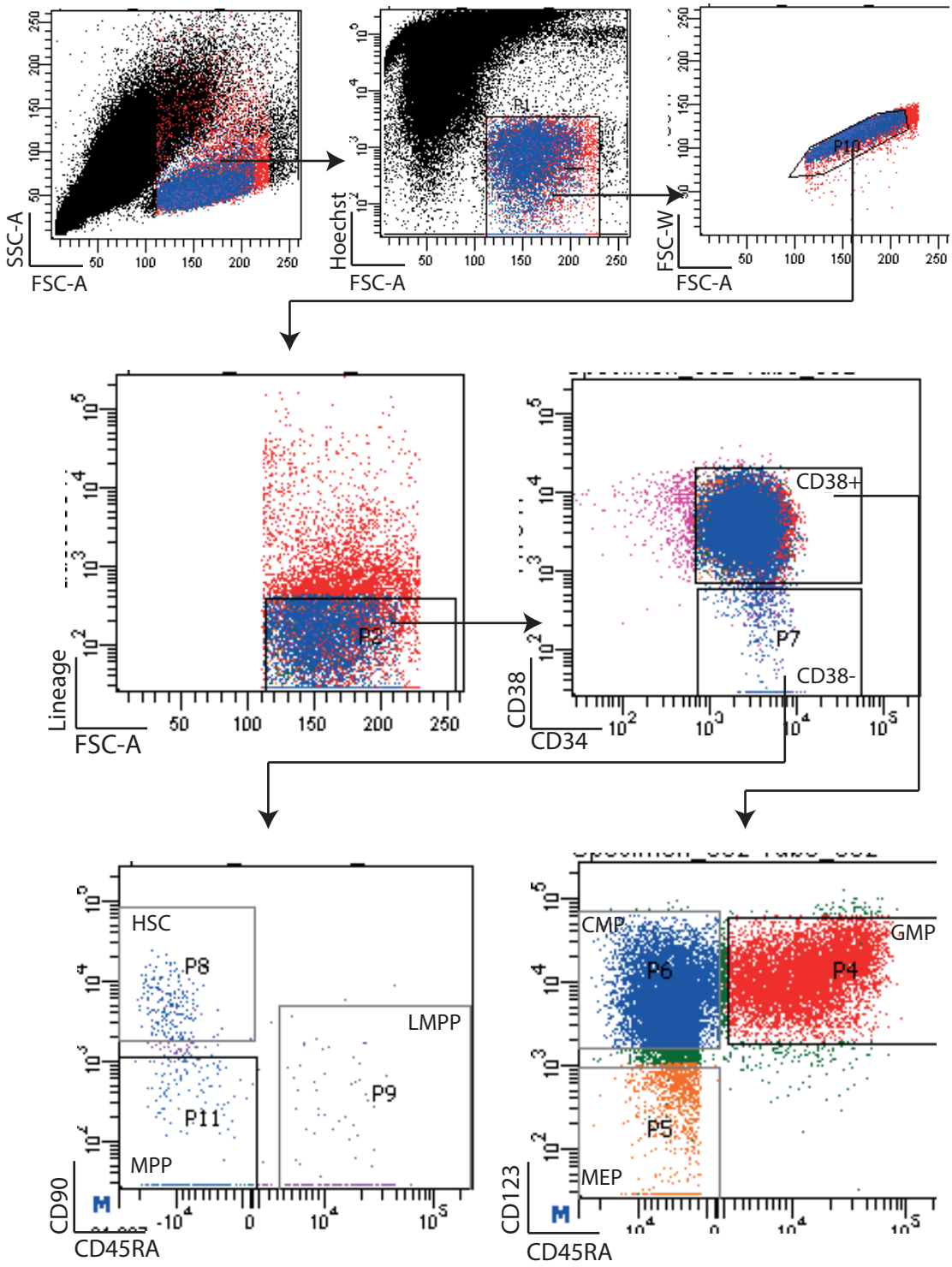


Fig 4.10 G-CSF mobilised peripheral blood as an alternative sample source to bone marrow. (G-CSF Mobilized peripheral blood experiment 2.)
Immunophenotype of mobilised peripheral blood from experiment 2 (Figure 4.18) after overnight culture.

sample	CD38+	CD38-	HSC	MPP	LMPP	CMP	GMP	MEP
	Population % of CD34+ cells							
ACK lysis (exp1)	76	24	13	8.7	0.29	29.5	5.9	30.7
Lymphoprep (exp1)	84	16	4.6	9.1	0.13	23.9	10.8	37.5
Clump (exp1)	73	27	15	9.2	0.38	27.5	5.7	29.7
exp 2	97.1	2.9	1	1	0.6	47	38.3	5.2
	Number of cells sorted							
ACK lysis (exp1)	1.1×10^6	ND	7.8×10^4	2.5×10^4	2000	ND	ND	ND
Lymphoprep (exp1)	9.1×10^4	ND	1328	1439	45	ND	ND	ND
Clump (exp1)	3.4×10^5	ND	2.4×10^5	7451	647	ND	ND	ND
exp 2	ND	ND	1300	1926	699	2.3×10^5	2.1×10^5	ND

Table 4.4 G-CSF mobilised peripheral blood as an alternative sample source to bone marrow.
Table showing information on immunophenotype and numbers of cells sorted from G-CSF mobilised peripheral blood experiments.

As it was not possible to increase the LMPP yield through optimising of FACS techniques nor was it possible to increase cell source for FACS through using different isolation kits or sample sources .The decision was made that it would not be technically possible to incorporate the LMPP population on the microarray.

4.4 Normal BM samples sorted for the microarray

Normal populations were sorted for the array as detailed in Figure 4.2. The immunophenotypic details each of the normal samples sorted for the microarray can be found in Table 4.5. The percentage each population made up of the CD34+ fraction is as follows; average HSC percentage of the CD34+ fraction 9.8%; MPP, 8.3%; LMPP 1.1%; CMP 24.7%; GMP 20.8%; and MEP 10.5. All of the sorted populations had high sort purities, all above 90%; the average purities for the populations were: HSC 99.2%, MPP 96.7%, CMP 98.1% and GMP 97.1% (Table 4.6)

4.5 AML BM samples sorted for the microarray

The immunophenotypes of the AML samples sorted for the array are shown in table 4.7. 52 AML samples in total were sorted for the microarray, 32 of which were of sufficient size and quality and to be used on the array. Sort purities of the AML populations used for the array are in table 4.8. These were high, with an average LMPP-like population purity of 98.4% and an average GMP-like population purity of 97.8%. The samples used for the array had a wide range of different cytogenetics and mutational statuses, as we did not wish to focus on a particular subtype of AML other than the GMP-like/LMPP-like expanded AML (Table 4.9).

GMP-like/LMPP-like expanded AML samples were sorted with the same FACS panel as normal BM samples. However the immunophenotypic profiles in AML samples were very different to what was seen in normal BM and were much more varied. Examples of the profiles seen in CD34+ GMP-like/LMPP-like expanded AMLs can be seen in Figure 4.11.

The CD38 expression within AML was highly variable and either the CD38+ or CD38- fraction could be expanded. In about 60% of samples both CD38 populations could be seen (Figure 4.11A ii); in around 30% of samples only a CD38+ fraction was present (Figure 4.11A iii); and in around 10% of samples only a CD38- fraction was present (Figure 4.11A iv).

The CD38- fraction can be further separated by CD90 and CD45RA into the HSC, MPP and LMPP populations (Figure 4.11B). The LMPP population in the AML samples was highly expanded compared to normal BM. In 80% of the samples there were no other populations present in the CD38- compartment (Figure 4.11B ii); in 18% of

samples there was also an MPP population present (see Figure 4.11B iii); and in only one sample were all of the CD38- populations still present (see Figure 4.11B iv).

Figure 4.11 C shows the CD38+ fraction further separated by CD123 and CD45RA into the GMP, CMP and MEP populations. The GMP compartment was also highly expanded in the AML samples compared to normal BM. Of these 76% had only, the GMP population remaining in the CD38+ fraction (Figure 4.11 C ii); in 17% of these samples, there was also a CMP population present (Figure 4.11 C iii); and in 7% of samples there was a small amount of MEP as well as the GMP and CMP populations (Figure 4.11 C iv). The different AML immunophenotypic profiles seen in the samples sorted for the microarray confirmed and expanded on what had previously been seen in the lab for GMP-like/LMPP-like expanded AMLs.

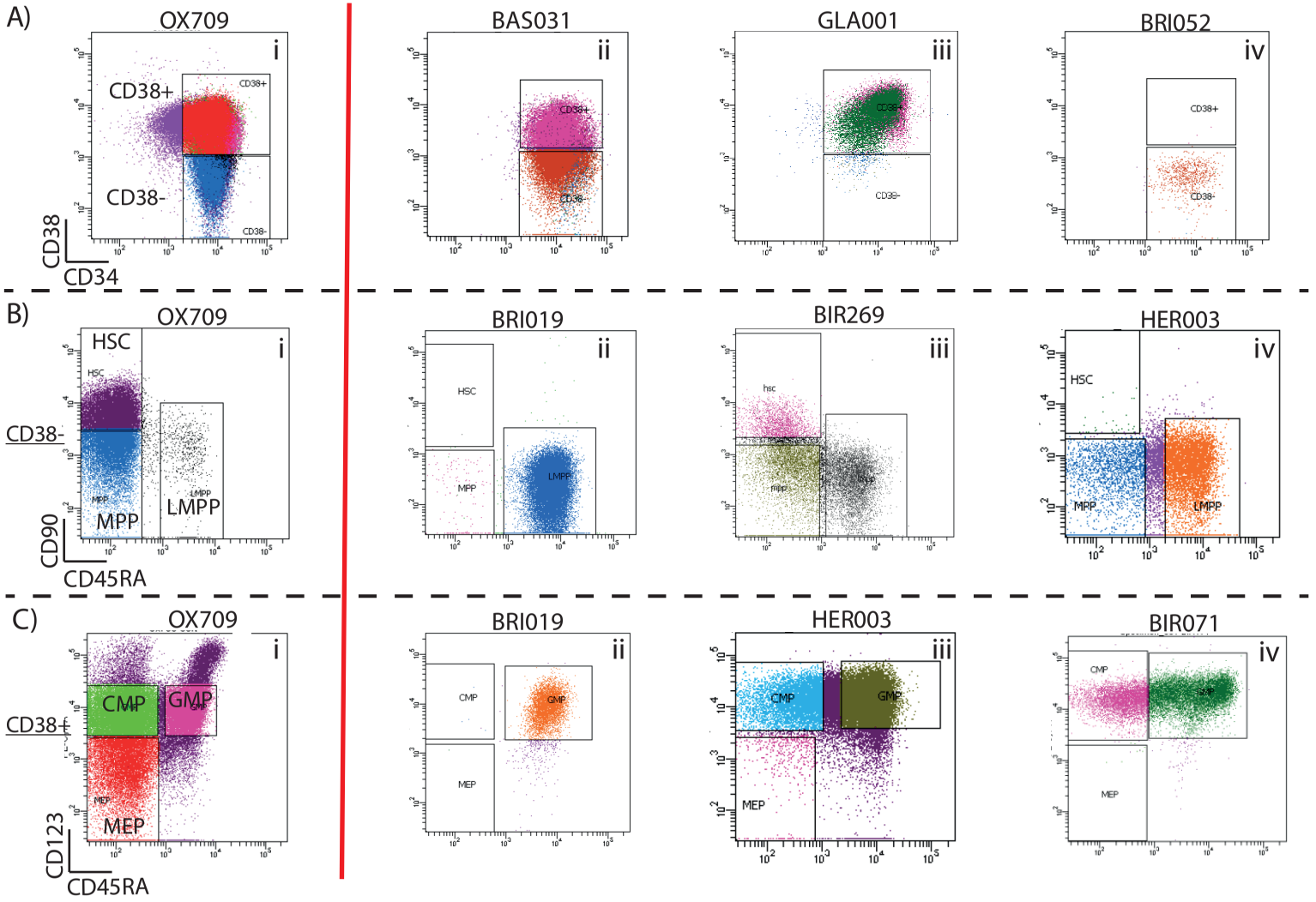


Fig 4.11 Examples of the different immunophenotypes seen within the GMP/LMPP like expanded AML.
A) CD38 expression in CD34 enriched samples, i) is an example of an age-matched normal control, ii-vi are AML samples representative of the different immunophenotypes seen. B) CD90/CD45RA expression profile in the CD38-fraction; highlights expanded LMPP populations in AML, i) is an example of an age-matched normal control, ii-vi are AML samples representative of the different immunophenotypes seen. C) CD123/CD45RA expression profile in the CD38+ fraction; highlights the expanded GMP populations in AML, i) is an example of an age-matched normal control, ii-vi are AML samples representative of the different immunophenotypes seen.

4.6 RNA extraction from purified populations

Once the samples had been sorted and high sort purity confirmed, cells were pelleted and stored at -80°C. RNA was extracted in batches from cell pellets, the RNA was then quantified using the Qubit fluorometer. The amount of RNA extracted from cells differed depending on cell type (Figure 4.12).The RNA content of the HSC and MPP populations was much lower than GMP and CMP populations. The HSC and MPP populations had on average 100ng/10^5 cells and the CMP and GMP populations had on average 300ng/10^5 cells. The range of RNA extracted per 10^5 cells was greater amongst the GMP/CMP cells than the HSC/MPP cells. The reason for more RNA being obtained from GMP/CMP cells, may be due to these cells being larger and more transcriptionally active than HSC/MPP cells.

LMPP-like AML cells had similar RNA amounts to the HSC/MPP cells. However there is no data available for normal LMPP cells to compare the LMPP-like AML cells to. The GMP-like AML cells had lower amounts of RNA than the normal GMP cells, the GMP-like AML cells had on average around 150ng/10^5 cells compared to 300ng/10^5 on average from normal GMP cells. The RNA levels in GMP-like AML cells were more similar to that seen in HSC/MPP cells than normal GMP cells.

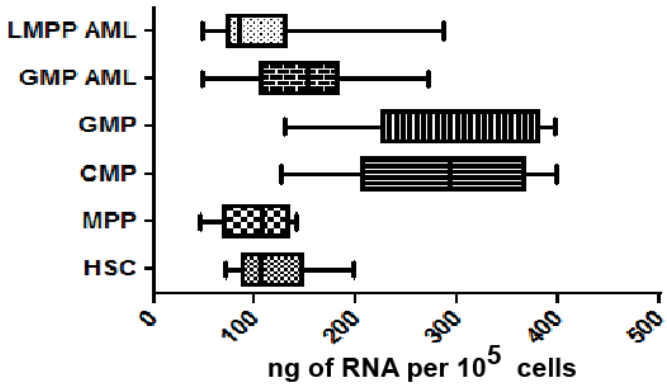


Fig 4.12 RNA quantity per cell type.
RNA was extracted from different cell numbers, amount per 10^5 cells was calculated to standardise comparison. Graph is a box and whisker plot showing the mean, 25th percentile, 75th percentile as well as the minimum and maximum values. Data is from 12 HSCs, 9 MPPs, 13 CMPs, 13 GMPs, 33 GMP-like AMLs, 26 LMPP-like AMLs

Sample ID	CD38- %	CD38+ %	HSC %	MPP %	LMPP %	CMP %	GMP %	MEP %	RNA used for array	RNA degraded	RNA not made	not enough RNA	RNA not used	RNA used for qPCR validation
OX496	16.1	83.9	8.6	5.0	2.1	37.4	26.0	16.8		HSC,MPP,CMP,GMP				
OX497	26.9	75.6	13.3	10.6	2.5	29.0	30.7	12.4	HSC,GMP	MPP,CMP				
OX533	26.8	73.2	3.3	22.1	0.4	27.3	11.2	6.8	HSC,MPP,CMP		GMP			
OX554	31.9	68.1	16.6	12.9	1.2	17.0	16.6	4.1		HSC	MPP,CMP,GMP			
OX530	11.7	88.3	6.3	4.6	0.3	26.0	26.8	16.7		HSC,GMP	CMP	MPP		
OX558	22.3	77.7	15.9	4.2	0.4	22.0	20.9	14.3		HSC	MPP,CMP		GMP	
OX548	23.4	76.5	8.6	7.5	2.3	19.9	21.8	9.6		CMP,GMP	HSC,MPP			
OX585	19.9	80.1	10.8	6.3	1.1	28.1	28.9	6.1		CMP,GMP	HSC,MPP			
OX586	20.7	79.3	10.3	8.4	0.70	26.6	20.1	4.5	GMP	MPP,CMP	HSC			GMP
OX598	10	90	5.7	2.7	0.9	17.2	16	12	CMP	HSC,MPP			GMP	
OX599	13.6	86.4	5.14	6.2	1.1	15	19	6		CMP,GMP	HSC,MPP			
Ox709	25.7	64.5	13.5	12.1	0.4	27.8	17	19.4	HSC,MPP,CMP,GMP					GMP
OX555	10.3	89.7	6.3	3.6	0.4	21.2	15.3	3.1	HSC		CMP,GMP	MPP		
OX564	24.5	75.5	12.8	9.8	1.3	31.2	21.4	15.2	HSC,GMP			MPP	CMP	
OX291	25.6	74.4	13.1	8.4	3.4	37.6	33.9	1.5	HSC,GMP	CMP		MPP		GMP
OX288	23.5	76.5	11	11.5	0.9	26.2	28.2	8.8	HSC,MPP,CMP,GMP					GMP
OX255	NA	NA	NA	NA	NA	NA	NA	NA	MPP					GMP
OX204	41	59	6.2	14.8	1.8	26	24.2	6.7	CMP					

Table 4.5 Immunophenotype data of normal age-matched control samples.
Details of immunophenotype for all samples used as well as what samples were used for.
%= what percentage each population is of the CD34+ cells.

Sample ID	HSC %	MPP %	LMPP %	CMP %	GMP %
OX497	98.5	97.9	98.5	98.4	97.7
OX533	98.0	97.0	98.1	98.4	99.4
OX586	100.0	98.8	100.0	100.0	99.8
OX598	100.0	96.0	100.0	99.5	96.3
OX709	99.4	96.1	100.0	96.5	96.3
OX555	100.0	96.0	100.0	99.0	ND
OX564	100.0	92.0	92.3	98.5	ND
OX291	NA	NA	NA	NA	NA
OX288	98	95	97	99	99
OX255	NA	NA	NA	NA	NA
OX204	NA	NA	NA	94	92
	99.2	96.1	98.2375	98.14444	97.21429

Table 4.6 Sort purity of age-matched normal control samples used in array and validation.

Sample	CD38- %	CD38+ %	HSC %	MPP %	LMPP %	GMP %	CMP %	MEP %	Used on array	GMP used in qPCR vaildation	RNA degraded	not enough cells sorted	not enough RNA extracted	array failed
AYL004b	0	100	0	0	0	75.2	6	3.2	GMP	x		LMPP		
BIR031	59.8	40.2	0	1	56.5	36.7	0.4	0	LMPP,GMP	x				
BIR073	1	99	0	0.2	0.7	34.6	64.2	0	GMP	x		LMPP		
BIR075	12.1	87.9	0	2.3	8.2	64.3	11.2	0.1	GMP				LMPP	
BIR091	41.1	58.9	0	0.1	41	57.8	0.4	0.1	LMPP,GMP	x				
BIR141	55.8	44.2	0	4.5	48	39.1	0.9	0.2	GMP	x				
BIR247	3.5	96.5	0	0.1	3.3	74.1	0	0	GMP			LMPP		
BIR257	36	64	0	1.8	32.8	52.2	5.3	0.3	LMPP,GMP	x				
BIR265	0.1	99.9	0	0	0.1	84.4	0	0	GMP	x		LMPP		
BIR269	22.2	77.8	2.4	7.2	11.1	60.7	3.8	0.4	GMP,LMPP					
BIR272	67.7	32.3	0	0.3	67.2	30.3	0	0	LMPP			GMP		
BIR277	20.4	79.6	0	1.6	18.3	63	12.1	0.2	GMP	x			LMPP	
BIR295	0.1	99.9	0	0	0.1	90.1	0	0	GMP	x		LMPP		
BIR305	5.7	94.3	0	0.3	5.3	76	0.3	1.2	GMP	x		LMPP		
BIR312	63.2	36.8	0	1.9	60.6	34.9	0.1	0	GMP	x			LMPP	
BOU002	62.8	37.2	0	0.3	60.6	31.9	0.1	0	LMPP,GMP	x				
BOU014	NG								LMPP,GMP	x				
BRI019	91	9	0	0.3	90.1	8.6	0	0	LMPP			GMP		
BRI023	11.4	88.6	0	0.5	10.6	75.4	0.7	0.3	LMPP					GMP
BRI024	99.7	0.3	1.9	41	55	0.1	0.1	0	LMPP			none from NG		
BRI080	21.6	78.4	0	1	20.4	76.4	0.1	0	GMP,LMPP	x				
GLA001	1.2	98.8	0	0.5	0.6	85.4	7.5	0	GMP	x		LMPP		
HER003	37	63	0	11.4	24.1	38.8	10	2	GMP	x				LMPP
LEE060	3.1	96.9	0	2.2	0.7	57.9	28.6	2.9	GMP	x		LMPP		
OX195	98.3	1.7	0.1	2.4	90.2	1.4	0.2	0.1	LMPP,GMP	x				
OX208	47.4	52.6	0	5.3	30.2	42.6	3.1	0.1	LMPP,GMP					
OX325	80.4	19.6	0.1	0.8	78.6	18.3	0.3	0	LMPP,GMP	x				

OX476	9.6	90.4	0	0.2	9.1	82.5	0.4	0.1	GMP	x			LMPP	
OX512	0.9	99.1	0	0.1	0.7	69.1	2.7	1.4	GMP	x		LMPP		
OX561	0.3	99.7	0	0	0.3	99.2	0	0	GMP	x		LMPP		
SMD0003	75.9	27.4	0.3	0.3	74.6	26.9	0.1	0.02	LMPP				GMP	
SW031	94.7	5.3	0	0.4	4.8	93.6	0.7	0	LMPP			none from NG		
BAS031	43	57	0	2.8	36.6	49.6	1.2	0.4			LMPP,GMP			
BIR071	57.4	42.6	0.4	14.8	40.1	29.3	13.0	0				LMPP,GMP		
BIR088	2	98	0	1.1	0.8	97.5	0	0			GMP	LMPP		
BIR093	30.5	69.5	0	1.2	27.1	65.4	1.5	0				GMP,LMPP		
BRI121	4.9	95.1	0	0.1	4.4	93.3	0.4	0.1			GMP	LMPP		
BIR125	33	67	0	0.9	32	64.6	0.4	0.1			LMPP,GMP			
BIR191	13.1	86.9	0	0.1	12.6	81.9	0.1	0.1				LMPP,GMP		
BIR193	6.4	93.6	0	0.1	6	90.7	0	0				LMPP,GMP		
BIR194	63	36.7	0	0.06	60.9	34.8	0.2	0				LMPP,GMP		
BIR200	16.1	83.9	0.1	1	13.5	77.8	0.4	0.5				LMPP,GMP		
BIR226	18.6	81.6	0	0.3	16.5	67.9	0.2	0.08				LMPP,GMP		
BIR314	27.2	72.8	0.4	2.7	23.6	70.2	0.4	0.1				LMPP,GMP		
BRI052	99.4	0.6	0	0.5	98.6	0.5	0	0				LMPP,GMP		
BRI089	53.1	47.1	0	4.3	47.1	43.2	1.6	0.2			GMP,LMPP			
BRI094	9.7	90.3	0.1	0.6	8.9	85.5	2.3	0.1				LMPP,GMP		
BRI104	56.8	43.2	0	0.1	55.4	43.1	0	0				SORT FAILED		
BRI105	81.5	18.5	0	6.2	70.3	11.3	3	0.6				GMP	LMPP	
BRI135	32.1	67.9	0	0.1	30.4	65.4	0	0				LMPP,GMP		
LEE040b	81.3	18.7	0.7	1.1	54.6	16.8	0.8	0			LMPP	GMP		
PLY004	78.8	21.2	0.9	1.3	75.2	21.1	0	0				GMP/LMPP		

Table 4.7 Immunophenotype data of AML samples.

Details of immunophenotype for all samples used. All samples above double line were used in array and samples highlighted in blue were samples used in qPCR validation. %= what percentage each population is of CD34+ cells.

Sample	LMPP %	GMP %
AYL004b	NS	100
BIR031	100	99
BIR073	96.9	97.1
BIR075	100	98.5
BIR091	100	96.1
BIR141	97.5	94
BIR247	97.6	96.3
BIR257	99.6	98
BIR265	NS	100
BIR269	100	100
BIR272	95	100
BIR277	98.9	95.4
BIR295	NS	98.6
BIR305	100	95.9
BIR312	100	97.4
BOU002	97.4	96.5
BOU014	NG	
BRI019	100	100
BRI023	100	98.4
BRI024	NA	NA
BRI080	99.4	97.8
GLA001	92.9	100
HER003	100	100
LEE060	NS	98.4
OX195	NS	98.2
OX208	100	99.7
OX325	100	92.1
OX476	100	100
OX512	86.4	100
OX561	NS	100
SMD0003	100	100
SW031	99.6	87

Table 4.8 Sort purity of AML samples used in array and validation.

Sample	Study	AGE	Diagnosis	cytogenetics	simplified cytogenetics	FAB	Blast%	FLT3	NPM1	CEBPα	CKIT
AYL004b	MDSBIO	NA	AML	NA	NA	M4	NA	NA	NA	NA	NA
BIR031	other	62	AML	46 XY	normal	M1	66%	ITD mt	WT	ND	WT
BIR073	Other	67	AML	96,XXYY,+13,+13,+13,+13[4]/46,XY[6]	complex	M5a	73%	WT	WT	NA	WT
BIR075	Other	71	AML	45,XX,t(3;6)(q26;q25),-7[10] MECOM (EVI1) gene rearrangement positive	EVI1 gene rearranged	M4	65%	WT	WT	NA	WT
BIR091	Val aza	67	AML	47,XX,+11[8]/46,XX[22]	complex	M5b	25%	WT	WT	WT	ND
BIR093	Val aza	69	AML	46 XY	Normal	M4	40%	WT	WT	WT	ND
BIR141	Val aza	50	Secondary AML	92,XXYY[31]/46,XY[29]	complex	NA	ND	WT	WT	ND	ND
BIR257	Other	1	AML	46,XY,inv(16)(p13q22)[2]/46,XY,del(7)(q31q36),inv(16)(p13q22)[16]/46,XY[2]	MYH11/CBFβ	NA		WT	WT	ND	NA
BIR265	Other	56	AML	46,XX,t(6;19)(q25;q13)?c[30]	t(6;19)	NA	NA	WT	ND	WT	NA
BIR269	Other	78	AML	46,XY,inv(3)(q21q26),add(12)(q15)[10]/46,XY,inv(3)(q21q26),der(12)idic(12)(p11)add(12)(q15)[5]/46,XY[4]	complex	NA	NA	WT	WT	WT	NA
BIR272	Other	78	AML	45,X,-Y,t(15;17)(q22;q21)[10]/46,XY[10]	PML-RARA	NA	NA	WT	ND	ND	NA
BIR277	Other	60	AML	46,XX[20]	normal	NA	NA	WT	ND	ND	NA
BIR295	Val aza	27	AML	NA	NA	NA	NA	NA	NA	NA	NA
BIR305	Other	33	AML	46,XY,t(8;21)(q22;q22)[10]	AML-ETO	NA	NA	WT	WT	WT	NA
BIR312	Other	27	AML	46,XX,t(15;17)(q22;q21)[7]/46,XX[3]	PML-RARA	NA	NA	39bp ITD MT	WT	ND	NA
BOU002	MDSBIO	NA	Secondary AML	46 XY	Normal	NA	ND	WT	WT	NA	WT
BOU14	MDSBIO	85	Secondary AML	ND	ND	NA	65%	WT	WT	NA	WT
BRI019	MDSBIO	71	AML	46,XX,inv(3)(q21q26),del(5)(q3q3),del(7)(q2q3),ins(12;?)(q13;?)[10]	complex	M4	90% (63%)	WT	WT	NA	ND
BRI023	MDSBIO	64	AML	47,XY,+8[3]/48,XY,+11,+13[2]/46,XY[5]	trisomy 8	M0/M1	95%	ND	WT	NA	ND

BRI024	NA	79	AML transformed from MDS	46,XY,del(5)(q13q33)[9]/46,XY[1]	del 5q	M0	90%	ND	ND	NA	ND
BRI080	AML 17	22	AML	NA	NA	NA	NA	NA	NA	NA	NA
GLA001	Other	54	AML	46,XX,del(12)(p11.2)[20]	del 12	NA	NA	NA	NA	NA	NA
HER003	MDSBIO	75	AML	54, XX,+1,+4,-5,+6,+8,+10,+11,+13,+14,+16,add (16)(q2), add(17) (p1), +18,i(22)(q10), +mar[cp14]	complex	M2	77%	FLT D835	WT	NA	WT
LEE060	MDSBIO		AML	NA	NA	NA	NA	NA	NA	NA	NA
OX195	MDSBIO	38	AML	46, XX	normal	M1	90%	ITD mt	MT	NA	WT
OX208	MDSBIO	75	AML	47,XY,+14[3]/46,XY[7]	trisomy 14	M0	55%	ND	WT	NA	WT
OX325	MDSBIO	41	AML	46, XY, inv (16) (p13q22)[6]/46, XY[14]	MYH11/CBFβ	M4	79%	WT	WT	NA	ND
OX476	MDSBIO		AML	45, X, -Y, t(8;21)(q22,q22) [10]	AML-ETO	M2	71%	WT	WT	NA	WT
OX512	MDSBIO	75	AML	-y	del Y	NA	NA	NA	NA	NA	NA
OX561	MDSBIO		AML	45,XX,-7,INC(6)	monosomy 7	NA	NA	NA	NA	NA	NA
SMD003	MDSBIO	47	AML	46 XX	normal	ND	27%	WT	WT	NA	WT
SW031	MDSBIO	NA	AML	46XYt(3;3)(p21q23)del(7)(q11q32) [28]/46XY[2]	EVI1 gene rearranged + monosomy 7	M1	66%	NA	NA	NA	NA

Table 4.9 Details of cytogenetics and mutational status of AML samples.
(Where available for all samples used in array and validation).

4.7 Discussion II

The aim of this section of work was to sort and prepare all samples needed for the microarray analysis. GMP-like and LMPP-like AML populations and normal HSC, MPP, CMP and GMP populations were sorted to high purity and the RNA was extracted for the microarray.

The original aim was also to include the normal LMPP population on the microarray; however it became apparent that sorting enough LMPP cells was going to prove difficult. So a large amount of time was spent trying to optimize FACS procedures and looking at ways to increase the number of cells available for sorting, however this did not generate a method to increase LMPP yield.. The decision was therefore made to exclude the normal LMPP from the microarray analysis, as due to sample availability and sorting limitations. I would not be able to sort the required amount of these cells.

Whilst sorting samples for the microarray, I was able to make observations on the immunophenotypic profiles of AML cells. There was still a large amount of heterogeneity within GMP-like/LMPP-like AMLs. With only 60% of samples having both a GMP-like and LMPP-like population present. Previous work in our lab had demonstrated that both populations were LSCs, but also that there was a hierarchy present where LMPP-like cells gave rise to GMP-like cells [15]. However in 30% of samples only a GMP-like population was present. This may be because the LMPP-like population was too small to have been detected; or that LMPP-like population gave rise to the GMP-like cells but these cells have since been outcompeted; or that there was never any LMPP-like AML cells present in these patients. This brings into question the origin of the GMP-like cells in these patients; are they the cells of origin or have they arisen from an alternative leuekmic population? here was never any LMPP-like AML cells it may have been that the GMP-like AML cells arose from a different cell type or that they are the cell of origin in that patient.

It is more often observed that GMP-like cells are the only cells present in the CD38+ fraction and that LMPP-like cells are the only cells present in the CD38- fraction. But in samples where other populations remain, it raises the question what is the nature of these cells; are they remaining normal cells or are they additional leuekmic populations. To determine in these cells are leukemic, chromosomal abnormalities or gene mutations associated with AML could be looked for in these remaining populations.

Very rarely is a notable amount of HSCs observed. Making it unlikely that a leukemic HSC population exists, however this possibility cannot be excluded. As leuekmic HSC levels may be too low to be detected by FACs. It also does not exclude the possibility that the first pre-leukemic mutations arose in these cells.

Chapter 5: Results III Microarray

The aim of the microarray was to produce data on the microRNAomes of highly purified leukemic cells and normal stem and progenitor cells in BM. We hypothesised that the microRNAome in leukemic cells would be disturbed as compared to the microRNAome in normal cells. To address this, several different bioinformatic analyses were used to investigate the similarities and differences seen at a microRNA level between the different cell populations.

5.1 Microarray supports and probes

Agilent microarray supports were used (Figure 5.1a). The support is a 25mm x 75mm glass slide and holds eight microarrays in two rows of four. Thousands of spots were printed on each microarray, with each spot containing multiple identical probes. RNA is labelled with a fluorescent Cyanine-3 dye conjugated to a cytosine; RNA is then allowed to hybridize to probes on the array surface (Figure 5.1b). Post hybridization, the arrays are scanned and the level of Cyanine-3 fluorescence detected at each spot is then used to calculate how much of each microRNA was present in the sample.

Each probe is printed across several different spots to create replicates, so if a problem arises with a particular spot, the spot can be excluded without losing information on that particular microRNA. There are either 16 replicate spots of one probe, eight replicate spots of two different probes recognising the same microRNA, or four replicate spots of four different probes recognising the same microRNA.

The design of the array probes allows for efficient and specific binding of microRNAs. Each probe contains a hybridization sequence that binds to specific target microRNA.

A guanine is included as part of the hybridization sequence to stabilize the binding of cytosine-labelled microRNAs. A hairpin is also incorporated as part of the probe to allow for size specificity, as the hairpin prevents large non-specific RNAs from stably binding probes (Figure 5.1b).

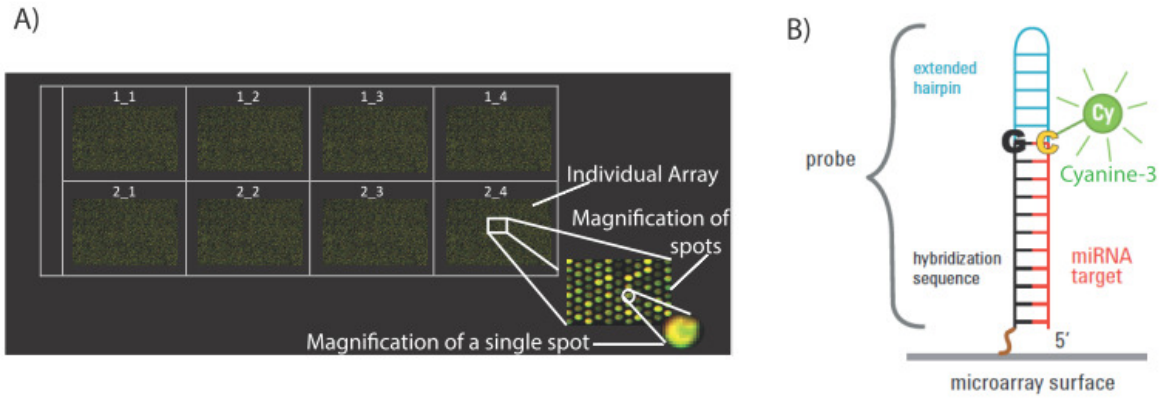


Fig 5.1 Microarray and probe schematic.
A) Representation of one microarray support with magnification of a microarray spots. Each microarray support holds 8 microarrays. B) Illustration of a microRNA binding an array probe. Multiple identical probes will be printed on one array spot.

5.2 Microarray process

5.2.1 Production of array data

A flow chart outlining the different stages involved in producing array data is shown in figure 5.2. The first stage involved purifying the different cell populations and extracting RNA, details of which can be found in Chapter 3.

The next stage was to evaluate RNA. RNA was first quantified using the Qubit fluorometer. If a suitable quantity of RNA had been extracted, RNA integrity was analysed using the Agilent bioanalyser. Further information on the analysis of RNA

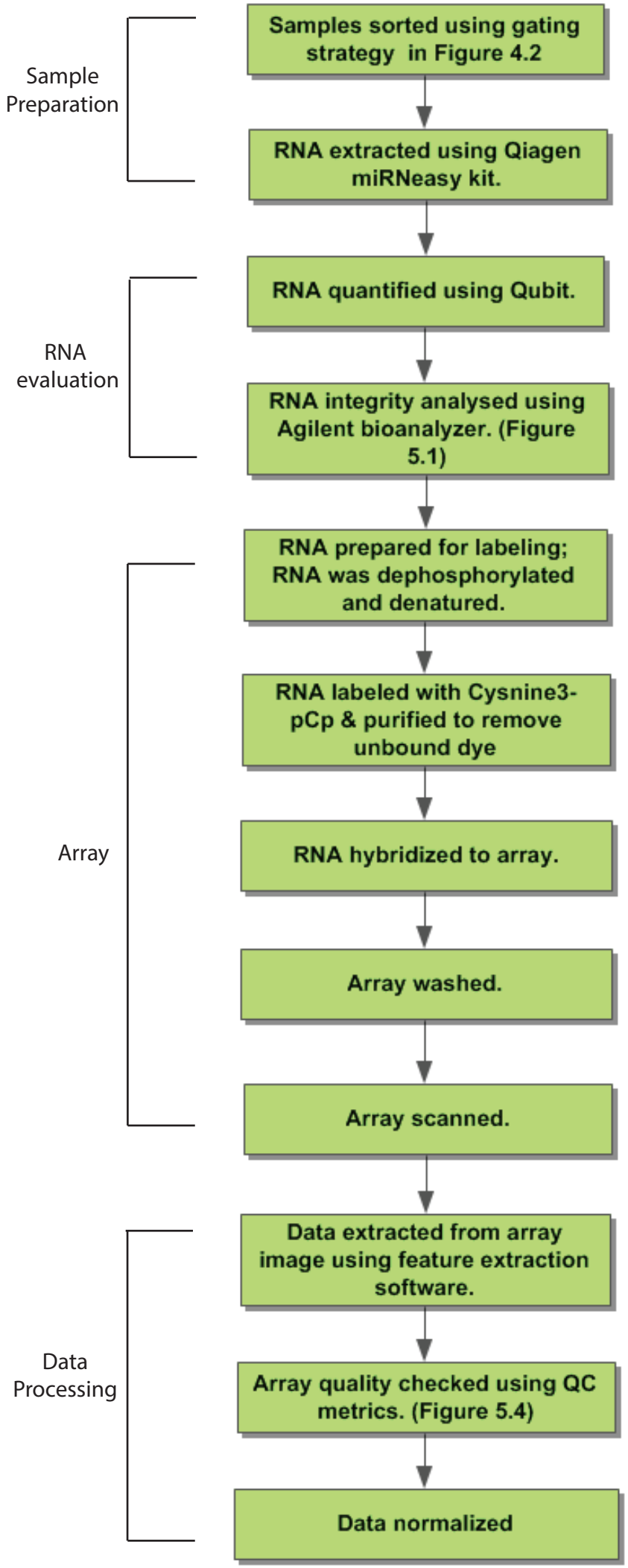


Fig 5.2 Flow Chart Outlining the key experimental steps invovled in producing array data

integrity can be found in Section 5.2.2. Details of the quantity and integrity of RNA samples run on microarrays are in Table 5.3.

RNA of suitable quantity and Integrity was then run on a microarray. Firstly RNA was dephosphorylated and denatured to ensure efficient unbiased labelling. The RNA was then labelled by T4 RNA ligase with Cyanine-3-pcp and then purified to remove unbound dye. Purified RNA was loaded onto array slides and baked overnight to allow the RNA to hybridize to array probes. Once hybridization was complete, the array slides were washed to remove unbound RNA and any residual dye. The arrays were then scanned using the Agilent microarray scanner, which produced a high resolution TIFF image for each array support.

To process the data produced, Feature extraction (FE) software was used to extract the data from the TIFF image. The FE software also produced a QC report for each array. This was examined to determine the quality of the array data. If the data from an array was deemed poor quality, the array was excluded from any further analysis. Further information on the QC report can be found in Section 5.4. The final step of data processing was data normalization, and the bioinformatician in our lab conducted this following the methology [122].

5.2.2 RNA integrity

It was important degraded RNA was not used on arrays, as this would lead to erroneous data. To prevent this RNA integrity was determined using the Agilent bioanalyser, which runs the equivalent of an agarose gel on a chip using microfluidics. This produces an electrogram (Figure 5.3a) as well as a corresponding gel-like image (Figure 5.3b). The Agilent bioanalyser also calculates a RNA integrity number (RIN)

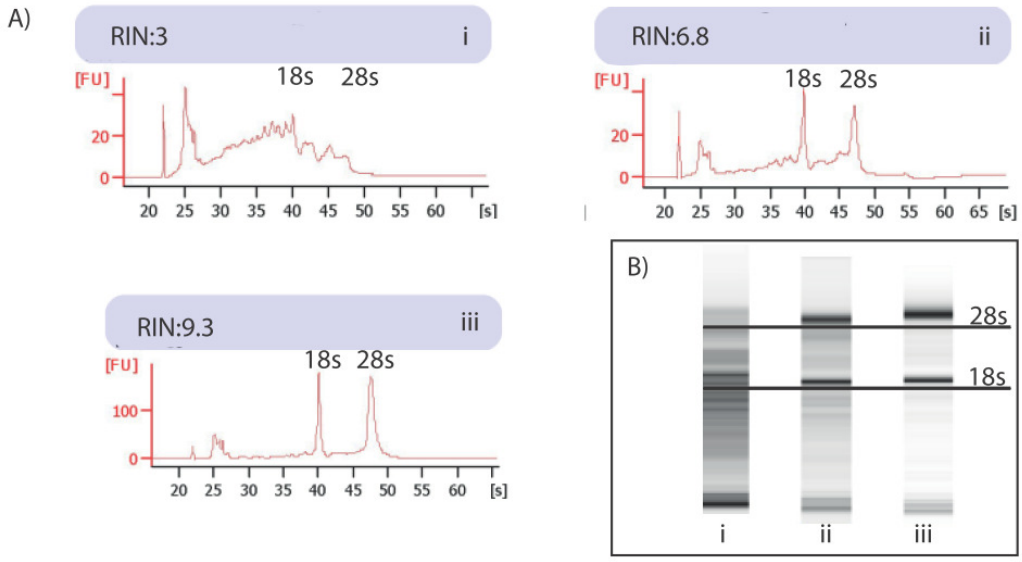


Fig 5.3 RNA integrity analysis.
After extraction RNA integrity was examined using the Agilent bioanalyser. A and B) data output after running samples on bioanalyser. A) Electrograms. B) The corresponding gel-like images. i) An example of a degraded RNA sample; that would not be used on the array. ii) An example of an RNA sample at the lowest integrity level that would be considered for use on the microarray. iii) An example of a RNA sample at a high level of integrity. RIN scores range from 1-10. (RIN: RNA intergrity number)

score, which is a measure of RNA integrity. The RIN score is calculated based on the ratio between the area underneath the electrogram line at the 18s peak and 28s peak. The area under the 28s peak should be roughly twice that of the area underneath the 18s peak if the RNA is intact. The area under the electrogram line not under the peaks is also used as part of the RIN calculation, as this is representative of degradation products. The RIN score ranges from 1 to 10, with 10 indicating RNA with the highest integrity and 1 indicating RNA with the lowest integrity.

An example of bioanalyser outputs for a range of different RNA integrities can be seen in Figure 5.3. The RNA sample (i) in Figure 5.3 is an example of degraded RNA: the 28s and 18s peaks can no longer be discriminated on the electrogram, and the gel-like image is smeared, with degradation bands present. The RIN value for this sample was 3. The RNA sample (iii) in Figure 5.3 is an example of a sample with high integrity: very clear 18s and 28s peaks can be seen on the electrogram; with twice the area

underneath the 28s band compared to the 18s band; and there are also no spikes on the electrogram representing degradation products. A small spike can be seen at 25 seconds which represents small RNA species. The gel-like image shows clear 28s and 18s bands without the presence of smearing. The RIN value for this sample was 9.3. The RNA sample (ii) in Figure 5.3 is a sample with lower integrity than (iii) but is not entirely degraded like (i); clear 28s and 18s bands can been seen on both the electrogram and the gel-like image and the area under the 28s peak is roughly twice the area under the 18s peak. The presence of a small amount of degradation can be seen as small jags on the electrogram and some smearing on the gel-like image. The RIN value for this sample was 6.8. Sample (ii) was not entirely degraded but the beginnings of degradation were present. However, the integrity was still classified as high enough for the array, as small RNAs are less susceptible to degradation than large RNAs.

A RIN value of above 6.5 was used as a cut-off for RNA samples to be used on the array. However, most of the samples used had a RIN value above 7. The gel-like images of samples were also inspected manually to ensure RNA integrity. The details of the RIN values of all RNA samples used on arrays can be found in Table 5.1.

5.2.3 Array QC report

To assess the quality of the array it was important to examine the QC report produced during data extraction by the FE software. An example of the QC report produced by the FE software is shown in Figure 5.4.

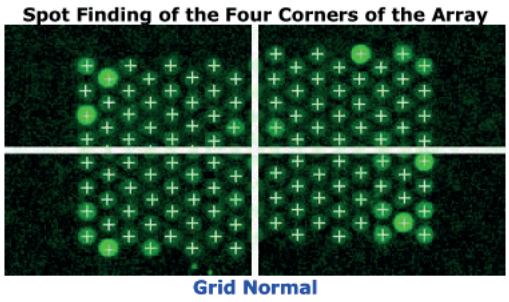
Before the FE software can extract data it must first align the spots found on the array image to a grid template. The QC report is divided into sections A-E, section A shows a

data image of the four corners of the array with the grid template overlaid. If the grid has been aligned correctly, the white crosses are found within the centre of the spots. This example shows a grid template aligning correctly. If a grid does not align to the spots, it may be that the wrong template has been used or that there are problems with the array image, and therefore data cannot be extracted from the array image.

Section B is an image displaying the spatial distribution of outliers on the array. There are two types of outliers: population outliers and non-uniform outliers. Each probe has several replicates. If a probe is not within a predefined range of the other replicate probes, it is classed as a population outlier. When the noise surrounding a particular spot is higher than the threshold, it is classified as a non-uniform outlier. The distribution of the outlying probes should be random. If the distribution of outliers is not random, this indicates an artefact on the array surface such as a bubble. If this occurs, all the probes found in this region should be removed from further analysis; if this region is too large, the data from this array should be excluded. Outlier probes are flagged in the data output from the FE software and are removed as part of the normalization process.

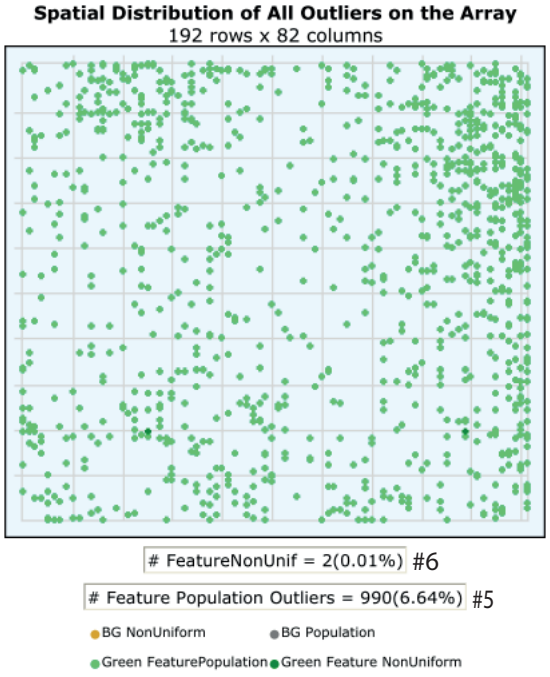
Section C shows the net signal statistics of the non-control probes. This gives an indication of the dynamic range of fluorescent signal on the array. The first bar in the table gives the number of saturated spots on the array. There should not be any saturated spots as this indicates the full range of expression is capped at the upper limit. No saturated spots were found on any of the arrays. An important value to be considered in this table is the signal level at the 99^{th} percentile; this gives an indication of the level of signal of bound probes on the array. If this value is low it indicates there is not much signal being detected on the array, which could indicate a problem with labelling or hybridization. It may also be that there is a lower level of total miRNAs present in the sample. When there has been a problem with either the labelling or the

A)



B)

	Feature	Local Background
Non Uniform	2	1
Population	990	2

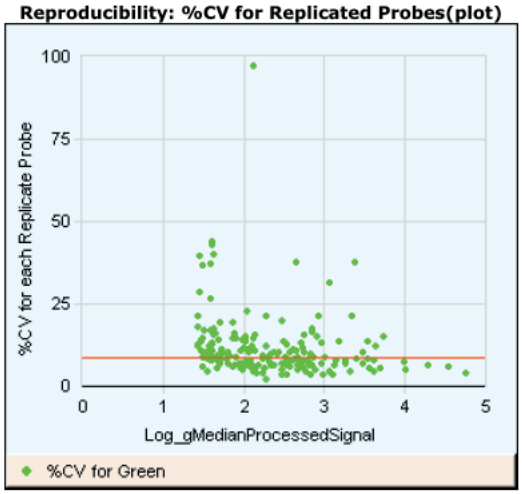


C) **Net Signal Statistics**

Non-Control probes:

# Saturated Features		0
99% of Sig. Distrib.	#8	2301
50% of Sig. Distrib.	#9	88
1% of Sig. Distrib.	#10	78

D)



E) **Evaluation Metrics for miRNA_QCMT_Sep09 :**

Excellent (4) ; Good (2)

Metric Name		Value	Excellent	Good	Evaluate
IsGoodGrid	#3	1.00	>1	NA	<1
AddErrorEstimateGreen	#4	2.73	<5	5 to 12	>12
AnyColorPrcntFeatPopnOL	#5	6.64	<8	8 to 15	>15
gNonCtrlMedPrcntCVBGSu...	#7	8.82	0 to 10	10 to 15	<0 or >15
gTotalSignal75pctile		0.10			
LabelingSpike-InSignal	#1	3.16		>2.50	<2.50
HybSpike-InSignal	#2	4.15		>2.50	<2.50
StringencySpike-InRatio		1.96			

Fig 5.4 Example of a QC report produced by Feature extraction software.
Data is extracted from the TIFF image of the arrray by feature extraction software; which produces a QC report. Key areas from QC report are shown. A) Grid alignment image. B) Outlier distribution across array surface. C) Net signal statistics for non-control probes. D) %CV for replicate probes is plotted. E) QC metrics report. (# for reference in later figures)

hybridization, the value at the 99th percentile falls below 300. The net signal statistic values for all arrays can be found in the QC columns 8-11 in Table 5.3.

Section D shows the coefficient of variation (%CV) of replicate probes. The lower the %CV, the higher the correlation between the replicates. Ideally these values should be as low as possible.

Section E shows the metric values assigned for different aspects of the QC report. The metric values for all arrays can be found in the QC columns 1-7 in Table 5.3. The metric values can be classified as ‘excellent’, ‘good’ or ‘evaluate’. In the table, a value that falls in the range of ‘excellent’ is purple, ‘good’ is blue, and ‘evaluate’ is red. Any arrays that had a QC metric of ‘evaluate’ were excluded from further analysis. IsGoodGrid (labelled #3) is related to section A of the QC report. If the template grid aligned correctly to the spots on the array, a value of 1 is given for this metric. If the grid did not align, a value of 0 is given for this metric. The AddErrorEstimateGreen (labelled #4) value relates to the background signal on the array. If there is an issue with the background signal on the array, this often indicates that there was a problem at the washing stage of the array. If the background is too high the array data cannot be used as the signal at the spots may be affected. AnycolourPrcntFeatPopnOL (labelled #5) is the percentage of spots that are population outliers; this is related to section B of the QC report. Outlier spots are removed from analysis as part of normalization but the number of outliers should be low. If the number of outliers is high, it indicates that there may be a problem with the array or the sample. gNonCtrMedPrcntCVBSu (labelled #7) is the median CV% of replicate probes and is related to section D of the QC report. The median CV% of replicate probes is used as a measure of intra-array signal reproducibility. If the median CV% value is high, it often indicates that the hybridization solution had not fully covered the array surface. Arrays with poor intra-array reproducibility must be excluded from further analysis. Labelling

spike-InSignal (labelled #1) and Hyb spike-InSignal (labelled #2) are the metric values assigned to spike-ins. Spike-ins are synthetic small RNAs added during the array process to assess the labelling and hybridization reactions. If the labelling spike-in value is below 2.50, it indicates that the labelling reaction has failed. If both the hybridization and the labelling spike-in values are below 2.50, it indicates that hybridization has failed. If either of these steps fails, the array data cannot be used.

Information in the QC report was used to decide whether the array data should be used or excluded from further analysis. It also allowed us to identify where problems had occurred during the array process, which enable optimisation of subsequent arrays.

5.3 Pilot microarray experiment

Before RNA samples from leukemic and normal BM cells were run on arrays, a pilot microarray experiment was done using RNA from K562 cells and AML cells. The aim of the pilot was to highlight any potential issues that may occur during the array process. It also gave us an opportunity to see if any technical bias occurred from the RNA preparation. One array support was used for the pilot which allowed for eight samples to be run: three AML samples and five K562 samples. Details of the RNA preparation and array QC results for the pilot experiment can be seen in Table 5.1.

RNA from K562 cells was prepared in a number of different ways to investigate whether this would have any impact on the data. RNA was extracted from a range of cell numbers (10^6, $2x10^5$ and $5x10^4$); certain RNA samples were vacuum concentrated; and instead of labelling 100ng of RNA as was the standard protocol, only 70ng of RNA was labelled for one sample.

Before the RNA was labelled for the array, the RNA integrity was assessed on the Agilent bioanalyser (see Figure 5.5a). The number of cells from which RNA was extracted did not have an effect the RIN score; RNA extracted from 10^6 cells had RIN scores of 9.2 and 9.4; $2x10^5$ cells had a RIN score of 8.9; and $5x10^4$ cells had a RIN score of 9.5. The process of vacuum concentration also did not affect the RIN score. RNA extracted from 10^6 cells was divided into two aliquots and one of the aliquots was vacuum concentrated. The aliquot that was not vacuum concentrated had a RIN score of 9.2 (Figure 5.5a(i)) and the aliquot that was vacuum concentrated had a RIN score of 9.4 (Figure 5.5a(ii)). The RNA samples, both from the AML and K562 cells, all had good integrity, with RIN scores above 7.

The RNA was labelled and hybridized to arrays. After the arrays had been scanned, the QC report was examined (Table 5.1). The QC report indicated that there was a problem with the background signal on the array surface. Four of the arrays had a background value within the 'evaluate' range, and the other four arrays had a background value within the 'good' range, but the background value was still high for these arrays. As the background issue was common to all arrays, it suggested that the problem was affecting the whole array support. The raw data image was also examined and this revealed swirls of aberrant signal on the array surface. This led us to believe the problem lay at the washing step. The swirls did not appear to affect the array data as they were mainly found on the edges of arrays rather than on the arrays themselves, also the other QC metrics fell within the 'good' or 'excellent' ranges. To prevent the background issue occurring in subsequent arrays we discussed this issue with Agilent, who also identified the wash steps as the most likely cause of this problem. No errors in the technique used to wash the slides could be identified, so Agilent replaced the wash buffers. Once fresh buffers were used, the background problem was not seen again, suggesting that this initial background problem was due to the wash buffers.

The values of the spike-in controls were also examined. The hybridization spike-in values for all the arrays fell within the ‘good’ range, signifying hybridization had occurred successfully in all arrays. The labelling spike-in values of six of the arrays fell within the ‘good’ range indicating the labelling reaction had occurred successfully in six of the arrays. However, two of the arrays had labelling spike-in values within the ‘evaluate’ range, indicating that, for these two arrays, the labelling reaction had not worked efficiently.

The data from the pilot array was then normalized and analysed. Figure 5.5b shows coefficient of variation (%CV) of replicate probes after normalization. The %CV of replicate probes is low across all of the arrays, signifying high inter-array data reproducibility post normalization.

Hierarchical cluster analysis was performed on the data from the arrays post normalization (Figure 5.5c). The first cluster division was between the K562 cells and the AML cells, indicating that the AML cells were closer to each other on a miRNA level than they were to K562 cells. The red boxes in the figure highlight samples that had a labelling spike-in value in the ‘evaluate’ range (see Table 5.1). The AML sample with inefficient labelling clustered away from the other AML samples. This could be because of the inefficient labelling or because of patient variation, as this sample was from patient BAS031 while the other two samples were from the same patient, BIR093. However, the K562 sample with inefficient labelling was also clustering away from the other K562 samples. A Pearson correlation analysis was conducted on the array data from the K562 samples to determine the level of correlation between arrays (Table 5.2). The correlation between the K562 arrays was very high with correlation coefficients of ~0.98. However, the level of correlation was lower between the array with the inefficient labelling and the other K562 arrays, which confirmed what was seen

Array position	Cell type	Sample information	Vacuum concent-rated	RIN score	Labelling spike-in signal #1	Hybridization spike-in signal #2	Background value #4	% population outliers #5	% non-uniform outliers #6	% CV for replicate probes #7	99% of signal distribution #8
1_1	K562	RNA extracted from 10^6 cells; 75ng of RNA labelled		9.2	3.36	3.83	14.11	5.11	0.01	12.88	2985
1_2	K562	RNA extracted from 2 x 10^5	X	8.9	3.23	3.85	12.86	6.21	0	9.66	2266
1_3	K562	RNA extracted from 10^6 cells		9.2	3.30	3.84	10.68	5.67	0.05	9.97	3258
1_4	AML	LMPP BIR93	X	7.3	3.42	3.87	11.89	7.80	0.02	8.54	4190
2_1	K562	RNA extracted from 5x10^4	X	9.8	3.34	4.04	19.29	5.71	0.01	13.72	2493
2_2	K562	RNA extracted from 10^6 cells	X	9.4	2.36	3.79	11.77	5.63	0	10.93	989
2_3	AML	GMP AML	X	8.2	2.33	3.79	10.84	6.08	0	10.96	852
2_4	AML	GMP AML	X	9.4	3.43	3.70	13.2	4.92	0	7.95	2848

QC score	Labelling spike-in signal	Hybridization spike-in signal	Background value	% population outliers	% CV for replicate
Excellent			<5	<8	0 to 10
Good	>2.50	>2.50	5 to 12	8 to 15	10 to 15
Evaluate	<2.50	<2.50	>12	>15	<0 or >15

Table 5.1 RNA sample details and QC information of pilot microarray experiment.
Numbers in purple represent a QC metric rating of excellent, numbers in blue represent a QC metric rating of good, numbers in red represent a QC metric rating of evaluate. (# relates to QC report Figure 5.4

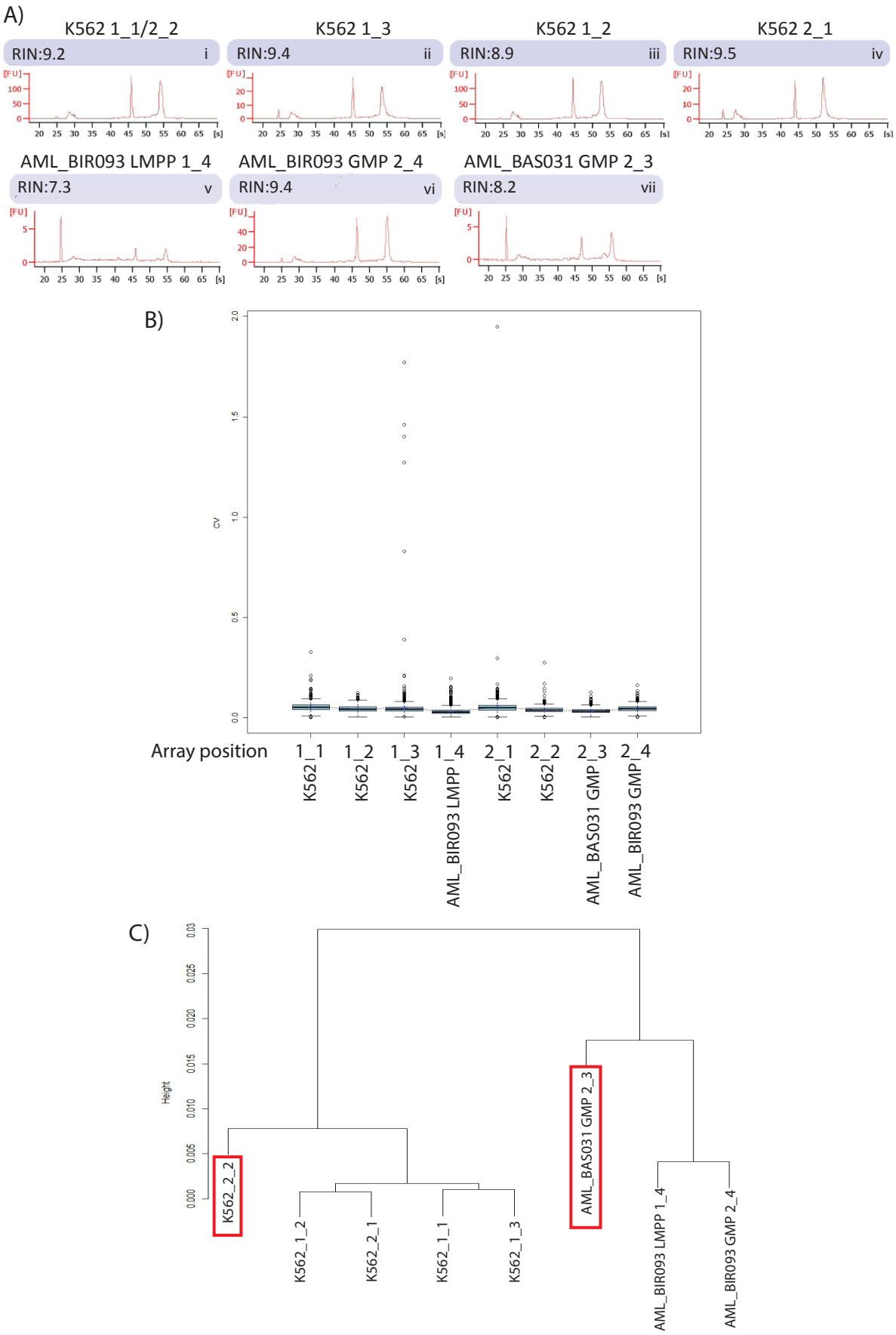


Fig 5.5 Pilot microarray experiment data analysis.
A trial experiment of the microarray system was performed using K562 cells and AML cells. A) RNA was run on the Agilent bioanalyser to check RNA integrity before it was used on microarrays. Electrograms for each sample are shown, i-iv K652 cells i) RNA extracted from 10^6 cells,pre-vacuum concentration ii) RNA extracted from 10^6 cells, post vacuum concentration iii) RNA extracted from 2x105 cell, iv) RNA extracted from $5x10^4$ cells, v-vii AML cells. B) Plot showing %CVs for replicate probes after normalization. C) Hierarchal clustering of arrays after normalization (Pearson distance). Boxes highlight arrays with a labelling spike-in value within the evaluate range.

on the hierarchal clustering (Figure 5.5B). This indicated that inefficient labelling had an adverse effect on the data produced.

Array 2_2 was the K562 array with the labelling issue. The RNA used on this array was also used for arrays 1_1 and 1_3. As neither of these arrays had a problem with the labelling, the problem was therefore not caused by the input RNA but by a technical issue. The most likely cause was that the RNA had not been fully denatured prior to labelling. Steps were taken to prevent this issue from occurring in the main array experiments. However in future arrays if the value of labelling spike-in was in the 'evaluate' range, array data was excluded from further analysis.

The RNA preparation did not impact on array profiles as the level of correlation was high between samples, independent of vacuum concentration or cell number (Figure 5.5c and Table 452). The K562 sample, where 70ng as opposed to 100ng of RNA was labelled, also gave a comparable profile to the other arrays after normalization (K562_1_1, Figure 5.5c and Table 5.2). This gave us confidence that the RNA preparation would not pose any bias in the main array experiments, provided the same extraction method was used throughout.

	K562_1_1	K562_1_2	K562_1_3	K562_2_1	K562_2_2
K562_1_1		0.982	0.989	0.984	0.917
K562_1_2	0.982		0.990	0.992	0.928
K562_1_3	0.989	0.990		0.985	0.915
K562_2_1	0.984	0.992	0.985		0.941
K562_2_2	0.917	0.928	0.915	0.941	

Table 5.2 Correlation between K562 arrays.
A Pearson pairwise correlation analysis was done between each of the K562 arrays, using data on all microRNAs after normalization. The correlation coefficients from each analysis are shown. Array with labelling spike-in value within the evaluate range is highlighted

5.4 Running the main microarrays

If RNA samples from the sorted leukemic and normal BM cells were judged to be of a suitable quantity and integrity, they were run on arrays. RNA samples were processed in batches of eight, as this was the number of samples that could be loaded per array support. Eleven array supports/ 88 arrays were run in total. Six HSC, three MPPs, seven GMP, seven CMP, 19 LMPP-like and 26 GMP-like samples in total were run on arrays. However, any arrays with a metric value within the 'evaluate' range or that had irregularities on the array image were excluded from further analysis, so six HSC, three MPP, five CMP, five GMP, 16 LMPP-like and 25 GMP-like samples were used for further analysis. Details on samples and QC results for array runs can be found in Table 5.3.

5.4.1 Synthetic spike-ins

Synthetic labelling and hybridization spike-ins were added to each sample during the array process to evaluate the efficiency of labelling and hybridization. The spike-ins values from the different array supports are plotted in Figure 5.6. Spike-in values within the 'evaluate' range were not included. Array supports 10 and 11 had hybridization spike-in values within the 'evaluate' range, but this is not because of inefficient hybridization, as the labelling spike-in values were within the 'good' range. The problem was therefore with the spike-in rather than the array. The spike-in values from the different supports were all at a similar level, suggesting that the efficiency of labelling and hybridization was similar across all array supports.

5.4.2 Correlation of RNA integrity and quantity to QC metrics

In the pilot experiment, neither the RNA integrity nor the quantity of RNA labelled had any effect on the data. To investigate if this was also true for the main array

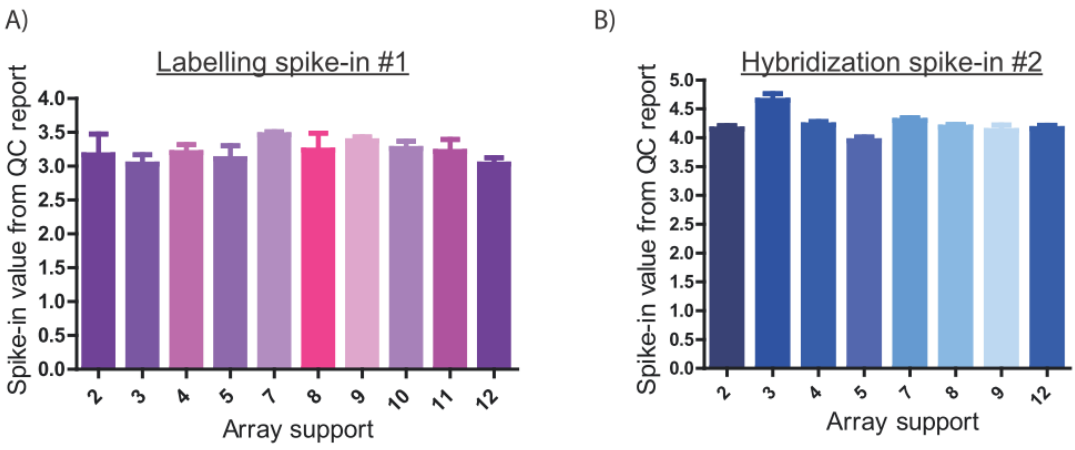


Figure 5.6 Synthetic spike-in controls.
Synthetic RNA controls known as spike-ins were added to RNA samples during the arrray process to identify problems with labelling and hybridization steps. Bars represent average spike-in value of each array support (Figure 5.4 E), one standard deviation is shown by error bars. Data from arrays that had spike-in values within the evaluate range were excluded. A) Labelling spike-in. B) Hybridization spike-in. (# relates to QC report Figure5.4)

experiment, the RIN score and the RNA quantity were compared to the QC metrics. A Pearson correlation analysis was used to determine if there was any correlation between the quantity and integrity of RNA to the QC metrics. The correlation co-efficients from the correlation analyses can be found in table 5.4. There is no correlation between the RIN score and any of the metrics, and there is also no correlation between the RNA quantity labelled and any of the metrics. If correlation had been observed, this would have indicated that the level of RNA integrity or quantity had caused a bias in the data; however, as there is no correlation, it is unlikely that a bias occurred due to the integrity or quantity of RNA.

QC metric	Labelling spike-in #1	Hybridization spike-in #2	background #4	Population outliers #5	%CV of replicate probes #7	Signal at 99% #8
RIN score	0.07	0.18	0.1	0.08	0.12	0.24
Quantity of RNA labelled	0.013	0.17	0.12	0.2	0.12	0.33

Table 5.4 Correlation between quantity of RNA labelled, RIN score and QC metrics.
A Pearson pairwise correlation analysis was conducted between the RIN score and each QC metric and a Pearson pairwise correlation was conducted between the quantity of RNA labelled and each QC metric. The correlation coefficients from each analysis are shown. (# relates to the QC report Figure 5.4)

5.5 Bioinformatic analysis of array data.

Once all the arrays had been run and the arrays that were to be used for analysis selected, the data was subjected to normalization. There were two stages involved in normalization: intra-array normalization and inter-array normalization. The normalization process removed any anomalies, adjusted for background, and enabled comparison between the different arrays. Once normalization was complete, the miRNA expression levels in each of the samples was then ready to be investigated. There were probes against ~900 miRNAs present on arrays but only 150-200 miRNAs were found to be expressed in any of the populations analysed.

5.5.1 Hierarchal Clustering

As it is standard when conducting analyses such as clustering to use a reduced gene set rather than all 900 miRNAs analysed. Hierarchal clustering of all populations was conducted based on an Anova gene list (Figure 5.7). Using a reduced gene set reduces the noise and thus aids clustering. An Anova test was conducted on the normal populations to determine the most differentially expressed miRNAs, which produced a list of 74 significantly differentially expressed miRNAs.

The first observation of the clustering was that the AML and normal populations did not cluster as two separate populations. The HSC and MPP populations clustered together, indicating that these two immature populations were most similar to each other at an miRNA level. The CMP and GMP populations also clustered together indicating that these two more mature populations were similar to each other. The first branch division on the hierarchal tree is the HSC/MPP populations clustering away from the other populations. This reveals that the AML populations were therefore more

similar to the CMP/GMP populations than the HSC/MPP populations at an miRNA level.

The normal populations tended to cluster together by population type, however the AML GMP-like and LMPP-like populations were dispersed throughout the hierarchal tree. The clustering of normal populations by population type is not surprising, as you would predict that the miRNA profiles would be most similar between cells of the same immunophenotype. However, other aspects of the AML samples were influencing their clustering.

LMPP-like and GMP-like samples that were from the same patient were stared and numbered, with the number indicating which patient the sample was from (Figure 5.7). GMP-like and LMPP-like samples that were from the same patient clustered together, implying that the patient from which the sample originates had a large impact on the clustering of leukemic populations.

The genetic variation between normal individuals will be minimal; however the genetic variation between leukemic individuals will be much greater due to the patient specific mutations and chromosomal abnormalities. It is likely that patient-specific mutations and chromosomal abnormalities influenced the clustering of leukemic populations. Unfortunately, there was not enough information on the mutational status of samples arrayed to determine whether this was having an impact on clustering. However, the cytogenetic information was known for most samples arrayed and therefore the cytogenetic status of each of the samples was annotated underneath the hierarchal tree. This revealed that AML samples did appear to cluster by cytogenetic subtype.

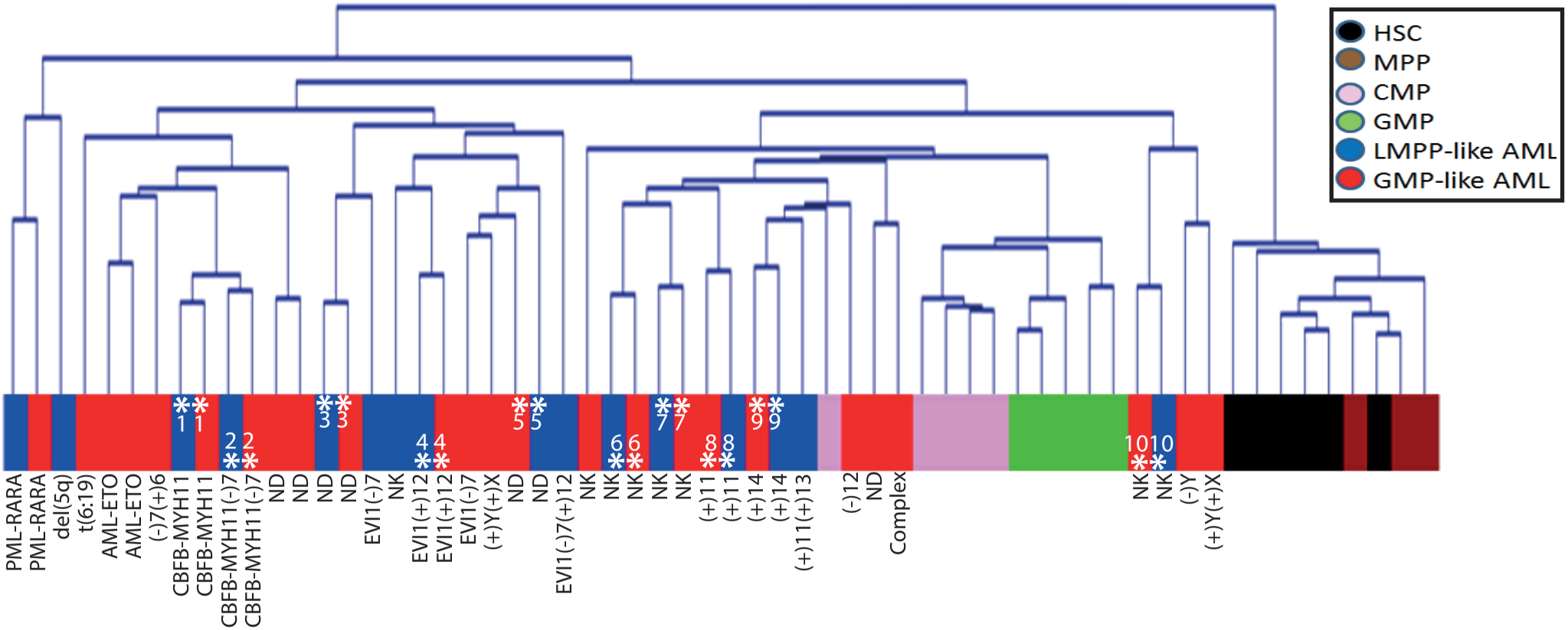


Fig 5.7 Hierarchical clustering of normal stem and progenitor and AML LMPP-like and GMP-like populations.
Supervised hierarchical clustering of microRNA expression profiles of normal FACS sorted stem and progenitor populations (six HSC, three MPPs, five CMPs and GMPs) and FACS sorted AML populations (16 LMPP-like and 26 GMP-like). An ANOVA gene set of differentially expressed microRNAs in normal populations was used. The clustering was done using the Euclidean distance. Writing underneath the cluster indicates cytogenetic phenotype of AML patient. White stars signify where the LMPP-like and GMP-like populations come from the same patient, number reflects patient.

5.5.2 Three dimensional (3D) principle component analysis (PCA)

To further explore how the microRNA profiles of our populations related to each other, a 3D PCA was conducted. The 3D PCA is a mathematical transformation that allows an infinitive number of data points to be displayed as one node in a three dimensional plot. Array data from a sample can therefore be displayed as a single node on a 3D plot. Array data from multiple samples is transformed to nodes as part of the analysis, and how the nodes lie in relation to each other reflects how similar the expression profiles of the samples are. A 3D PCA displaying the miRNA expression profiles of our populations was generated using the Anova gene set (Figure 5.8a). Each node on the plot represents an miRNA expression profile from one sample and the colour of the node represents the population of the sample.

The nodes on the PCA plot clustered by population type, signifying that miRNA profiles were most related between the samples within the same population. The nodes of HSC and MPP samples clustered together and the nodes of the CMP and GMP samples clustered together. This confirms what was seen in hierarchal clustering. However, there was a large distance between the HSC/MPP cluster and the CMP/GMP cluster, indicating that the miRNA profiles of the HSC/MPP populations were largely distinct from the miRNA profiles of the CMP/GMP populations.

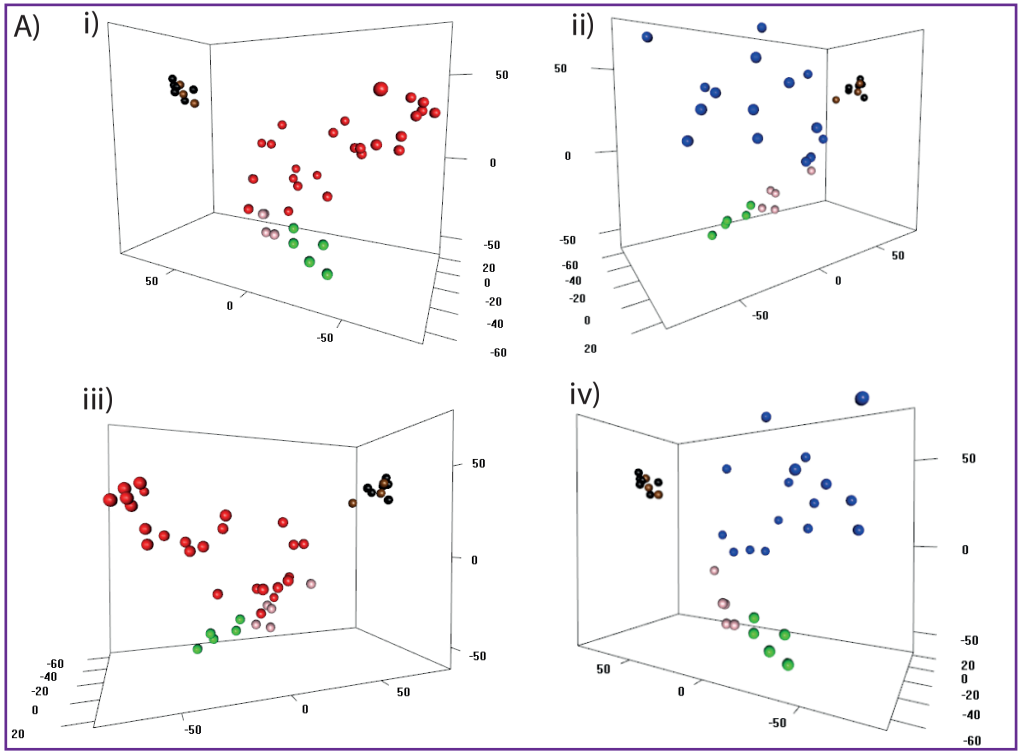
The nodes of the AML samples also clustered together, but the clusters were more dispersed than the normal populations. It is likely that this reflects the heterogeneity of leukemic samples. The GMP-like AML samples clustered closer to the GMP/CMP samples than the HSC/MPP samples. This reiterates what was seen in the hierarchal clustering that, on a microRNA level, the GMP-like AML samples were more similar to the GMP/CMP normal populations than the HSC/MPP populations. Although the GMP-

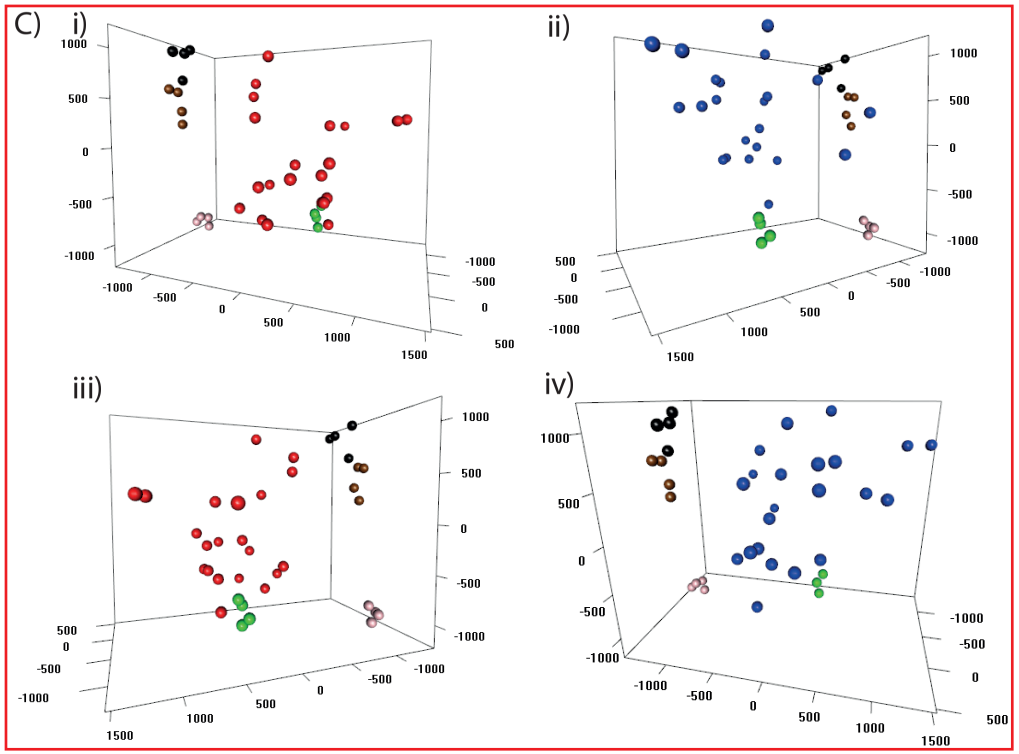
like samples clustered closer to the GMP/CMP samples, they did not cluster with them. This indicates that while the miRNA profiles of leuekmic cells are closet to the cells they share an immunophenotype with, the profiles are distinct from normal cells. The LMPP-like AML samples did not cluster near either the HSC/MPP cluster or the CMP/GMP cluster but appeared to cluster midway between the two groups.

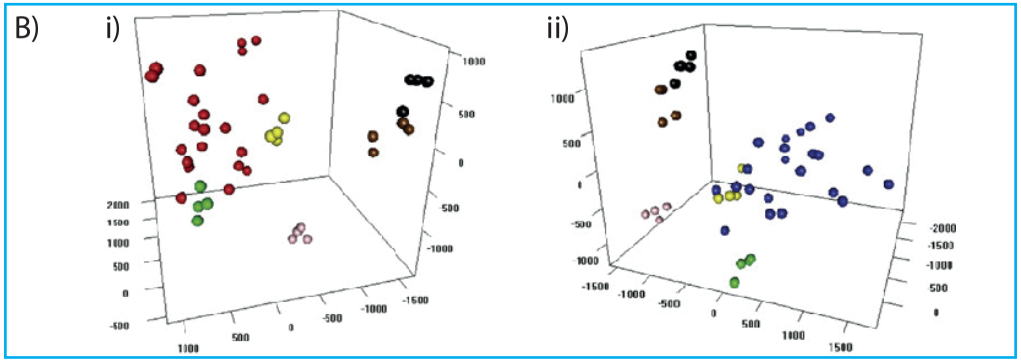
Previously in our lab 3D PCA plots were generated using mRNA expression profiles of normal HSC, MPP, LMPP, CMP, GMP and AML LMPP-like and GMP-like samples (Figure 5.8b) [15]. This allowed us to compare what was seen on an miRNA level with what was seen on an mRNA level. In previous work mRNA PCA plots contained the normal LMPP population. As this was not included in the miRNA analysis, the mRNA PCAs were repeated but excluding the normal LMPP samples (Figure 5.8c).

Like the miRNA plot, the HSC and MPP samples clustered together on the mRNA plot. The CMP and GMP populations lay closer to each other on the mRNA plot than they did to the HSC/MPP cluster, indicating that the GMP and CMP populations were more similar to each other on an mRNA level than they were to the HSC/MPP populations. However, in the mRNA plot, the CMP and GMP samples did not cluster together as they did on the miRNA plot. This implies that the CMP and GMP populations were less distinct from each other on an miRNA level than they were on an mRNA level.

On an miRNA level, the GMP-like samples clustered closest to the GMP/CMP cluster, and on an mRNA level the GMP-like samples clustered closest to the GMP cluster. Therefore, the GMP-like samples on a molecular level most closely resembled the population they shared an immunophenotype with. The GMP-like samples clustered closest to GMP clusters, but still formed separate distinct clusters on both the mRNA and miRNA plots. Therefore even though GMP-like cells were most similar to GMP







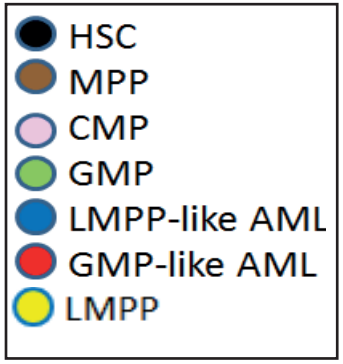


Fig 5.8 3D principle component analysis (PCA) of normal stem and progenitor and AML LMPP-like and GMP-like populations.
3D PCAs displaying expression profiles of normal FACS sorted stem and progenitor populations and FACS sorted AML populations. A) miRNA expression profiles i & iii normal populations with GMP-like AML (two different angles of the same plot are shown), ii & iv normal populations with LMPP-like AML (two different angles of the same plot are shown). B) Original mRNA expression profiles from Goardon et al paper (Figure 4C) (ref) i normal populations with GMP-like AML, ii normal populations with LMPP-like. C) mRNA expression profiles using data form Goardon et al paper (ref) without the normal LMPP population, i & iii normal populations with GMP-like AML (two different angles of the same plot are shown), ii & iv normal populations with LMPP-like AML (two different angles of the same plot are shown).

cells, these data show that alterations have occurred to both the mRNA and miRNA expression profiles in the leukemic cells.

At the mRNA level, the LMPP-like samples clustered closest to the normal LMPP population. However, when the normal LMPP population was not present, the LMPP-like samples clustered midway between the GMP cluster and the HSC/MPP cluster, similar to what was seen on an miRNA level.

Overall our populations relate to each other in a similar way at the both the miRNA and mRNA level. However the CMP and GMP populations, exhibit greater similarity at a miRNA level than at an mRNA level.

5.5.3 T-tests

The PCA and hierarchal cluster analysis suggested that the microRNAome in leukemic cells was altered compared to normal cells. To further investigate changes of the microRNAome in the leukemic samples, t-tests were performed. As the normal GMP population was the most similar to the GMP-like AML population on an immunophenotypic and molecular level, a t-test was performed to generate a list of miRNAs that were significantly different between these two populations. Using a significance level of <0.01 p value, 25 miRNAs were found to be significantly differentially expressed between the GMP-like samples and the GMP samples. 16 miRNAs were more highly expressed and eight miRNAs had lower expression in the GMP-like population compared to the GMP population. The *p* value and expression fold change of these miRNAs can be seen in Table 5.5. As significantly differentially expressed miRNAs can be identified, this supports the theory that the microRNAome in leukemic cells is altered.

					Additional t-tests				
gene name	GMP-like AML mean expression	GMP mean expression	p-value	expression fold change	HSC/MPP v CMP/GMP	LMP-like AML V GMP	GMP v CMP	AML v NORMAL	GMP-like AML v LMPP-like AML
hsa-miR-26a	7.36	6.41	1.20E-06	1.94	HSC/MPP	LMPP-like			
hsa-miR-363	5.39	8.20	2.65E-06	7.02	CMP/GMP	GMP		Normal	
hsa-miR-20b	7.53	9.12	3.15E-06	3.01	CMP/GMP	GMP		Normal	
hsa-miR-30b	7.22	6.40	3.83E-06	1.77		LMPP-like		AML	
hsa-miR-195	3.37	2.62	9.39E-05	1.68		LMPP-like			
hsa-miR-720	12.76	13.75	1.71E-04	1.99	CMP/GMP	GMP		Normal	
hsa-miR-551b	4.41	6.63	1.80E-04	4.67	HSC/MPP		CMP	Normal	
hsa-miR-26b	8.44	7.54	2.59E-04	1.87	HSC/MPP	LMPP-like			LMPP-like
hsa-miR-342-5p	3.39	2.90	2.69E-04	1.40	HSC/MPP	LMPP-like			
hsa-miR-210	3.23	2.43	2.83E-04	1.74		LMPP-like			
hsa-miR-92a	8.79	9.57	4.94E-04	1.71	CMP/GMP	GMP			
hsa-miR-21*	3.96	4.75	0.00135	1.72		GMP			
hsa-miR-17	8.64	9.54	0.001656	1.86	CMP/GMP	GMP			GMP-like
hsa-miR-23a	9.31	8.90	0.002218	1.32	HSC/MPP	LMPP-like			
hsa-miR-142-5p	8.46	7.69	0.002229	1.71		LMPP-like		AML	
hsa-miR-361-3p	4.18	3.26	0.002331	1.89		LMPP-like		AML	
hsa-miR-19a	9.29	10.02	0.002969	1.66	CMP/GMP	GMP			
hsa-miR-222	5.08	4.44	0.003243	1.57				AML	
hsa-miR-886-3p_v15.0	4.29	7.26	0.003269	7.82	CMP/GMP	GMP		Normal	
hsa-miR-146b-5p	6.66	5.59	0.003863	2.11	HSC/MPP	LMPP-like			
hsa-miR-181c	5.37	4.04	0.004126	2.52	HSC/MPP	LMPP-like			
hsa-miR-30d	7.18	6.79	0.00441	1.31				AML	
hsa-miR-151-5p	4.55	3.13	0.004711	2.68	HSC/MPP		CMP		
hsa-miR-186	3.93	3.24	0.008259	1.61	HSC/MPP	LMPP-like			
hsa-miR-34a	6.67	5.11	0.009395	2.93					GMP-like

Table 5.5 miRNAs significantly differentially expressed between normal GMP samples and GMP-like AML samples.
A t-test analysis was conducted to identify miRNAs that were differentially expressed between normal GMPs and GMP-like AMLs (green indicates miRNA more expressed in GMP-like and red indicates miRNA more expressed in GMP). Mean expression is log2 transformed. Additional t-tests were conducted comparing different populations and if a miRNA came up significant in that analysis it is highlighted in table (the population that the miRNA is more expressed in is indicated in the box). A P-value cut off of <0.01 was used.

To investigate the differences between normal cell populations, t-tests comparing them were performed. Only two miRNAs were found to be significantly differentially expressed between HSCs and MPPs and eight miRNAs were found to be significantly differentially expressed between CMPs and GMPs. This reflects what was seen on the PCAs and the hierarchal cluster, that these populations are highly similar to each other on an miRNA level.

To gain an understanding of the differences between the immature HSC/MPP cells and the more mature CMP/GMP cells, a t-test was performed comparing HSC/MPP cells to CMP/GMP cells. 70 miRNAs were found to be differentially expressed; 44 miRNA were more expressed in HSC/MPP cells than in GMP/CMP cells; and 26 miRNAs were less expressed in HSC/MPP cells than in GMP/CMP cells. There is a high degree of miRNA variation between HSC/MPP cells and CMP/GMP cells which is consistent with what was seen on the hierarchical cluster and the PCAs. The 70 differentially expressed miRNAs represent the miRNA differences between stem cells and more mature progenitor cells, and so we termed this miRNA signature a stem cell-like miRNA signature.

We hypothesised that the GMP-like AML cells may have acquired a more stem cell-like miRNA profile. To investigate this, the miRNAs that were found differentially expressed between GMP-like cells and GMPs cells were compared with the miRNAs that were found differentially expressed between HSC/MPP cells and CMP/GMP cells. Of the 16 miRNAs that were over-expressed in GMP-like cells compared to GMP cells, nine were expressed at significantly higher levels in HSC/MPP cells compared to GMP/CMP cells (Figure 5.9a). Of the 14 miRNAs that were less expressed in GMP-like cells compared to GMP cells, 12 had significantly lower expression in HSC/MPP cells compared to the GMP/CMP cells (Figure 5.9a). These data suggest that some of the alterations that have occurred in the miRNA profile of GMP-like cells are stem cell-like.

There was no normal LMPP array data available so it was not possible to compare LMPP-like AML cells to normal LMPP cells in order to investigate the leukemic aberrations in the LMPP-like cells. It is thought that LMPP-like cells may give rise to GMP-like cells. A t-test was conducted comparing LMPP-like AML cells to GMP cells, to explore whether the differences seen between to GMP-like cells and GMP cells would also be seen between LMPP-like cells and GMP cells. There were 37 miRNAs differentially expressed between LMPP-like cells and GMP cells. 20 miRNAs were over-expressed in LMPP-like cells compared to GMP cells and of these, 12 miRNAs were also over-expressed in GMP-like cells compared to GMP cells. 17 miRNAs had lower expression in LMPP-like cells compared to GMP cells and of these, 13 miRNAs had lower expression in GMP-like cells compared to GMP cells (Figure 5.9b). Most of the miRNAs found to be differentially expressed between GMP-like cells and GMP cells were also differentially expressed between LMPP-like cells and GMP cells. A paired t-test was also conducted comparing GMP-like cells to LMPP-like cells. There were 12 miRNAs significantly differentially expressed between these populations.

Many of the previous papers looking at microRNA expression in AML compared non-purified leukemic populations to non-purified normal populations. To investigate how different our data would be using non-purified populations, a t-test was conducted comparing the two AML populations to the normal HSC, MPP, CMP and GMP populations. 52 miRNAs were differentially expressed between AML cells and normal cells. 36 miRNAs were over-expressed in AML cells compared to normal cells and of these, only six miRNAs were found over-expressed in GMP-like cells compared to GMP cells. 16 miRNAs were less expressed in AML cells compared to normal cells and of these, only five miRNAs were found to be less expressed in GMP-like cells compared to GMP cells (Figure 5.9c). This suggests that comparing non-purified

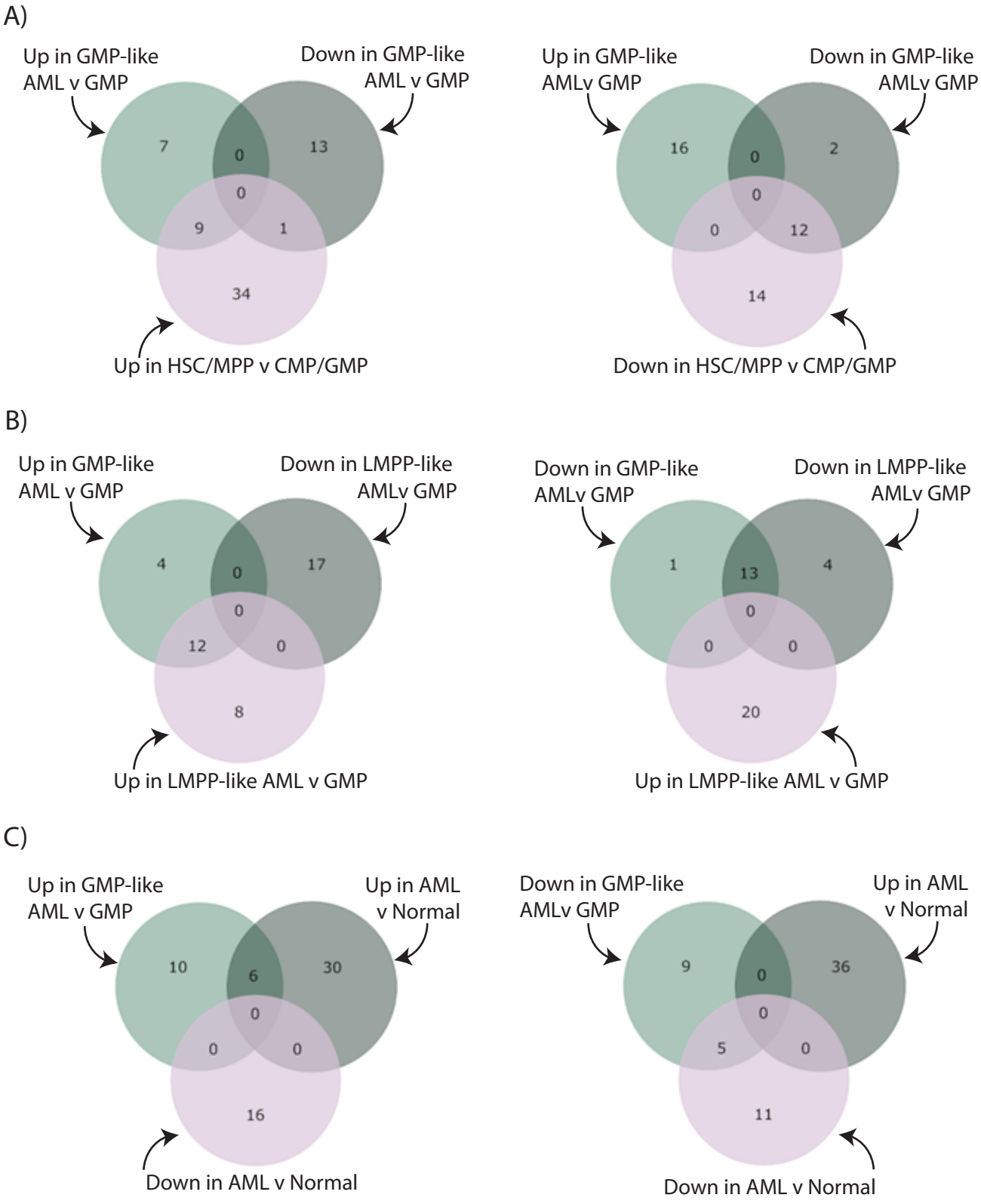


Fig.5.9 Venn diagrams of significant miRNA overlap
A t-test was conducted to identify miRNAs that were differentially expressed between GMP samples and GMP-like samples. T-tests were also conducted between other populations to identify differentially expressed miRNAs. Venn diagrams show overlap of miRNAs that were identified as significantly differentially expressed between GMP versus GMP-like t-test and the other t-tests performed.

populations would lead to a different interpretation of the aberrations in the microRNAome of leukemic cells.

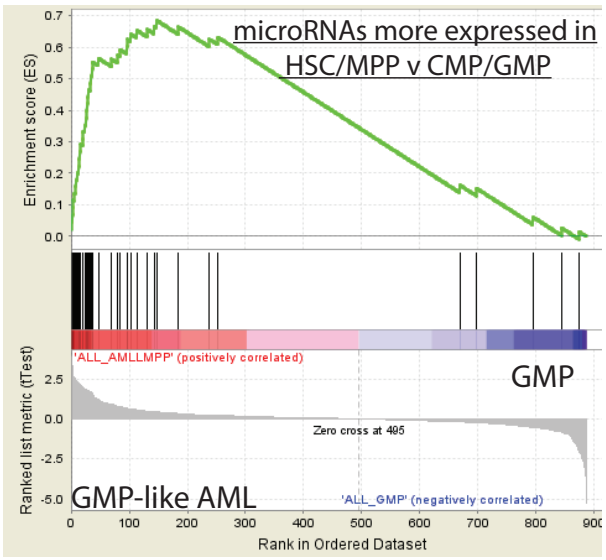
5.5.4 Gene Set Enrichment analysis

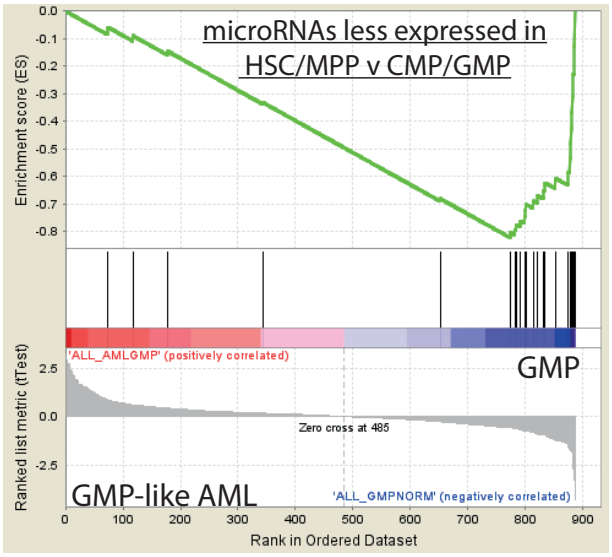
We hypothesised that the GMP-like AML cells may have acquired a more stem cell-like miRNA profile, and examination of the differentially expressed miRNAs in GMP-like cells compared to GMP cells supported this. To investigate further, gene set enrichment analysis was performed. Previously a miRNA stem cell-like signature was derived by identifying miRNAs that were significantly differentially expressed between HSC/MPP cells and GMP/CMP cells. GSE analysis was used to investigate whether there was any significant enrichment of the stem cell-like miRNA signature in GMP-like cells compared to GMP cells (Figure 5.10a). These data showed that miRNAs that were more expressed in HSC/MPP cells were significantly enriched in GMP-like cells, whereas miRNAs that were less expressed in HSC/MPP cells were significantly enriched in GMP cells. This demonstrates that there was an enrichment of the stem cell-like miRNA signature present in GMP-like cells.

GSE analysis was also used to investigate whether the stem-cell-like miRNA signature was more enriched in LMPP-like cells compared to GMP cells (Figure 5.10b). miRNAs that were more expressed in HSC/MPP cells were significantly enriched in LMPP-like cells, whereas miRNAs that were less expressed in HSC/MPP cells were significantly enriched in GMP cells. The GSE analysis thus demonstrates that there is also an enrichment of the stem cell-like miRNA signature in LMPP-like cells.

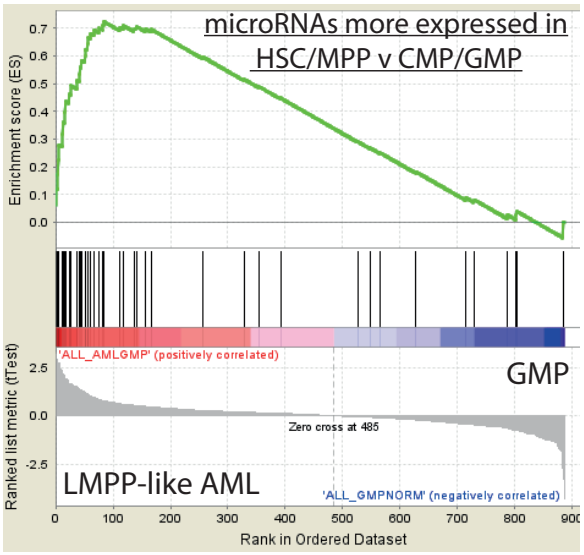
The frequency of LSCs in the LMPP-like population is higher than the GMP-like population, therefore it was postulated that the stem cell-like miRNA signature would be more enriched in the LMPP-like population than the GMP-like population. GSE

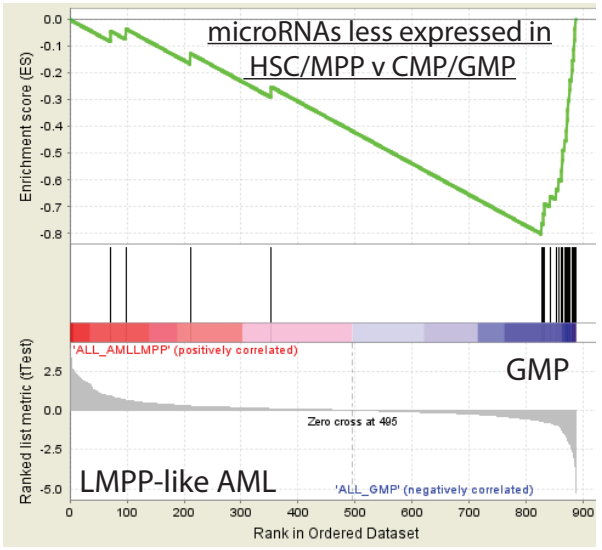
A)



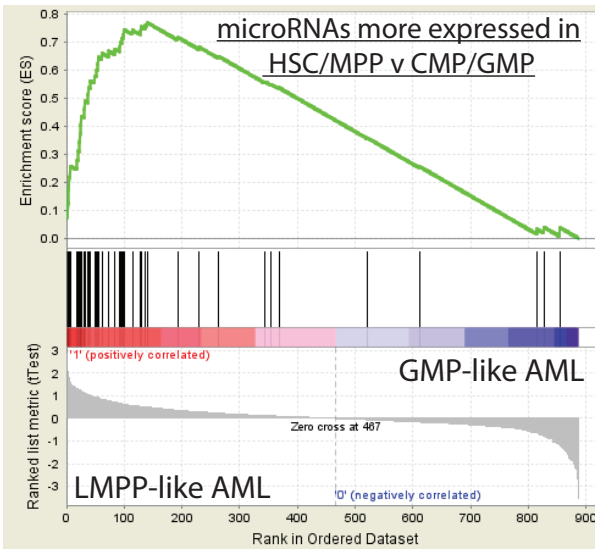


B)





C)



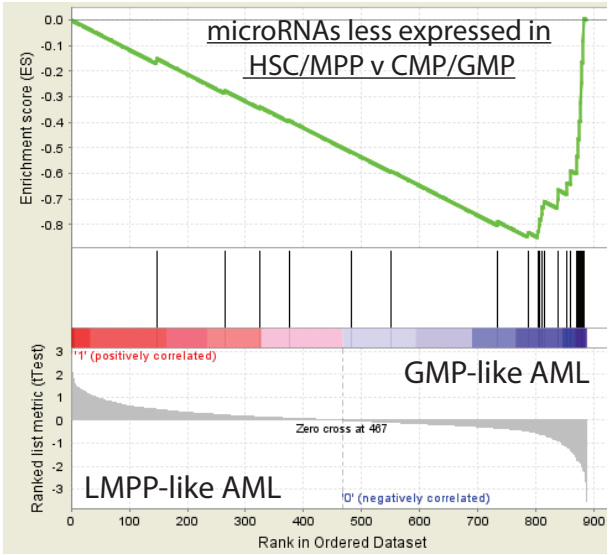


Fig 5.10 Gene set enrichment analysis.
A stem cell-like miRNA signature was indentfied; gene set enrichment analyses were conducted to establish if this signature was more enriched in one population compared to another A) GMP-like population and GMP population compared B) LMPP-like population and GMP population compared C) LMPP-like and MGP-like population compared

analysis revealed that the stem cell-like miRNA signature was indeed more enriched in the LMPP-like population than the GMP-like population (Figure 5.10c).

5.5.5 Summary of bioinformatic analysis

In summary bioinformatic analysis revealed that on a miRNA level the GMP-like population most closely resembled the normal GMP/CMP populations. However, it is evident that the miRNA profile of the GMP-like cells was altered as compared to normal cells, with the GMP-like cells acquiring an aberrant stem cell-like miRNA signature. GMP-like cells were also heterogeneous at a microRNA level, and it is likely that this reflects the different cytogenetics and gene mutations of patients.

5.6 Differentially expressed miRNAs in GMP-like cells

A t-test comparing the miRNA expression profiles between GMP-like and GMP cells identified 25 miRNAs as significantly differentially expressed between the two populations. To further examine these microRNAs, their expression in each population was plotted. MicroRNAs that were over-expressed in GMP-like cells can be seen in Figure 5.11 and microRNAs that were less expressed in GMP-like cells can be seen in Figure 5.12.

The expression levels of these miRNAs within the normal cell populations were similar. The expression levels were much more varied in AML cell populations. However, the expression pattern was often similar between the GMP-like samples and the LMPP-like samples.

16 miRNAs were over-expressed in GMP-like cells and of these, nine miRNAs (miR-151, miR-181c, miR-186, miR-195, miR-342-5p, miR-23a, miR-29a, miR26b and miR-146b-5p) had higher expression in both AML populations than the GMP and CMP

populations, but lower or equivalent expression in both AML populations compared to the HSC and MPP populations. These data support the GSE analysis, where the GMP-like and LMPP-like cells were shown to have a more stem cell-like microRNA profile. In six of the over-expressed miRNAs (miR-30b, miR-30d, miR-34a, miR-142-5p, miR-222 and miR-361-3p) the expression in both AML populations was higher than the expression in all of the normal populations. These miRNAs changes do not reflect a transformation to more stem cell-like microRNA phenotype, and so may provide to a different advantage to the leukemic cells. The final over-expressed miRNA (miR-210) was expressed at a higher level in both AML populations compared to the GMP population; however, its expression in the other normal populations is highly variable. This may be due to a problem the reliability of the array probes used to measure the expression of miR-210 (Figure 5.11).

miRNA families are defined as mature miRNAs that only differ from each other by a couple of nucleotides and that contain the same seed sequence. miR-26a and miR-26b are both from the same microRNA family. The expression pattern of miR26a and 26b over the different cell types is very similar. miR30b and miR-30d also belong to the same microRNA family, but their pattern of expression over the different cell populations varies slightly. (Figure 5.11).

In the AML samples, a bimodal distribution of expression was seen for some of the differentially expressed miRNAs. In the over-expressed microRNAs, the most obvious examples of this were miR-151 and miR-181c. The AML samples expressing the lowest level of miR-151 and miR-181c were at a similar expression level to GMP cells, and the AML samples expressing the highest level of miR-151 and miR-181c were at a similar expression level to the HSC and MPP cells (Figure 5.11). In microRNAs with lower expression in GMP-like AML, the most obvious examples of this were miR-551b and miR-363. The AML samples expressing the highest level of miR-551b and miR-

363 were at a similar level of expression to the normal samples, and the AML samples expressing the lowest level of miR-151 and miR-181c expression were below the level of expression in all of the normal samples (Figure 5.12).The source of this bimodal distribution in AML samples is unknown. It is likely that AML samples over-expressing or with reduced expression of the same miRNA share some unknown factor.

Nine miRNAs had expression in GMP-like cells and of these, seven miRNAs (miR-17, miR-92a, miR-720, miR-19a, miR-20b, miR-21* and miR-886-8p) had lower expression in both AML populations than the GMP and CMP populations, but higher or equivalent expression in both AML populations compared to the HSC and MPP populations. These data support what was seen in the GSE analysis, that GMP-like and LMPP-like cells were shown to have a more stem cell-like microRNA profile. In two of the less expressed miRNAs (miR-363 and miR551b) the expression in both AML populations was lower than that in all normal populations (Figure 5.12).

The miRNAs with lower expression in AML (miR-17, miR-19a and miR-92a) are part of the mir-17-92a polycistronic cluster, and the miRNAs with lower expression (miR-20b and miR363) are part of the polycistronic cluster mir-106a-363. The mir-17-92a and the mir-106a-363 clusters are paralogous clusters. miR-17 and miR 20b also share the same seed sequence and miR-92a and miR-363 share the same seed sequence (Figure 5.12). A Pearson correlation analysis was performed to investigate whether there was any correlation in the expression of these miRNAs in AML (Table 5.6). The highest level of correlation was between miRNAs that were part of the same polycistronic cluster; this is not surprising as miRNAs within the same cluster are transcribed together. The highest level of correlation was seen between miR-17 and miR-19a (Table 5.6). The level of correlation between the polycistronic clusters was more ambiguous, as miR-20b showed correlation to the miRNAs within the mir-17-92

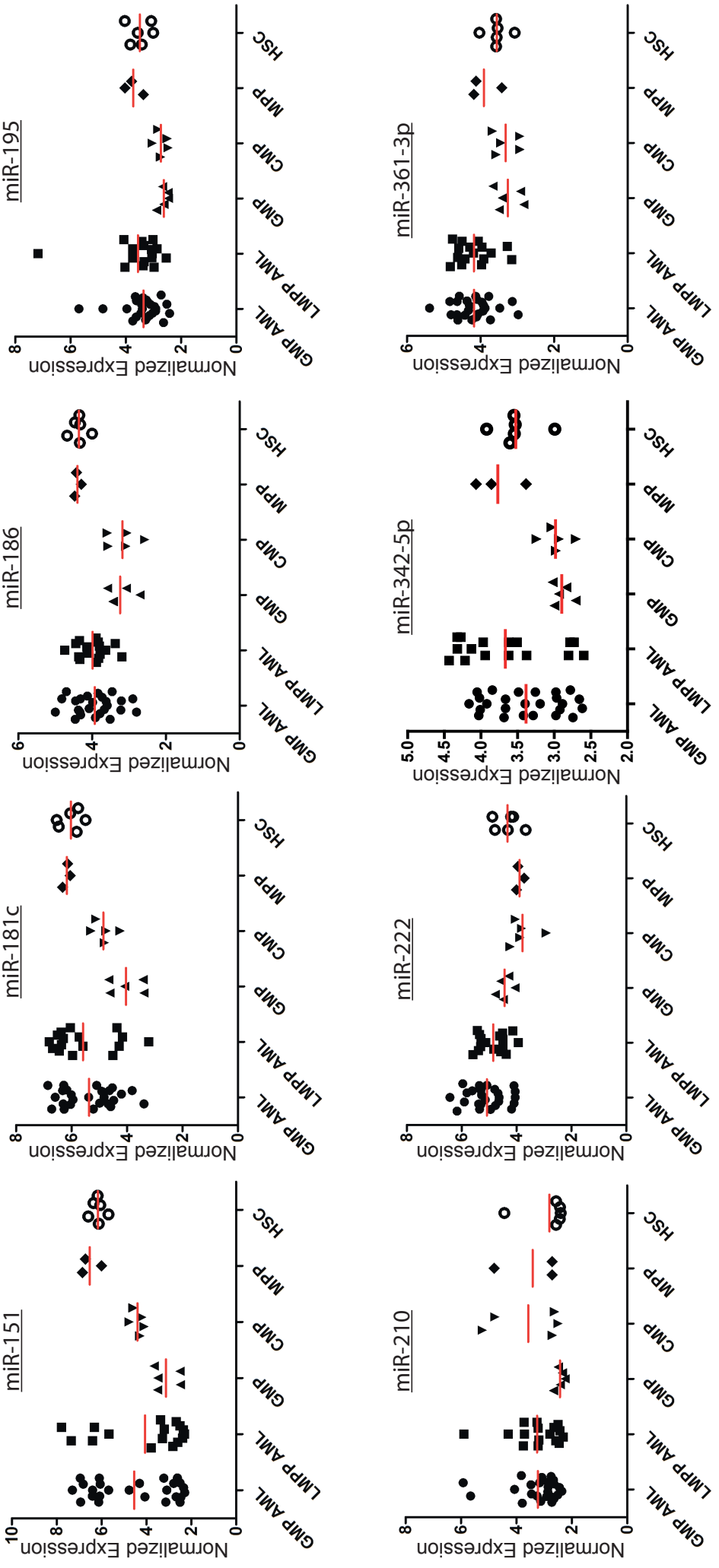


Fig 5.11 MicroRNA expression profiles of miRNAs significantly more expressed in GMP-like AML compared to normal GMP.
Normalized expression is plotted for each sample analysed on an array. Normalized expression value is log2 transformed. Red line represents population mean.

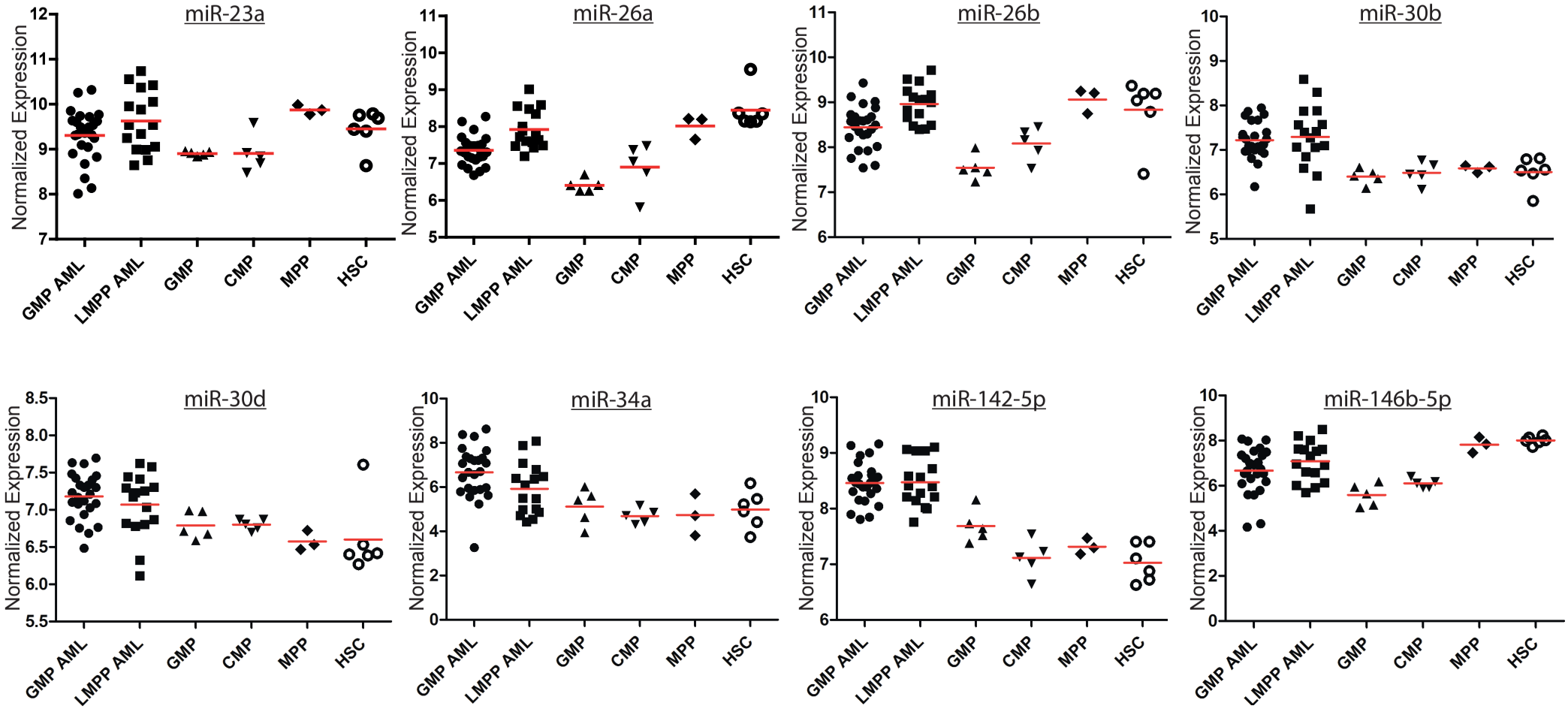


Fig 5.11 MicroRNA expression profiles of miRNAs significantly more expressed in GMP-like AML compared to normal GMP.
Normalized expression is plotted for each sample analysed on an array. Normalized expression value is log2 transformed. Red line represents population mean.

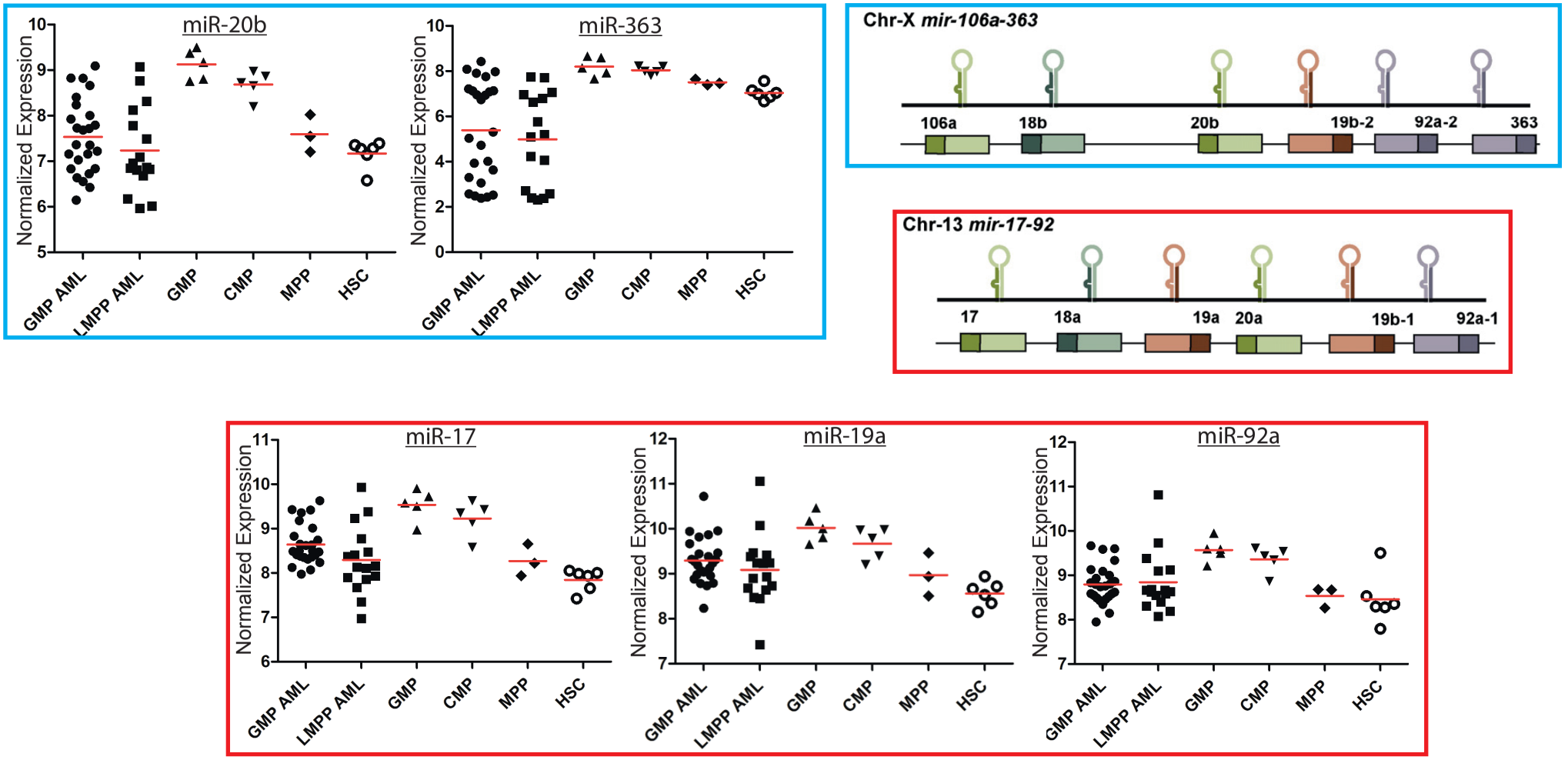


Fig 5.12 MicroRNA expression profiles of miRNAs significantly more expressed in normal GMP compared to GMP-like AML.
Normalized expression is plotted for each sample analysed on an array. Normalized expression value is log2 transformed. Red line represents population mean. Blue box indicates miRNAs part of the 106-363 cluster and red box indicates miRNAs part of the 17-92 cluster.

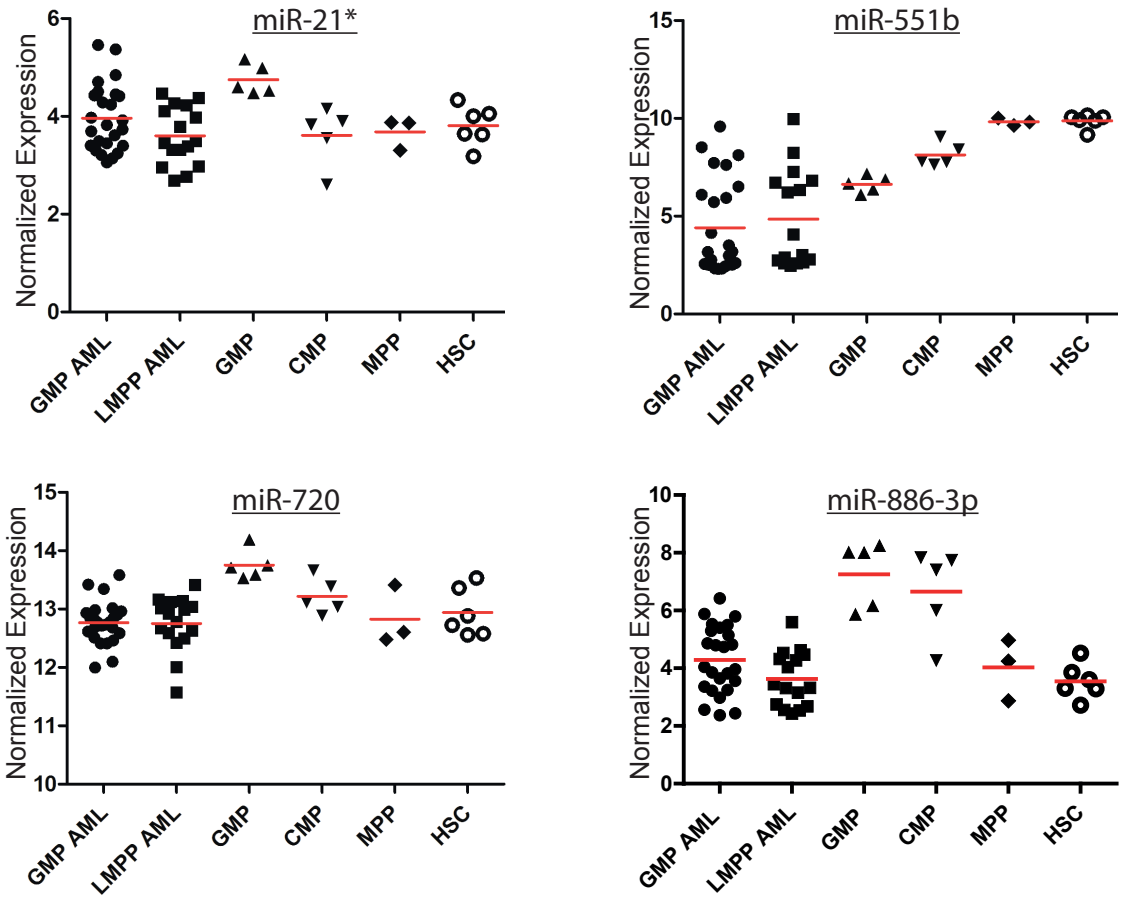


Fig 5.12 MicroRNA expression profiles of miRNAs significantly more expressed in normal GMP compared to GMP-like AML.
Normalized expression is plotted for each sample analysed on an array. Normalized expression value is log2 transformed. Red line represents population mean.

luster but miR-363 did not. However overall these data suggest there may be a common mechanism involved in the reduced expression of these miRNAs.

	miR-92a	miR-17	miR-20b	miR-363	miR-19a
miR-92a		0.76	0.66	0.4	0.7
miR-17	0.76		0.74	0.4	0.87
miR-20b	0.66	0.74		0.71	0.71
miR-363	0.4	0.4	0.71		0.4
miR-19a	0.7	0.87	0.71	0.41	

Table 5.6 miRNA correlation in AML
miRNAs found within clusters were subject to Pearson pairwise correlation analyses. Correlation analyses were performed between miRNAs comparing their expression in AML samples. The correlation coefficients from each analysis are shown.

5.7 miRNAs and mRNA targets

As miRNAs exert their function through binding to and repressing or down-regulating mRNAs, we investigated the mRNAs' targets of the differentially expressed miRNAs. Only a small number of miRNA target interactions have been confirmed by experimental investigation, so identifying real target interactions is difficult. However, databases exist of predicted target interactions, and complex algorithms are employed by these databases which use the microRNA seed sequence as well as other aspects of the miRNA, to predict potential mRNA targets. TargetScan is one of the highly accredited prediction databases which we used to give us lists of predicted target mRNAs of the differentially expressed miRNAs (TargetScan ref). However this generated hundreds of potential target mRNAs, this made it very difficult to interpret this data.

Previous work in our laboratory generated mRNA profiles from our populations (NG), we decided to investigated whether any of the targets identified from TargetScan were found differentially expressed on an mRNA level between GMP-like and GMP cells. Five of the GMP-like samples used on the miRNA array were also used on the mRNA array. This allowed us to compare data directly between the same AML samples. Unfortunately, none of the normal GMP samples used on the miRNA array were used on the mRNA array, but the variation between normal samples was much lower than observed in AML, therefore using data from different normal samples posed less problems than comparing data from different AML samples. A t-test comparing the miRNA expression levels between the five matching GMP-like AML samples and normal GMP samples was performed. The differentially expressed miRNAs was then input into TargetScan to generate a list of potential mRNA targets. A t-test was then performed comparing the mRNA expression levels between the five matching GMP-like samples and normal GMP samples. We then examined whether any of the miRNAs that were significantly over-expressed had mRNA targets that were significantly less expressed and vice versa. A bioinformatic approach was used to mine the data, to match up over-expressed miRNAs and less expressed mRNA targets (Figure 5.13), and to match up less expressed miRNAs and over-expressed mRNA targets (Figure 5.14). Multiple target matches were identified and can be seen as red dots on the plots.

The next question was whether the number of target interactions identified was significant. To investigate the significance, the same bioinformatic mining approach was taken, however rather than inputting a list of 100 significantly over-expressed or less expressed mRNAs, a list of a 100 random mRNAs was used as the input. The number of target matches identified using a random list of mRNAs was recorded, and this was then repeated 1000 times with random mRNAs. Histograms displaying the frequency of the number of target matches that occurred by chance using lists of

random mRNAs can be seen in Figures 5.15 (for over-expressed miRNAs) and Figure 5.16 (for less expressed miRNAs). The number of target matches using the top 100 significantly less expressed or over-expressed mRNAs as the input is seen on the histograms as a red line. The number of target matches using the top 100 significantly less expressed or over-expressed mRNAs with a fold change greater than 1.5 as the input is seen on the histogram as a blue line. The position of the red/blue lines on the histograms can be used to judge whether the number of target matches is significant. If the red/blue line is lying to the right of the histogram, the number of target matches is significantly more than would have been expected by chance. If the red/blue line is lying in the middle of the histogram, the number of target matches is what would have been expected by chance. If the red/blue line is lying to the left of the histogram, the number of target matches is less than what would have been expected by chance. Most of the over-expressed miRNAs had more mRNAs in target matches than would have been expected by chance, whereas most of the less expressed miRNAs had fewer mRNA target matches than would have been expected by chance.

It is therefore difficult to reach a conclusion on the results of the target matches. There are caveats with performing this kind of analysis. The first is that target prediction databases offer predicted targets and many of the targets are unlikely to be real. Many real targets may also not have been identified by prediction. The second caveat is that although the samples for the mRNA and miRNA arrays were from the same patient sample, the populations were sorted on different days by different people, and different RNA extraction methods were also used. The third caveat is that many different alterations occur in cancer cells, so even if an miRNA is over-expressed and its target is less expressed, this does not mean the alteration in the mRNA level was caused by an alteration in the miRNA level. The fourth caveat is that there is a lot of noise in any system such as this, and so it is difficult to identify real effects from noise.

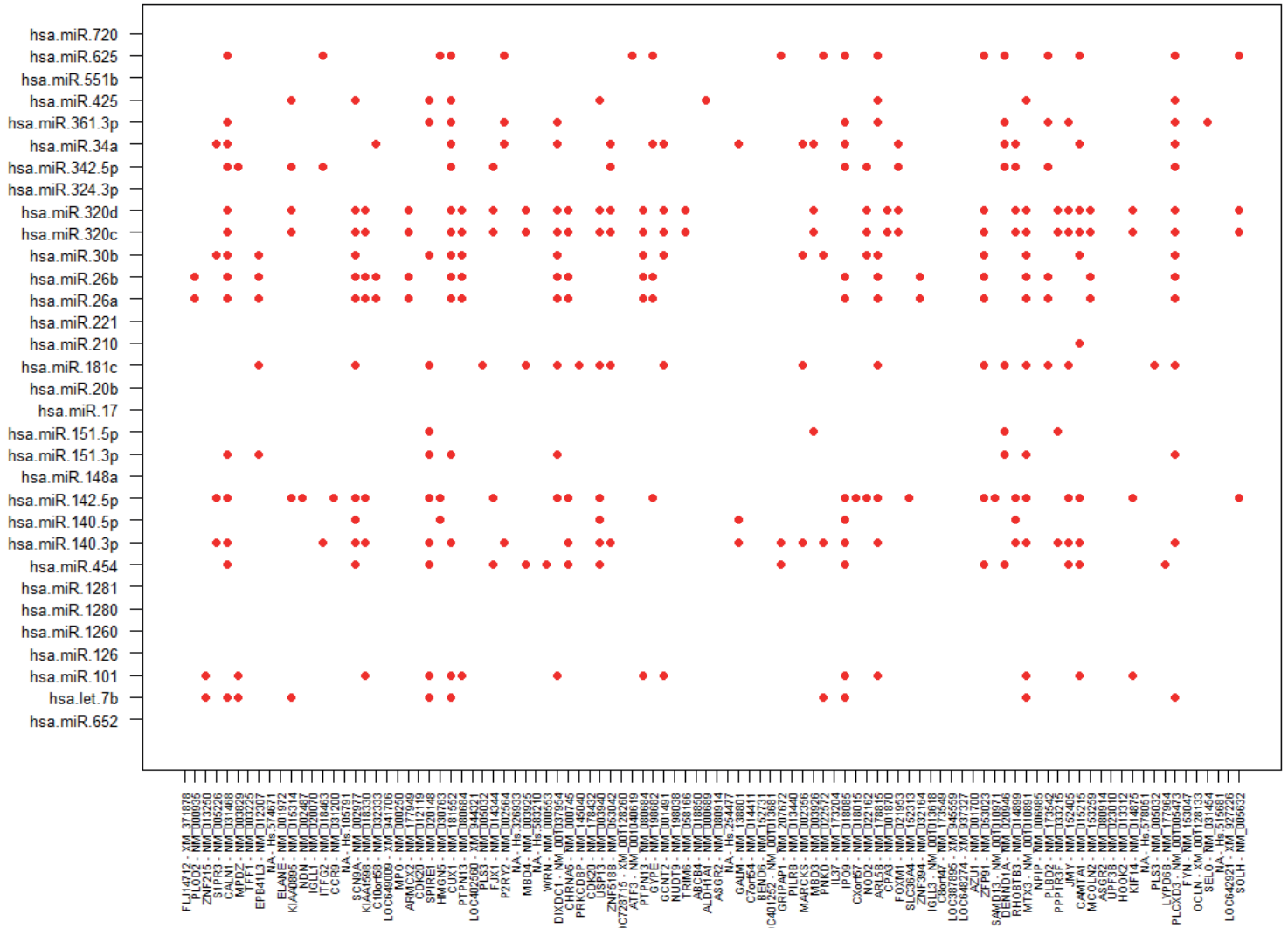


Fig 5.13 miRNA and target mRNA interactions; miRNA more expressed in GMP-like AML compared to GMP.

5 GMP-like AML samples were present on both the miRNA and the mRNA microarrays. A t-test was done to determine miRNAs/genes differentially expressed between these 5 GMP-like AML samples and the normal GMP control samples. A red dot represents a microRNA that is more expressed in GMP-like AML and a potential mRNA target that is less expressed in GMP-like AML. (all over-expressed miRNAs with significance of P-value <0.05 were used). (Top 100 signifcantly less expressed mRNAs were used). Potential target interactions were determined from TargetScan database.

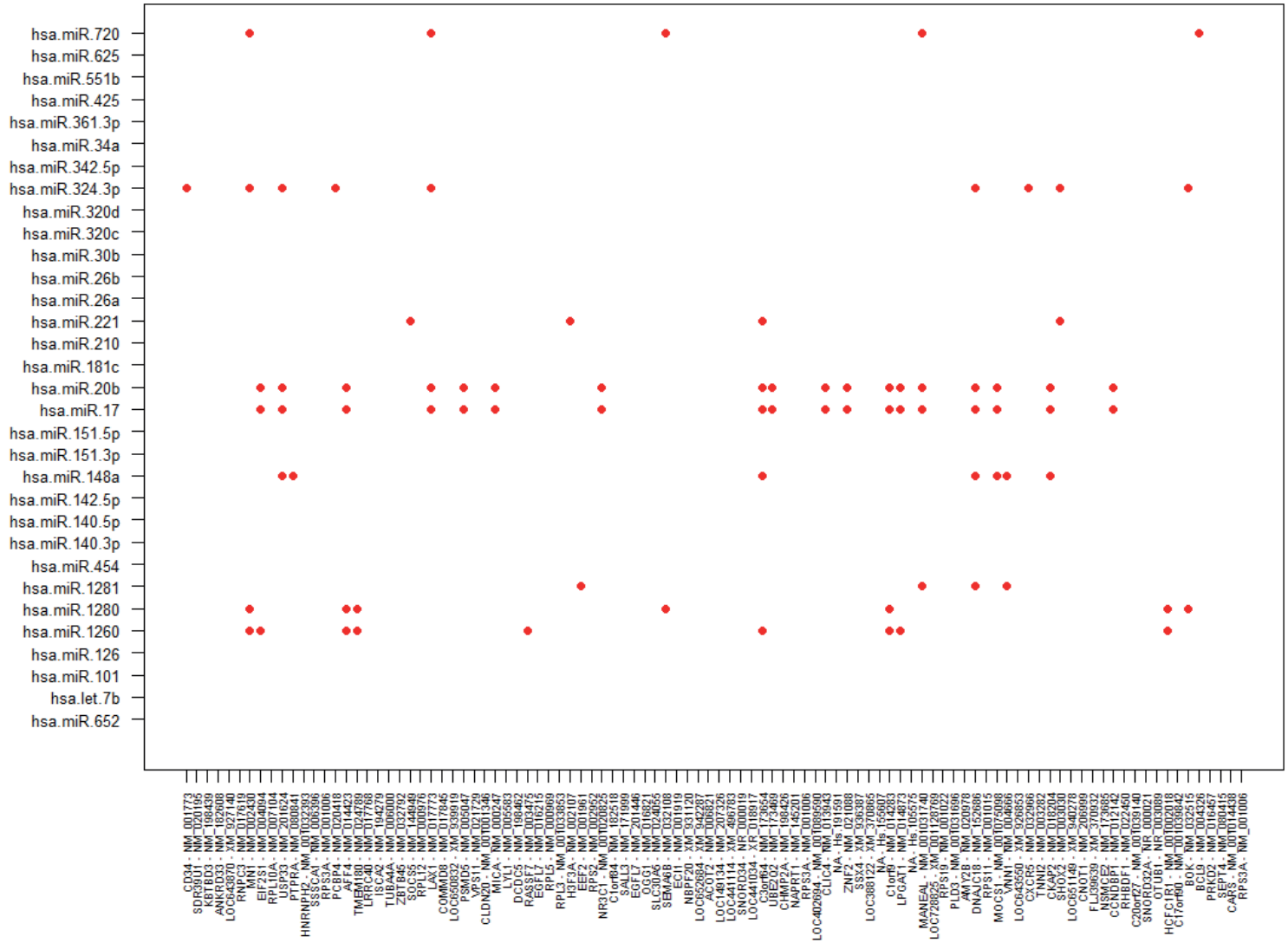


Fig 5.14 miRNA and target mRNA interactions; miRNA more expressed in GMP compared to GMP-like AML.
5 GMP-like AML samples were present on both the miRNA and the mRNA microarrays. A t-test was done to determine miRNAs/genes differentially expressed between these 5 GMP-like AML samples and the normal GMP control samples. A red dot represents a microRNA that is more expressed in GMP and a potential mRNA target that is less expressed in GMP. (all over-expressed miRNAs with significance of P-value <0.05 were used). (Top 100 signifcantly less expressed mRNAs were used). Potential target interactions were determined from TargetScan database.

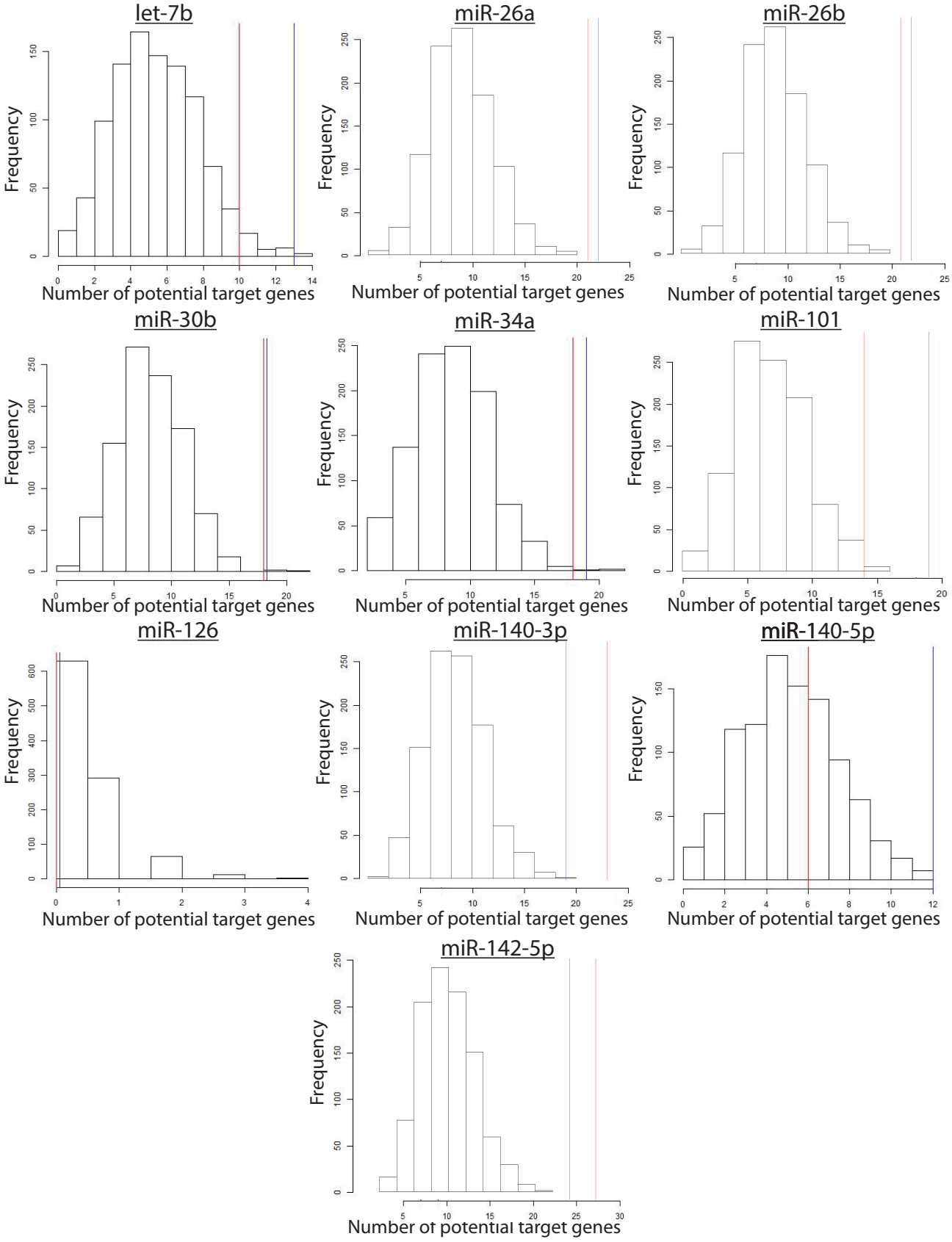


Fig5.15 miRNA and target mRNA interaction is it significant? miRNAs more expressed in GMP-like AML compared to normal GMP

The frequency of finding potential miRNA targets in a list of 100 random mRNAs is displayed as a histogram for miRNAs more expressed in GMP-like AML. The number of potential target matches using a list of the top 100 significantly less expressed mRNAs in GMP-like AML compared to GMP is seen on the histograms as a red line. The number of potential target matches using the top 100 significantly less expressed mRNAs in GMP-like AML compared to GMP with a fold change greater than 1.5 is seen on the histogram as a blue line.

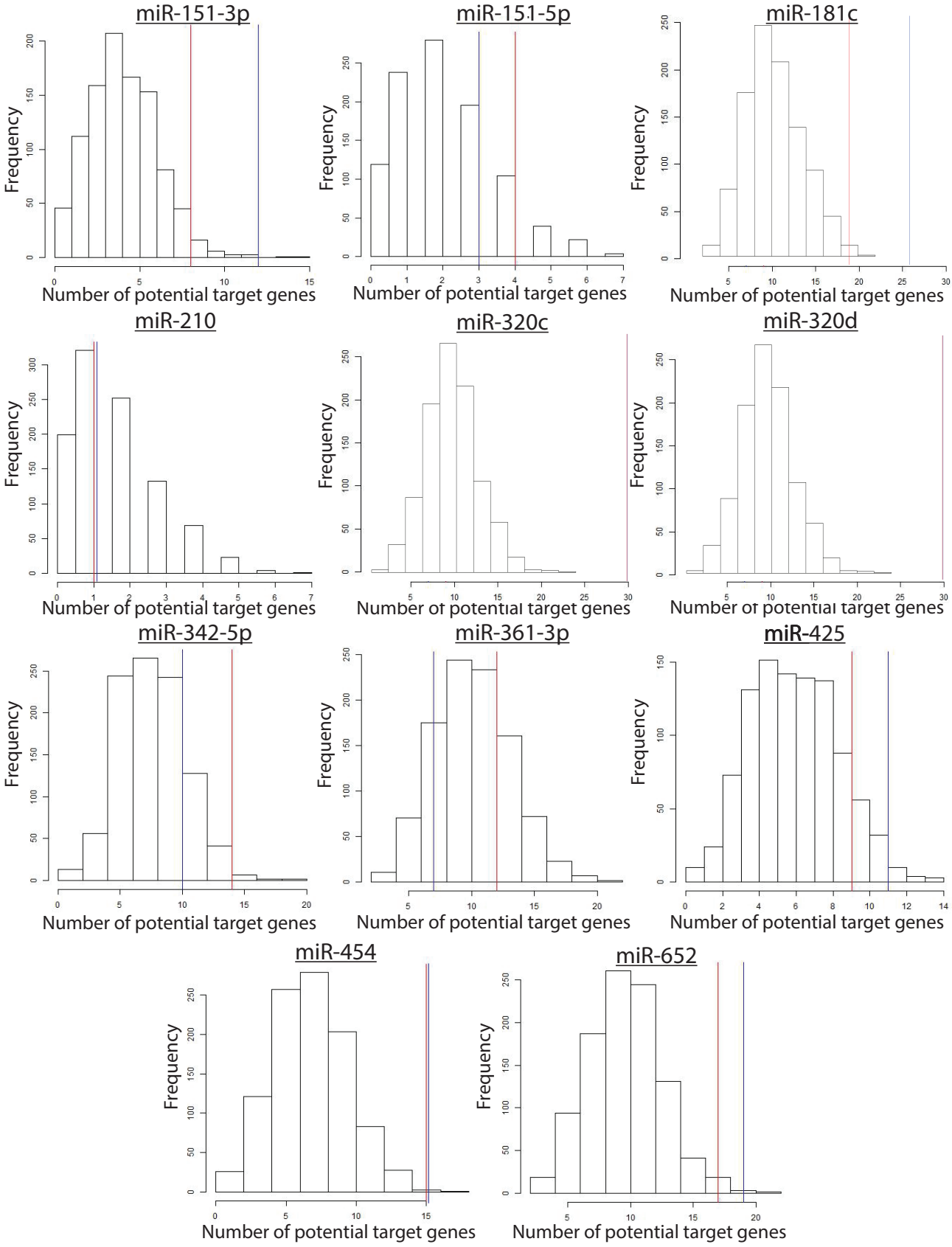


Fig5.15 miRNA and target mRNA interaction is it significant? miRNAs more expressed in GMP-like AML compared to normal GMP

The frequency of finding potential miRNA targets in a list of 100 random mRNAs is displayed as a histogram for miRNAs more expressed in GMP-like AML. The number of potential target matches using a list of the top 100 significantly less expressed mRNAs in GMP-like AML compared to GMP is seen on the histograms as a red line. The number of potential target matches using the top 100 significantly less expressed mRNAs in GMP-like AML compared to GMP with a fold change greater than 1.5 is seen on the histogram as a blue line.

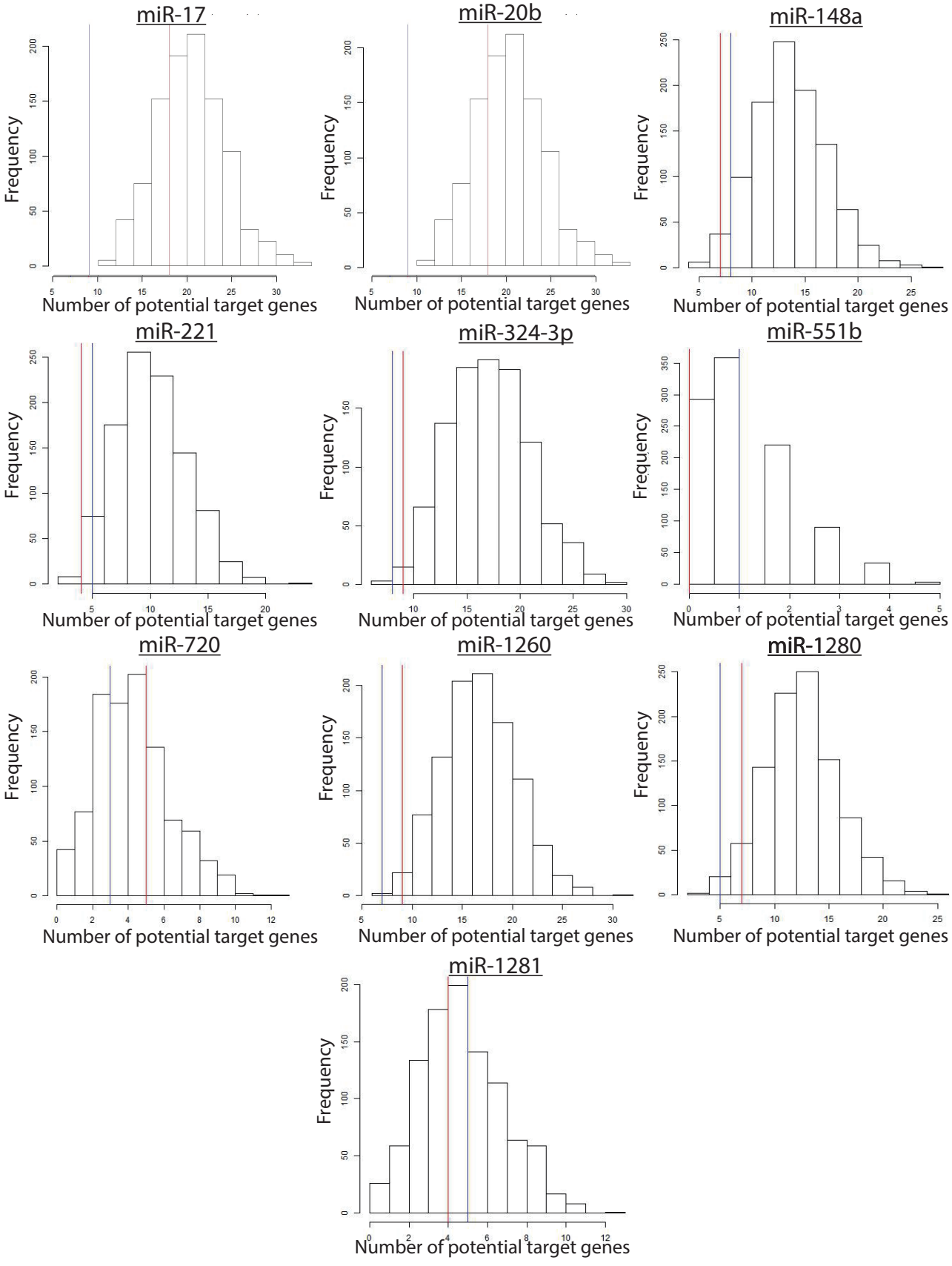


Fig5.16 miRNA and target mRNA interaction is it significant? miRNAs less expressed in GMP-like AML compared to normal GMP

The frequency of finding potential miRNA targets in a list of 100 random mRNAs is displayed as a histogram for miRNAs less expressed in GMP-like AML. The number of potential target matches using a list of the top 100 significantly more expressed mRNAs in GMP-like AML compared to GMP is seen on the histograms as a red line. The number of potential target matches using the top 100 significantly more expressed mRNAs in GMP-like AML compared to GMP with a fold change greater than 1.5 is seen on the histogram as a blue line.

(4) Array support :2529710112				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_1	BIR091	LMPP	A	4	273	68	7.8	85.5	5.81	4.23	1	2.25	6.71	0	7.27	3466	83	75
1_2	BOU014	GMP	A	10	1168	117	9.1	100.7	3.33	4.27	1	2.53	6.55	0.02	4.76	3299	82	73
1_3	OX291	HSC	N	0.7	95	136	8.9	83.3	3.26	4.28	1	2.77	6.45	0	5.91	1737	84	73
1_4	OX208	LMPP	A	20	1216	61	8.4	101.3	3.17	4.22	1	3.12	6.71	0	7.02	3580	87	74
2_1	OX533	MPP	N	2.8	322	115	8.4	100.5	3.29	4.26	1	3.23	6.46	0	4.76	3221	89	78
2_2	OX598	CMP	N	1.3	510	393	8	102.1	3.02	4.19	1	2.94	6.02	0	6.04	673	88	77
2_3	BIR091	GMP	A	3.3	640	194	7.5	100	3.26	4.23	1	2.79	6.79	0	4.29	3919	85	76
2_4	GLA001	GMP	A	2.7	277	103	8.1	98.9	3.07	4.1	1	3.04	7	0	10.5	1128	86	78
(5)Array support:2529710113				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_2	BIR141	GMP	A	1.4	238	170	7.2	99.3	3.32	4.01	1	2.55	6.26	0	6.04	1430	83	74
1_3	AYL004B	GMP	A	5	331	66	7.7	100.4	3.3	4.03	1	2.8	6.81	0	6.22	3706	83	76
2_1	OX476	GMP	A	7.1	1024	144	9.3	100.4	3.22	3.95	1	2.59	6.58	0.01	9.49	2823	84	75
2_2	BOU014	LMPP	A	2.5	135	54	9.1	84.5	3.19	3.88	1	2.65	6.51	0.04	6.57	1048	83	74
2_3	OX288	MPP	N	1.8	179	100	9	98.6	3.18	3.92	1	2.62	6.84	0.01	7.38	2310	78	70
2_4	OX204	CMP	N	1	125	125	8.2	78	2.86	3.84	1	6.79	6.75	0	6.21	948	111	93
(6)Array support:2529710114				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
2_4	BRI080	LMPP	A	1.9	144	76	7.8	89.9	2.87	3.71	1	2.85	6.98	0.01	8.58	1482	78	71

(7)Array support:2529710115				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_1	OX586	GMP	N	2.2	581	264	7.6	100.1	3.48	4.34	1	11.49	7.92	0	6.38	1349	119	96
1_3	BIR265	GMP	A	2.9	688	237	8.5	101.2	3.47	4.35	1	2.71	6.33	0	6.55	2288	94	85
2_1	BIR269	LMPP	A	1.1	83	75	8.1	72.3	3.44	4.32	1	2.73	7.16	0.03	7.02	2121	90	81
2_2	BIR312	GMP	A	0.6	162	270	7	80.8	3.43	4.23	1	2.47	5.93	0	6.4	1266	96	85
2_3	SMD003	LMPP	A	2	165	83	8.3	82.4	3.43	4.25	1	2.47	6.84	0	6.95	1985	92	82
2_4	OX561	GMP	A	5	650	130	8.7	101.5	3.54	4.3	1	2.46	7.12	0	6.15	1607	93	85
(8)Array support:2529710116				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_1	BIR277	GMP	A	2.5	490	196	9.1	102	3.28	4.14	1	11.19	8.65	0.03	14.03	1815	107	85
1_2	BRI019	LMPP	A	1.5	186	124	n/a	92.8	3.4	4.22	1	6.65	7.96	0.01	6.81	4469	96	80
2_1	OX288	HSC	N	1.8	189	105	9.1	94.4	3.11	4.14	1	3.12	6.8	0.02	7.25	762	93	84
2_2	OX598	CMP	N	1.3	510	393	8	102.1	3.42	4.16	1	11.44	7.31	0	5.53	1634	114	91
2_3	BIR257	LMPP	A	1.1	315	286	8.1	98.5	3.41	4.23	1	15.78	7.21	0.01	4.44	3238	119	89
2_4	OX325	LMPP	A	4	632	158	9	98.8	3.32	4.24	1	5.23	7.35	0	5.74	2054	95	81
(9) Array support:2529710117				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_1	OX325	GMP	A	2.5	344	138	9.4	101.2	3.34	4.06	1	2.84	7.17	0.21	14.35	3113	88	76
1_3	OX533	CMP	N	3.6	277	76.9	NA	98.9	3.31	4.17	1	3.49	6.29	0.01	6.11	1412	91	81
1_4	LEE060	GMP	A	1.3	238	183	7.2	99.3	3.37	4.18	1	5.96	7.04	0.33	6.67	2002	96	81
2_3	BRI023	LMPP	A	0.8	111	139	8.5	84	3.45	4.28	1	7.31	6.46	0	5.43	2572	97	84
2_4	OX555	HSC	N	NA	88	NA	7.1	69.5	3.35	4.1	1	2.39	6.34	0	6.37	1386	94	85

(10)Array support:2529710118				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_1	OX291	GMP	N	2.9	816	281	9	99.5	3.34	2.27	1	2.94	6.04	0	12.34	1326	75	66
1_2	SW031	LMPP	A	0.5	75.6	151	7.8	61.4	3.43	2.4	1	2.58	6.17	0.01	7.59	1649	77	70
1_3	HER003	GMP	A	1.9	408	214	7.9	102	3.25	2.31	1	2.6	6.71	0.02	7.74	2135	74	67
2_1	OX709	HSC	N	1.7	264	155	8.9	66	3.23	2.32	1	2.53	6.9	0.21	11.61	556	79	71
2_2	709 AND OX	MPP	N	1.0 & 0.2	70 & 41	73 & 178	8.5 & 8.9	80.1	3.36	2.4	1	2.84	6.96	0.06	7.01	1302	81	71
2_3	OX497	GMP	N	3.3	1312	394	7.6	100.9	3.18	2.32	1	2.46	6.5	0.01	7.29	1468	74	66
2_4	BRI024	LMPP	A	6	440	73	8.9	100	3.12	2.26	1	2.1	6.68	0.02	6.6	2093	78	71
(11) Array support:2529710119				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_1	OX709	CMP	N	4.7	1760	374	9	100	3.19	2.25	1	2.55	6.31	0	14.5	1062	79	69
1_2	BIR305	GMP	A	2.1	323	154	8.4	101	3.41	2.27	1	3.06	6.86	0	7.55	1902	77	69
1_3	497 AND OX	HSC	N	1.2 & 1	176 & 70.4	147 & 70.4	7.6 & 7.1	84	3.41	2.36	1	2.52	6.36	0.01	6.43	1384	77	70
2_2	OX288	GMP	N	2.4	310	129	9	100.1	3.35	2.42	1	2.5	6.49	0.04	7.99	1617	80	72
2_3	BIR272	LMPP	A	3	213	71	7.5	60	3.18	2.35	1	2.58	6.16	0.01	7.89	713	77	65
2_4	564 AND OX	HSC	N	0.5 &1.8	44 & 189	88 & 105	7.4+9.1	95	3.12	2.37	1	2.29	6.41	0.02	6.68	691	78	69
(3) Array support:252929710120				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_1	BOU002	LMPP	A	4.8	752	157	8.8	98.9	2.97	4.61	1	3.78	7.32	0.01	12.46	3186	102	86
1_2	OX288	CMP	N	3	640	213		100	3.2	4.63	1	3.06	6.23	0.01	7.13	1898	110	98
1_4	OX195	GMP	A	1.4	216	154	8.2	98.2	2.86	4.87	1	3.6	7.23	0	7.98	3207	109	84
2_1	OX512	GMP	A	3.2	491	153	9.1	104	2.96	4.57	1	3.42	6.95	0	13.22	2351	103	91
2_2	BOU002	GMP	A	2.2	400	182	6.8	100.5	3	4.63	1	3.11	6.74	0	8.01	2499	99	89
2_3	OX195	LMPP	A	8.5	832	98	8.5	100	2.96	5.99	1	3.45	6.14	0	8.43	2670	108	96
2_4	OX709	GMP	N	4.2	1507	359		98.8	3.04	5.91	1	3.47	6.23	0	7.78	2093	121	105

(2) Array support:2529297101121				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_2	BRI080	GMP	A	4.6	518	113	8.8	99.7	3.13	4.14	1	2.27	6.36	0	6.29	2825	79	72
1_4	BIR295	GMP	A	7.3	688	94	9	101.2	3.37	4.13	1	2.55	6.35	0	7.84	3016	79	71
2_1	BIR031	GMP	A	2.3	348.8	152	7.7	99.7	3.4	4.15	1	3.13	6.83	0.01	11.29	2115	84	74
(12)Array support:252929710122				RNA info					array info									
Array position	sample	population	AML/ Normal	Number of cells RNA extracted from ($x10^5$)	total RNA extracted (ng)	RNA per 10^5 cells (ng)	RIN	amount RNA of labelled (ng)	#1 Lab spike-in	#2 hyb spike-in	#3 is good grid	#4 back-ground	#5 % population outliers	#6 % non uniform outliers	#7 %CV for replicate probes	#8 99% of sig	#9 50% of sig	#10 1% of sig
1_2	BIR075	GMP	A	3.5	541	155	8.6	100	3.16	4.15	1	2.73	6.64	0.01	8.82	2301	88	78
1_3	BIR073	GMP	A	30	1424	47	9	101.7	3.09	4.08	1	2.65	6.74	0	9.11	2808	90	80
1_4	BIR031	LMPP	A	10	766.4	77	9.3	101.1	2.86	4.09	1	3.06	6.84	0.05	8.45	1231	87	77
2_1	OX208	GMP	A	20	2880	144	6.5	100	3.02	4.24	1	2.69	6.51	0.01	7	1408	94	83
2_2	BIR269	GMP	A	4.2	299	72	8.1	99.7	2.96	4.17	1	2.94	5.93	0	8.35	1479	87	76
2_4	BIR257	GMP	A	7.9	409.6	52	7.6	102.4	3.03	4.18	1	3.04	6.57	0.03	8.24	2749	101	89

Table 5.3 Sample and quality information for microarrays.

Only arrays where data was used in further analysis are shown

5.8 Discussion III

The aim of this section of work was to use the Agilent microarray system to generate miRNA expression profiles of populations of interest. Prior to running samples of interest a pilot experiment was completed. This demonstrated that the system as working well, and a high correlation was seen between array repeats of the same sample. Although there was an initial problem with the background on the pilot array, this was eliminated on further arrays, by using a new batch of wash buffers. This demonstrated how crucial it was that the reagents were working efficiently. They only samples in the pilot array that showed poor correlation, were samples with labelling issues. Labelling issues occurred in a few of the arrays on sorted samples; however the data from these arrays was not used in further analysis. It is not clear why the labelling reaction was inefficient in certain samples; however it did not appear to be connected to the sample quality as sample repeats did not yield the same problem. The issue was also not due to reagents as mastermixes were prepared and added to batches of samples and other samples from the same batch were not affected. The most likely cause of inefficient labelling is the sample not being fully denatured prior to the labelling reaction.

The extensive QC report produced on arrays allowed us to exclude any arrays with suboptimal performance, so only high quality data was were used in further analysis. As the RNA loaded onto arrays varied slightly in amount and integrity. I also compared the different QC results, to RNA amount and integrity and found that there was no correlation between this and array output. This indicated amount or integrity of RNA was not causing any bias on the array.

After the array data had been produced, the data had to be normalized to allow for comparison between different arrays. This was done by the bioinformatician in our

laboratory, Emanuele Marchi. Emanuele tried multiple different normalization methods and found that the one used produced the best overall normalization. This method had been developed specifically for looking at miRNA expression using the Agilent system [121]. A common problem with array data can be that arrays cluster by array support, indicating a level of bias based on sample run However after normalization arrays did not cluster by array support, indicating the array supports/batch was not causing artificial bias in the data.

The microRNA expression profiles were then analysed, this revealed that the cytogenetic subtype of the sample was playing a role on how the samples clustered. This effect has previous been seen by many other groups looking at miRNA expression in AML. Normal samples clustered by population type whereas AML samples clustered by patient. Demonstrating that patient specific mutations have more of an effect on the miRNA profiles than population type, a phenomena as observed at the mRNA level

The relationship of the different populations to each other, was similar on at miRNA level as it was on an mRNA level. However less miRNAs were differentially expressed than mRNAs. However the reason for this difference can most likely be explained by difference in the in the number of miRNAs profiled ~900 miRNAs compared to the number of mRNAs profiled in the magnitude of thousands. The only obvious difference between the miRNA expression pattern and the mRNA expression pattern, was that CMP and GMP populations areor much closer on an miRNA level than they are on a mRNA level, it may be that miRNAs are less important in this transition or that only a few key miRNAs are needed this transition.

Like on mRNA level, at a miRNA level the LSC populations more closely resemble more mature progenitor populations than HSC and MPP populations. However the

LSC populations display an aberrant miRNA stem cell-like signature. This has previous been reported for LSC populations for mRNAs [57].

We wanted to explore what miRNAs were differentially expressed in AML. As GMP-like cells appeared to be most similar to GMP cells on a miRNA and mRNA level and shared an immunophenotype. A t-test was used to determine if any of the microRNAs were differentially expressed between the two populations, to highlight which miRNAs were differentially expressed in AML Twenty five miRNAs in total were identified to be differentially expressed between GMP-like AML cells and normal GMP cells, indicating that miRNAs are aberrantly expressed in AML. Of the overexpressed miRNAs there were two miRNA families that were overexpressed; the miR-26 family and the miR-30 family. It is interesting that more than one member of the same family is overexpressed as they will recognise the same mRNA targets. The miR-26 family has been implicated in cancer however is found to be overexpressed in some cancers but also found to be less expressed in other cancers [138-141].

In miRNAs that were less expressed, three members of the miR-17-92 polycistronic cluster were found to have reduced expression in AML compared to normal cells as well as two members of the miR-106a-363 polycistronic cluster. miRNAs from polycistronic clusters are transcribed as one transcript. So it is not surprising that if the expression of one of these miRNAs is reduced, that other members of the clusters are also reduced. However the miR-106a-363 cluster and the miR-17-92 cluster are paralogous clusters and the miRNAs within the clusters target many of the same mRNAs. For example miR-92 and miR-363 have identical seed sequences so are likely to target many of the same miRNAs [142]. As both clusters are downregulated this suggests a common mechanism of action of the clusters and highlights that they may be playing a role in AML pathogenesis.

We also compared miRNAs and their targets mRNAs, using our miRNA expression profiling data set and the previous generated mRNA expression profiling data set. The data suggested that more downregulated mRNAs than would have been expected by chance were targets of miRNAs that were overexpressed in AML. Whereas mRNA targets of overexpressed miRNAs were less frequently found downregulated than would have been expected by chance. It is important not to assume that only the overexpressed miRNAs are having an effect in AML as there are many caveats with this type of analysis. Also the analysis uses predicted target interactions rather than proven biological targets interactions. It is likely many of the predicted targets are not real targets under normal biological conditions. It may be that overexpressed miRNAs are reaching much higher levels and are therefore targeting mRNAs that they would not normally target at normal physiological levels hence more targets being identified than by chance. Whereas the miRNAs with reduced expression will only have normal mRNA targets going up hence why less targets being identified than by chance the number of targets was identified than by chance.

Chapter 6: Results IV qPCR validation

In the previous chapter a set of miRNAs that were significantly differentially expressed between GMP-like AML cells and normal GMP cells were identified. The aim for the following section of work was to validate the differential expression of these miRNAs. The two commonly used methods to validate miRNA expression are northern blotting and RT-qPCR. As the amount of RNA material available for validation was limited, northern blotting was not a viable option as it requires a large amounts of RNA (microgram quantities). Very low amounts of material can be used with RT-qPCR (down to picograms), therefore RT-qPCR was selected to validate miRNA expression. RT-qPCR also works via a different method to arrays; it is useful when validating data from one system to choose a system that works via an alternative methodology.

Mature miRNAs do not contain polyA tails and are much shorter than mRNAs, therefore the standard RT-qPCR process has to be adapted to quantify miRNA expression. There are two different RT-qPCR technologies available to quantitate miRNA expression; the Taqman system and the SYBR green system.

The SYBR green system adds polyA tails to miRNAs then uses OligoT primers to reverse transcribe RNA to cDNA. Specific primers are then used during qPCR to amplify cDNA of the miRNA of interest. The SYBR green dye present in the qPCR reaction binds the double stranded DNA product and fluoresces, the level of fluorescence is measured to estimate the starting amount of miRNA.

The Taqman system uses specific primers during reverse transcription and so only creates cDNA of miRNAs of interest. Specific and universal primers are then used during qPCR to amplify cDNA of target miRNA. A probe which only binds target miRNA is also included in the qPCR reaction, and attached to the probe is a dye and a

quencher. The polymerase used in the qPCR reaction has 5'-3' exonuclease function, this means as the polymerase moves along the cDNA it digests the probe, this separates the dye from the quencher, allowing the dye to fluoresce, the level of fluorescence is then used to estimate the starting amount of miRNA. The advantages and the disadvantages of the two RT-qPCR systems can be seen in Table 6.1.

	Taqman	SYBR green
advantages	• Specific primers used during RT, adds another layer of specificity to system. • Can discriminate between the precursor miRNA and mature RNA. • Quantitates specific PCR product, rather than binding double stranded DNA. • Less RNA needed than SYBR green. • Due to system design a small amount of genomic DNA contamination should not interfere with quantification.	• All miRNAs converted to cDNA during RT, so cDNA can then be used to explore expression of any miRNA. • Reagents cheaper than Taqman.
Disadvantages	• Specific RT primers are used so cDNA can only be used to explore miRNAs that had RT primers included in RT mixture. • Regents more expensive than SYBR green.	• miRNAs may not be polyadenylated equally, increases bias in system • Cannot discriminate between precursor and mature miRNA, therefore expression level is the combined level of precursor and mature miRNA expression. • Quantifies double stranded DNA rather than specific PCR product, meaning things like primer dimers will also create fluorescent signal. • More RNA needed than Taqman. • Presence of a small amount of contaminating genomic DNA will interfere with results.

Table 6.1 Taqman System versus the SYBR green system; advantages and disadvantages.

The microarray quantified the expression of mature miRNAs, therefore the system used to validate the miRNA data must only quantify the expression of mature miRNAs, for this reason I selected the TaqMan system for validation.

6.1 Taqman RT-qPCR

To reverse transcribe miRNAs into cDNA the TaqMan system uses specific hairpin primers (Figure 6.1 B). 8 nucleotides (nt) within the RT primer specifically bind the target miRNA, the rest of the primer makes up the hairpin, this structure ensures that

the primer only binds the mature miRNA and not the precursor or primary miRNA. The resulting cDNA product is 58nt long, 36nt longer than the miRNA template.

In the qPCR reaction the forward primer binds to 16nt within the cDNA that are specific to the miRNA. The reverse primer is a universal primer that binds the RT primer section of the cDNA. The probe binds to 8nt that are specific to the miRNA within the cDNA and 10nt that are specific to the RT primer within the cDNA; this ensures the probe only binds to nucleic acid that has been amplified by RT primers (Figure 6.1 D). The forward primer adds a further 14nt to the PCR product, so that the final amplified PCR product is 72nt. Between the probe and the forward primer the whole of the miRNA is used for recognition adding extra specificity.

Prior to qPCR a preamplification step can be introduced, preamplification is often required when multiple targets are assayed from the same cDNA sample, and is required if the Fluidigm system is used. The same primers used in the qPCR reaction are used in the preamplification reaction. The RT and preamplification steps can be conducted as multiplex reactions, the qPCR reactions are always run as singleplex reactions.

I chose to use the Fluidigm system with TaqMan assays, this involved using multiplex primers during reverse transcription, then conducting a preamplification reaction also using multiplex primers. The resulting amplified cDNA was then loaded on to a Fluidigm chip which uses microfluidics to run upto 48 assays against 48 samples in a single qPCR experiment, allowing up to 2304 qPCR reactions to be run at a time. The Fluidigm system provide us with three advantages, it reduces the amount of RNA needed, reduced amount of reagents used and was a more time efficient system.

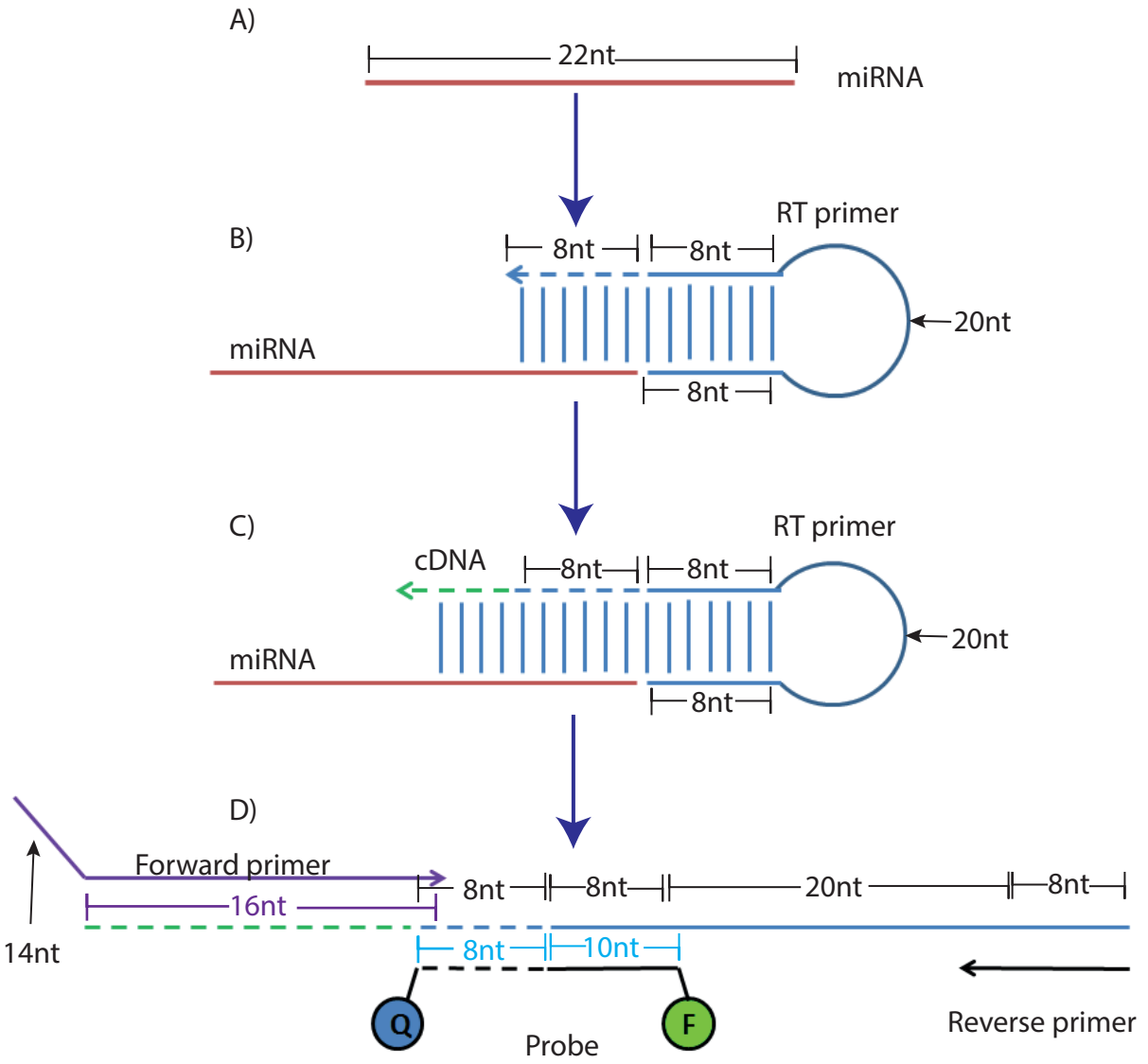


Fig 6.1. Schematic depicting miRNA RT-qPCR process.

A) microRNA before reverse transcription. B) Reverse transcription primer binding. C) Reverse transcription. D) qPCR primers and probe binding. (microRNA is depicted in red, RT primer is depicted in blue, cDNA is depicted in green, forward primer is depicted in purple and probe and reverse primer are depicted in black. The dotted line on cDNA and RT primer represent complementarity to microRNA. The dotted line on the probe represents where probe binds to the microRNA specific part of the cDNA and the straight line represents where probe binds to the RT-primer section of the cDNA.)

6.2 Selecting endogenous controls.

The same amount of RNA should be added to each RT reaction however slight inconsistencies in pipetting and RNA quantification may occur. To normalize for this and for varying RT efficiencies, endogenous controls are used. Endogenous controls should be expressed at similar levels across all samples. It is important when selecting endogenous controls for microRNAs, that the control is of similar size and stability to miRNAs, and that the assay used to amplify the control can be designed in the same format as the miRNA assays. For these reasons small nuclear RNAs (snRNA) or small nucleolar RNAs (snoRNA) are chosen as endogenous controls for miRNA expression experiments.

I conducted a literature search to identify what endogenous controls were routinely used in miRNA expression profiling of AML and Normal BM. This search provided me with a list of potential endogenous controls (RNU6, RNU6B, RNU24, RNU44, RNU48, RNU66).

Initially I selected RNU6 and RNU44 as endogenous controls. A selection of samples was run against the different assays on a Fluidigm chip to investigate the setup and reagents. The samples run on the chip were amplified cDNA from sorted normal GMP and GMP-like AML samples. No RT controls were included; these were samples that went through the same process as amplified cDNA except no reverse transcriptase was added during the RT reaction. These controls were present to test if any of the reagents amplified genomic DNA. In all samples tested there was amplification of both the endogenous controls. There was no amplification in the no RT controls for RNU44 however there was amplification in the no RT controls for RNU6. Therefore it was likely

that RNU6 was amplifying genomic DNA, so RNU6 was excluded as a potential endogenous control (Figure 6.2A).

As RNU6 was not suitable assay to use, more potential endogenous controls were tested. To ensure that the assays worked efficiently and to investigate whether they amplified genomic DNA. Freshly isolated RNA from a normal GMP sample was reverse transcribed using multiplex primers. The cDNA was then pre-amplified with multiplex primers. Primers from all of the non-control assays were included in multiplex reactions to ensure none of the selected assays would inhibit or interfere with the endogenous controls. Three qPCR reactions were carried out for each of the control assays using undiluted cDNA, 1/10 diluted cDNA and 1/100 diluted cDNA. A no RT control was also carried out for each control assay (Figure 6.2B). RNU66 had a high level of amplification in the no RT control and so it was excluded as a potential control (Figure 6.2B iv). There was amplification in the no RT control of RNU48 (Figure 6.2B v). However the level of amplification was low and the no RT control amplification curve crossed the threshold many cycles after the cDNA amplification curves; therefore RNU48 was not excluded as a potential control. There was no amplification of the no RT controls for any of the other assays (figure 6.2 B i, ii, v). The CT value at which each control assays crossed the threshold was plotted (Figure 6.2 C). RNU24, RNU44 and RNU48 all lay as parallel lines whereas RNU 6B lay at an alternative gradient, this suggested that the amplification efficiency was equivalent for RNU24, RNU44 and RNU48. However RNU6B did not show comparable amplification efficiency to the other controls and was excluded as a potential control. If a dilution factor of 10 is used there should be 3-4 cycles between each dilution at the threshold, this was the case for the dilutions of assays RNU24, RNU44 and RNU48. The cycles at which RNU24, RNU44 and RNU48 crossed the threshold was not too high or too low between 15-20 cycles. For these reason RNU24, RNU44 and RNU48 were selected to be used as the normalization controls.

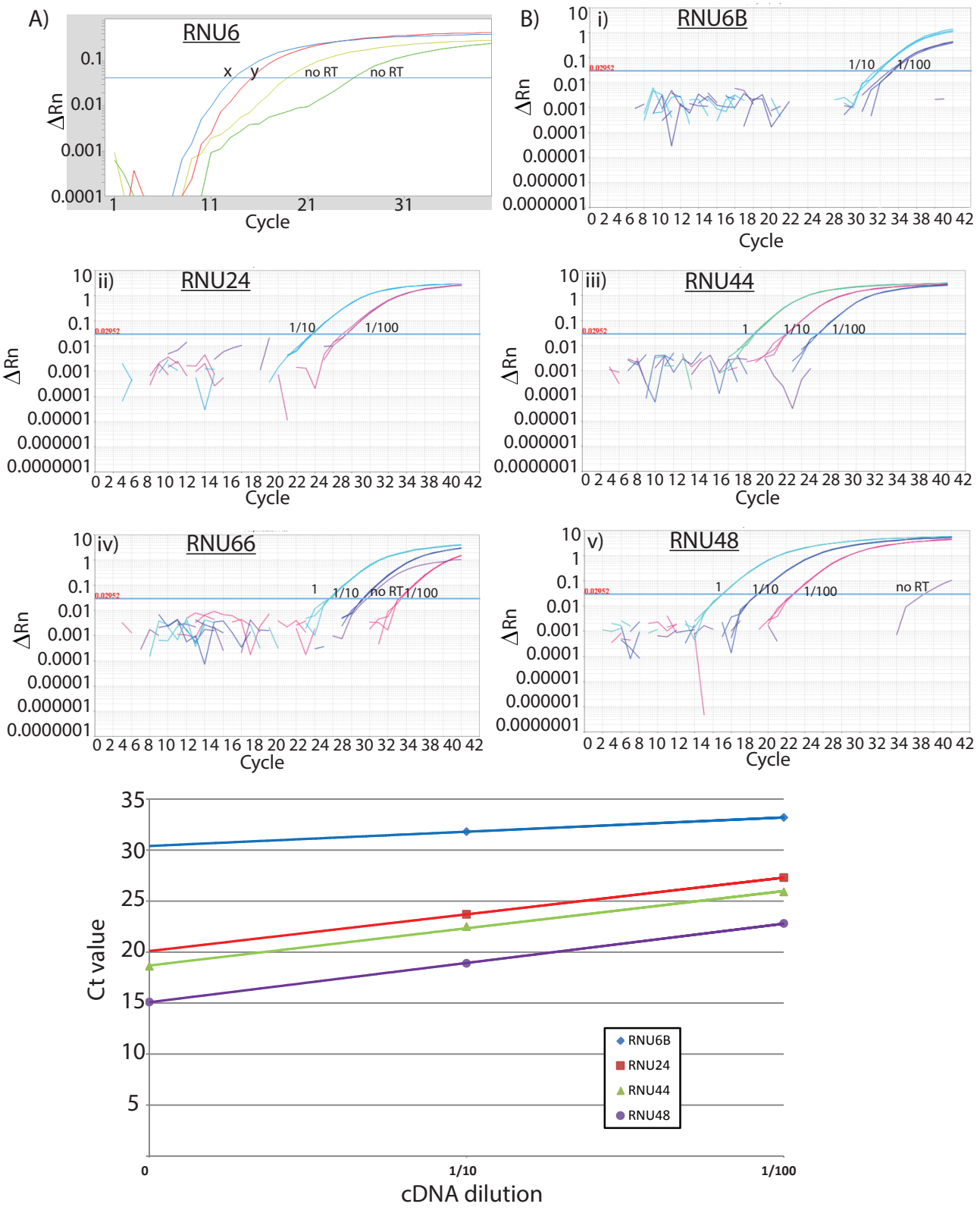


Fig 6.2 Endogenous controls for miRNA RT-qPCR.
Several different snRNAs/snoRNAs were preselected from the literature to be used as endogenous controls for qPCR. A) An example of RNU6 amplification from qPCR run on Fluidigm real time machine, X and Y are RNA samples that had been through multiplex RT and multiplex preamp reactions before qPCR. No-RT are controls RNA samples that went through the same process but no reverse transcriptase was added to reaction mix. B and C) Examples of further endogenous controls amplification from qPCRs run on ABI 7500 real time machine. RNA samples went through multiplex RT and a multiplex preamp reactions before qPCR. No-RT are controls RNA samples that went through the same process but no reverse transcriptase was added to reaction mix. Post amplification cDNA was added qPCR reactions either undiluted, in a 1/10 dilution or in a 1/100 dilution (amplification curves are labelled to indicate which dilution of cDNA was used). Each qPCR was done in triplicate. Where no line can be seen for no-RT there was no amplification. B) Amplification plots for each of the controls. C) The threshold Ct value was plotted for each control and dilution. (mean of triplicates is plotted) (no error bars are shown as stnadrad deviation is too low to be displayed.)

6.3 Validation of qPCR system

I next set out to test the non-control assays and validate the different steps of the qPCR system (Figure 6.3). Each RNA sample was reverse transcribed, then three replicate pre-amplification reactions were carried out. Each preamplification reaction was then added to three replicate qPCR reactions. The RT and pre-amplification reactions were done using multiplex primers. The replicates of the different reactions were all run on the same Fluidigm chip so that if any variation occurred in the replicates, it would not be as a result of different experimental runs. Each RT reaction would give 9 samples to run, as the Fluidigm chip can run 48 samples at a time, I chose to use to 5 RT reactions. On the Fluidigm chip there were 45 samples RT reactions plus 3 no RT reactions. The RT reactions were 1. RNA freshly isolated from GMP cells, 2. repeat of RT1, 3. repeat of RT1 but with 15 cycles of preamplification instead of 12 cycles, 4. RNA from a GMP-like AML sample, RNA had been previously on the microarray and 5. RNA from a GMP sample, RNA had been previously on the microarray. No RT controls 1. RNA from RT1, 2, 3, 2. RNA from RT4 3. RNA from RT5.

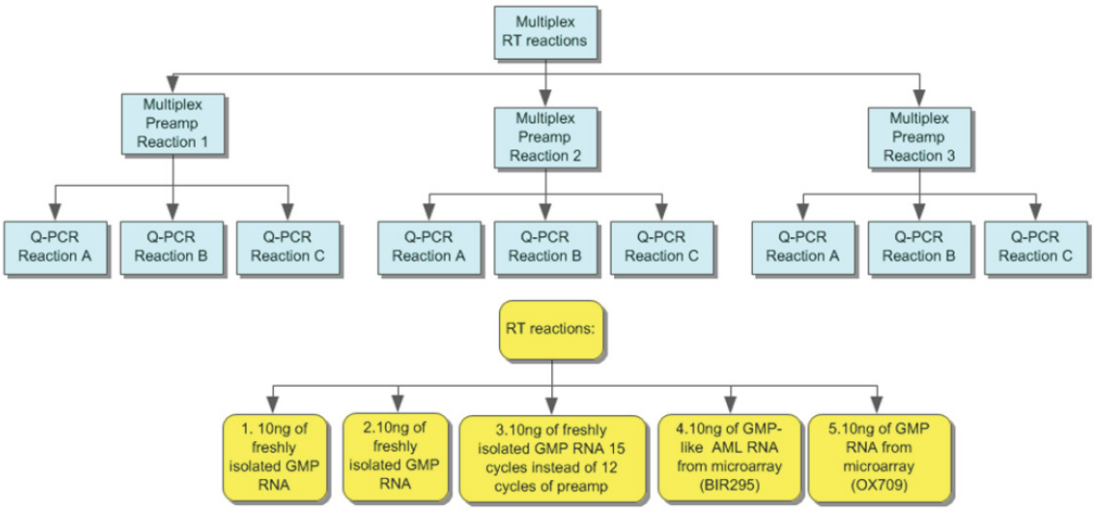


Fig 6.3 Validation of the qPCR system; schematic outlining experimental design.
An experimental plan was designed to test the fidelity of the microRNA assays at each stage. Diagram above illustrates the experimental outline. The RT and preamp reactions were multiplex reactions and qPCR was performed as singleplex reactions on the Fluidigm. (All RTs were done on the same day, all preamp reactions were carried out on the same day and all qPCRs were run on the same Fluidigm chip.)

Each of the amplification plots from the different assays were inspected to determine if amplification of the different assays had occurred efficiently. If the amplification reaction has worked well, the amplification curve shows three distinct phases a steep exponential phase followed by a linear phase then a plateau. Most of the assays showed good amplification curves an example of which can be seen in Figure 6.4A. However three assays gave poor amplification plots an example of this can be seen in Figure 6.4B.

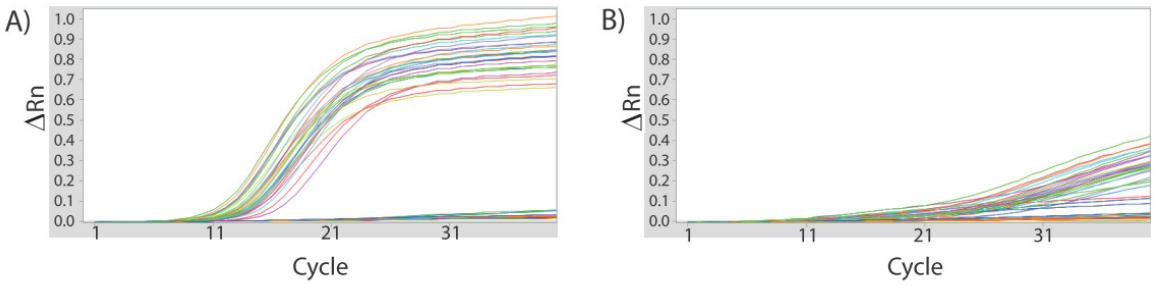


Fig 6.4 Examples of qPCR amplification plots.
The first stage in evaluating miRNA assays was to manually inspect amplification plots. A) Shows an example of what was deemed a good amplification plot (miR-17 assay) . B) Shows an example of what was deemed a bad plot (miR-181C assay)

There are several possible reasons for poor amplification such as insufficient amount of template available, inhibitors present in template, problems with design of the assay primers, problems with design of assay probe or primers being inhibited by other primers in multiplex reactions. For the assays with poor amplification this occurred across all samples tested, this suggest that the reason for poor amplification was due to assay design rather than template. When these assays were used across a larger number of RNA samples they still showed poor amplification and so were therefore excluded from any further analysis, these assays were against miR-361-5p, miR-181c

and miR-342-5p. The Ct values across the different replicates were then examined. In some assays the Ct values across the replicates was not consistent. As these assays were therefore not reliable they were also excluded from further analysis (miR-21*, miR-551b and miR886-3p). The rest of the assays all showed reliable amplification, and had Ct values ≤32, below the traditional cut off for exclusion of qPCR data which is a Ct value of ≥35. The mean Ct values of qPCR replicates for these assays can be found in table 6.1 of the appendices. In some of the assays amplification was not seen across all the RT templates. Two assays (miR-363 and miR-151-5p) only showed amplification in the template that had undergone 15 cycles of amplification, indicating that the amount of template may have been too low in other samples. Two assays (miR-195 and miR-34a) only showed amplification in the AML sample template, indicating the amount of template may be too low in normal samples. Two assays (miR-142-5p and miR-23a) only showed amplification in the normal sample with 15 cycles of amplification and the AML sample, also indicating the amount of template may be too low in normal samples. One assay miR-210 showed amplification in all RT templates except in RT 1 and 2. The standard deviation of the Ct value between the qPCR replicates can be seen for these assays in Table 6.2 in the appendices. The standard deviation of qPCR replicates for all assays was less than one Ct with an average Ct value for qPCR replicates of 0.29. Signifying a high level of reproducibly between qPCR reactions.

The mean Ct values for preamplification replicates are in table 6.3 in the appendices and the standard deviation of Ct values between preamplification replicates can be seen in table 6.4 in the appendices. The standard deviation between pre-amplification replicates is also low with most of the values falling below 1 Ct and average standard deviation being 0.18 Ct. Signifying that the level of reproducibility is also high between preamplification reactions.

6.4 A preamplification extension step of 4 minutes is better than an extension step of 1 minute.

The standard protocol when using multiplex preamplification with Taqman assays is to use a four minute extension step during preamplification. However four minutes is an unusually long extension step for small templates, after discussion with other Fluidigm users we decided to investigate if a one minute extension step would produce better data. The previous experiment outlined in Figure 6.3 was repeated using a one minute extension step during preamplification to investigate the effects of using a shorter extension step.

The results obtained when using a one minute extension step were compared to the previous experiment. All assays with poor or inconsistent amplification in the first experiment also displayed the same problem when a one minute extension step was used. A comparison of standard deviation between qPCR replicates for endogenous controls was done between the two experiments (Figure 6.5 A). The plots show that there is a higher degree of variation when a one minute extension is used. A comparison of the standard deviation between preamplification replicates for endogenous controls was also done (Figure 6.5 B), it was evident from these plots that the variation between the preamplification replicates was also higher when a one minute extension step was used.

The standard deviation of qPCR replicates across all assays was then compared between a one minute and four minute extension step. The min, max and the average standard deviation of qPCR replicates can be seen in Figure 6.6A. The standard deviation was higher for all assays when one minute extension step was used. The standard deviation of preamplification replicates across all assays was also compared

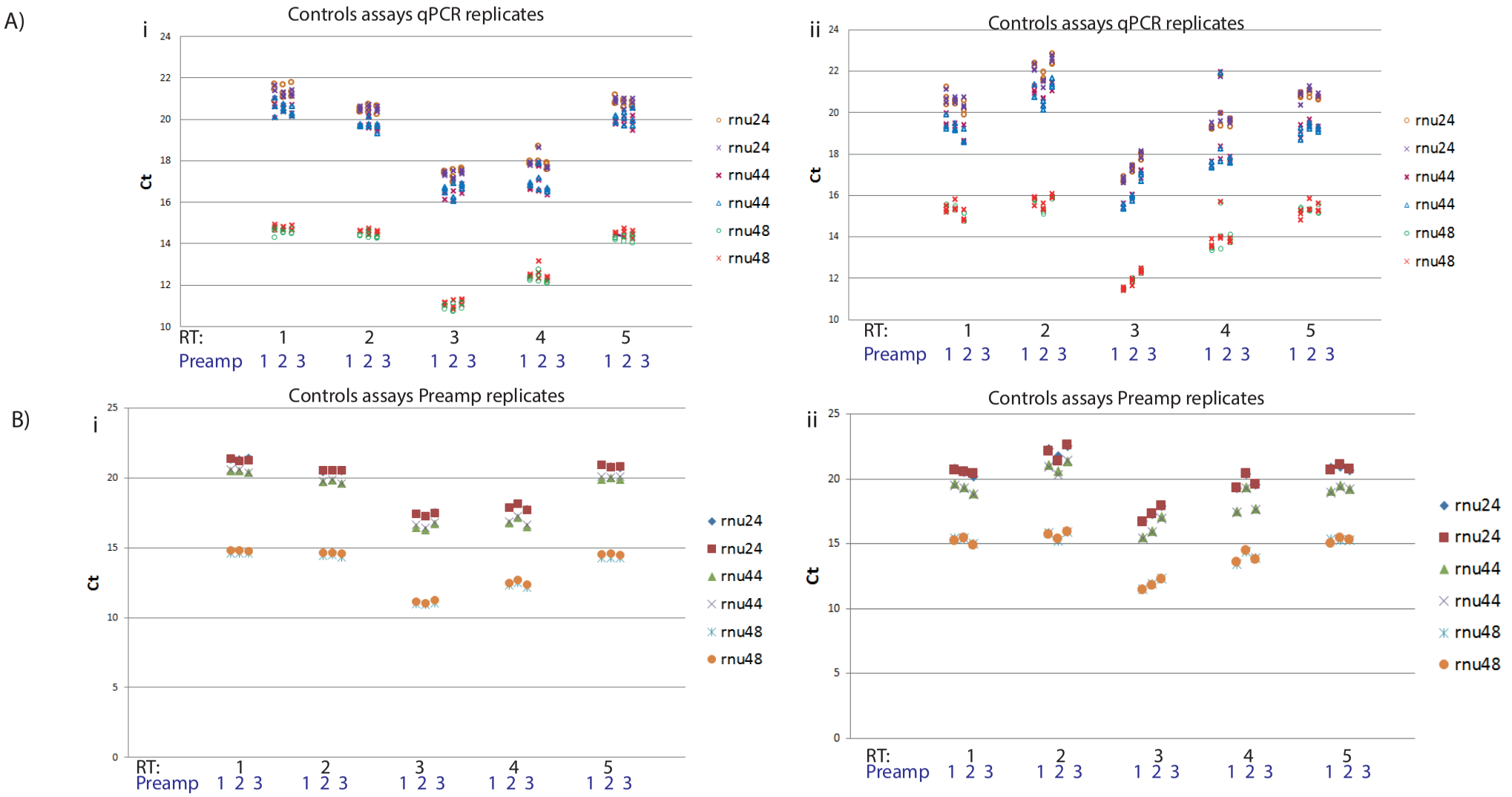


Fig 6.5 Comparison of different preamp extension steps on endogenous controls

Experiment in figure 6.3 was conducted using a (i) 4 minute extension step and (ii) a 1 minute extension step, plots above show results for control assays A) Ct values of qPCR replicates B) Ct values of qPCR replicates.

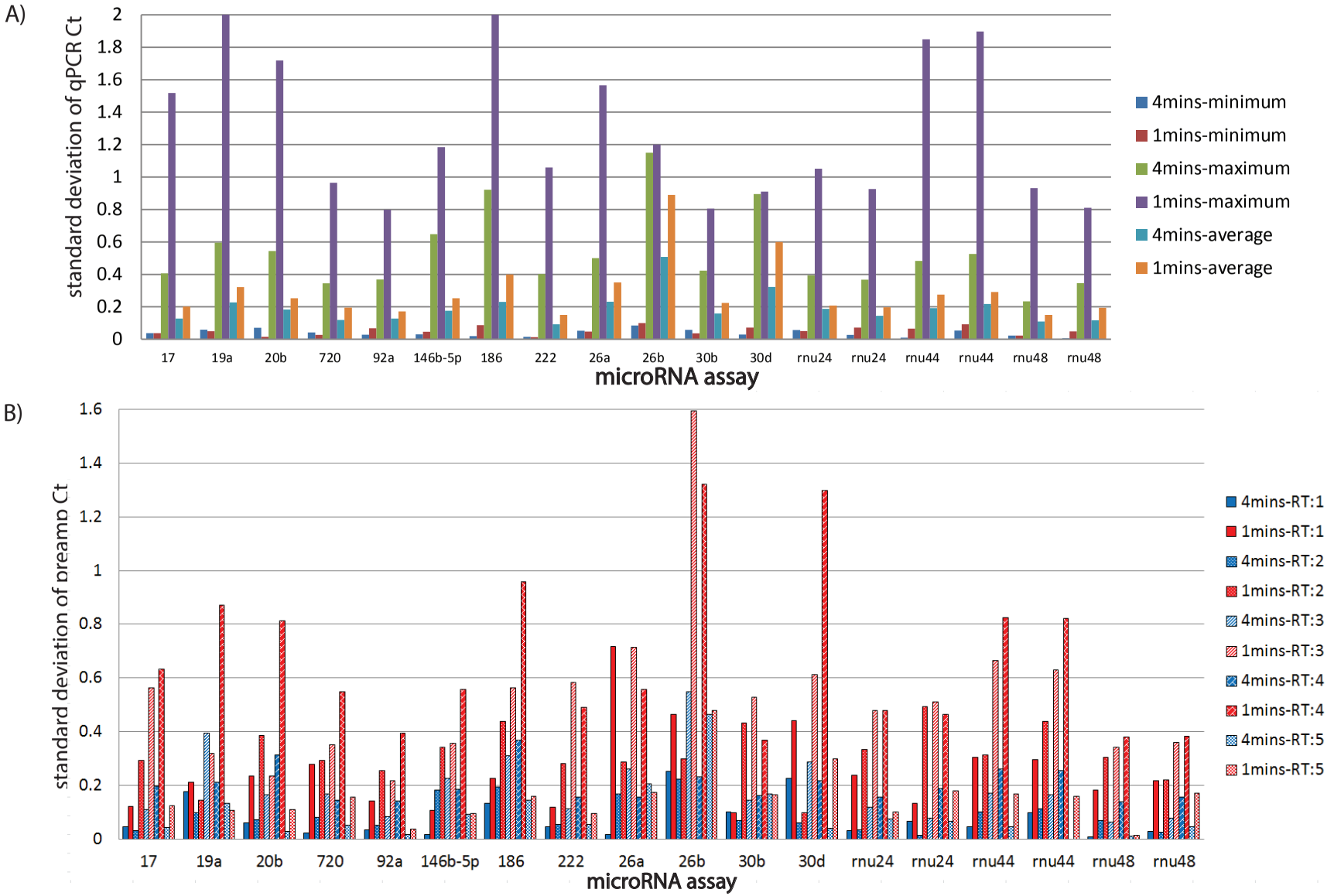


Fig 6.6 Comparison of different lenght of preamp extension step across all assays.
Experiment in figure 6.3 was conducted using a 4 minute extension step and a 1 minute extension step; the standard deviation of replicates for different assays were compared. A) The standard deviation of qPCR replicates was plotted for each assay. B) The standard deviation of preamp replicates was plotted for each assay. (Only assays that had consistent amplification across all RTs are included).

between a one minute and a four minute extension step (Figure 6.6B). The standard deviation was higher for all RTs across all assays when a one minute extension step was used.

Therefore the data is more reproducible when a four minute as opposed to a one minute extension step is used during preamplification. During preamplification primers are used at a limiting concentration. Also there are many different primers present as the reaction is multiplex. It is likely the longer time need is to allow the primers to find and bind their targets.

6.5 Assaying miRNA expression across a range of normal GMP and Leukemic GMP-like samples.

Following the validation experiments, miRNA expression quantification was carried out as follows; reverse transcription was performed using multiplex primers with RNA previously loaded on to microarrays; one RT reaction per RNA sample. Then each reverse transcribed cDNA sample was used in one multiplex preamplification reaction, with a four minute extension step and 15 cycles of preamplification. The pre-amplified cDNA was loaded on to a Fluidigm chip and run against all of assays of interest. The Fluidigm chip was repeated three times to give three replicates for each qPCR reaction. The mean expression of endogenous controls RNU24, RNU44 and RNU 48 was used to normalization miRNA expression between samples. In total 5 GMP samples and 23 GMP-like samples were assayed. Four of the five GMP samples assayed by qPCR were also analysed by microarray and all of the GMP-like samples assayed by qPCR were also analysed by microarray.

The normalized mean qPCR expression values of each of the miRNAs were plotted in graphs, alongside the normalized array expression values for lower expression in GMP-like AML compared to normal GMP (Figure 6.8), for overexpressed miRNAs in

GMP-like AML compared to normal GMP (Figure 6.9). GraphPad Prism software was used to determine if miRNA qPCR expression was significantly different, between GMP and GMP-like populations according to qPCR data. Table 6.5 summarizes the qPCR results and contains p value and expression fold changes. If the P-value was ≤0.05 the expression change was deemed significant and is highlighted in yellow. If the mean expression in GMP-like samples was 0.6 fold or less than the expression in GMP samples it was highlighted in red and if the mean expression in GMP-like samples was 1.5 fold or more than expression in GMP samples it is highlighted in green.

Nine miRNAs were significantly less expressed in GMP-like samples according to the array data. Of these 6 could be assayed by qPCR (Figure 6.7). Three miRNAs were not validated by qPCR as the assays had been shown previously to be unreliable (section 6.3). All 6 miRNAs that had lower expression in GMP-like AML, were also found to have lower expression by qPCR. The difference in expression of 5 of these miRNAs was also deemed as significant by qPCR. miR-20b was found to be less expressed in GMP-like samples by qPCR but had a P-Value of 0.0744 so expression difference was not deemed significant. Microarray analysis of miR-363 showed a bimodal expression pattern in the AML samples, this expression pattern was also seen by qPCR analysis. The expression levels between different miRNAs was not preserved between the array and qPCR data, for example miR-720 was expressed at a higher level than miR-92a according to array data but their expression level was similar according to qPCR data. This is likely due to different levels of efficiency of microarray probes and the different efficiency levels of qPCR assays. As we are not comparing levels of different microRNAs but comparing the expression level of one miRNA in different populations this does not affect our conclusions.

Sixteen miRNAs were overexpressed in GMP-like samples according to array data, of these 13 could be assayed by qPCR (Figure 6.8). Three miRNAs were not validated by qPCR as the assays had been shown previously to be unreliable (section 6.3).

microRNA	GMP-like AML mean expression	GMP-like AML StDev	normal GMP mean expression	normal GMP StDev	p value	Expression fold change in GMP-like AML compared to normal GMP
17	2.454	± 0.2489	5.152	± 0.3679	< 0.0001	0.48
186	0.04133	± 0.007246	0.04567	± 0.01205	0.796	0.90
195	0.000179	± 8.740e-005	0.000108	± 3.388e-005	0.4558	1.66
210	0.004198	± 0.001438	0.001339	± 0.0002584	0.0625	3.14
222	2.766	± 0.3887	2.022	± 0.3449	0.3939	1.37
363	2.00E-05	± 5.589e-006	9.02E-05	± 1.928e-005	< 0.0001	0.22
720	0.4985	± 0.05416	0.722	± 0.03293	0.0018	0.69
142-5p	0.000964	± 0.0001596	0.000796	± 0.0001129	0.402	1.21
146-5p	0.2664	± 0.04147	0.1553	± 0.01810	0.0214	1.72
151-5p	0.001679	± 0.0004357	0.000428	± 6.096e-005	0.0094	3.92
19a	0.01882	± 0.001956	0.03684	± 0.001844	0.0003	0.51
20b	1.449	± 0.2394	2.467	± 0.3907	0.0744	0.59
23a	0.000268	± 5.577e-005	0.000218	± 3.885e-005	0.4738	1.23
26a	0.01025	± 0.0009111	0.008008	± 0.001383	0.2899	1.28
26b	0.000517	± 6.512e-005	0.000361	± 6.344e-005	0.2903	1.43
30b	0.08232	± 0.005316	0.08362	± 0.008813	0.9158	0.98
30d	0.006408	± 0.0005953	0.004591	± 0.0009252	0.192	1.40
34a	0.006404	± 0.001643	0.002384	± 0.0006647	0.0322	2.69
92a	0.5376	± 0.07166	0.9422	± 0.1038	0.0194	0.57

Table 6.2 qPCR data on microRNA expression in GMP-like AML and normal GMP.
P-value was generated in graph pad using a t-test analysis. P-values lower than 0.05 were deemed significant and are highlighted. Fold changes greater than 1.5 are highlighted, red indicates microRNA more expressed in normal GMP and green indicate microRNA more expressed in GMP like AML

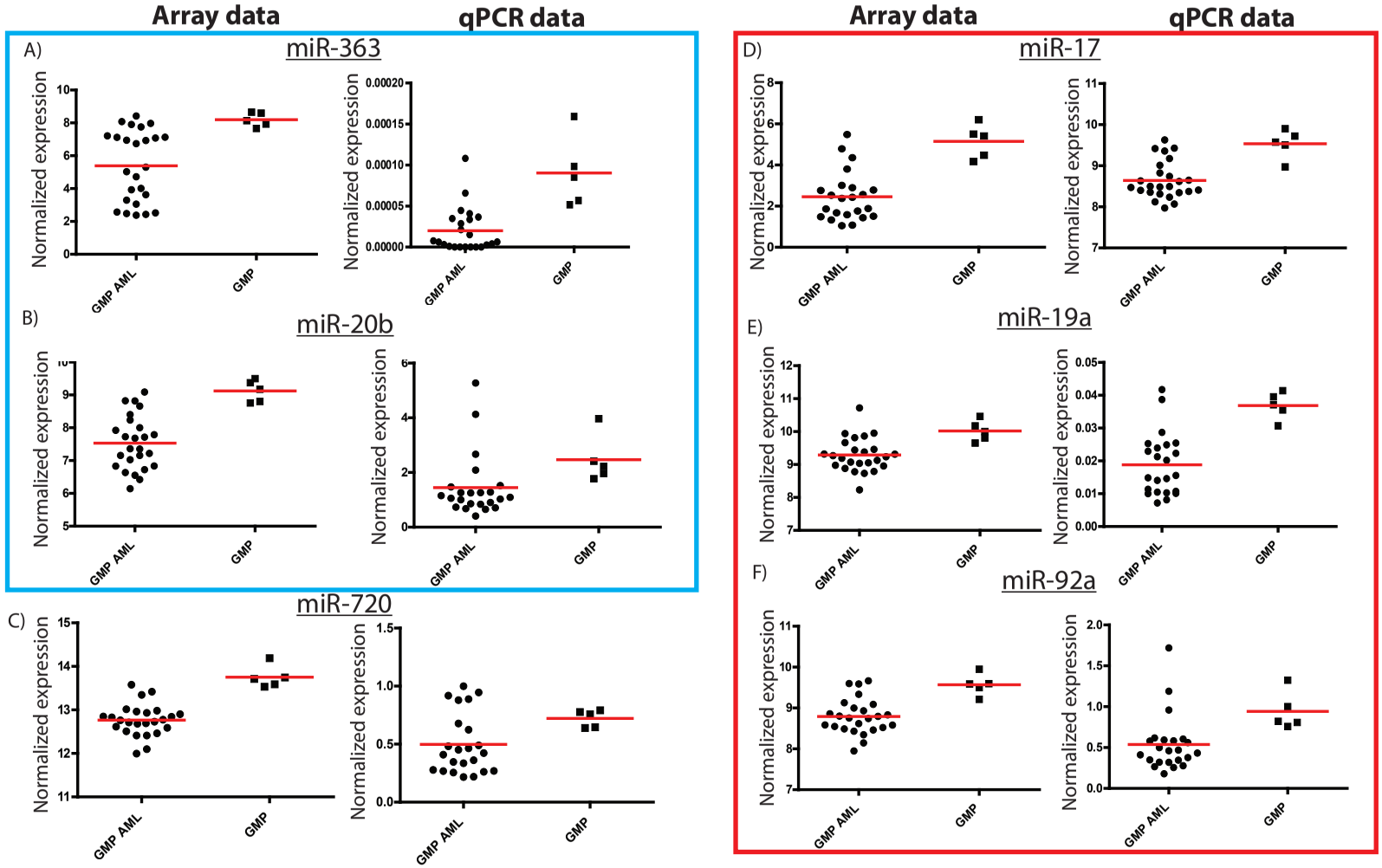
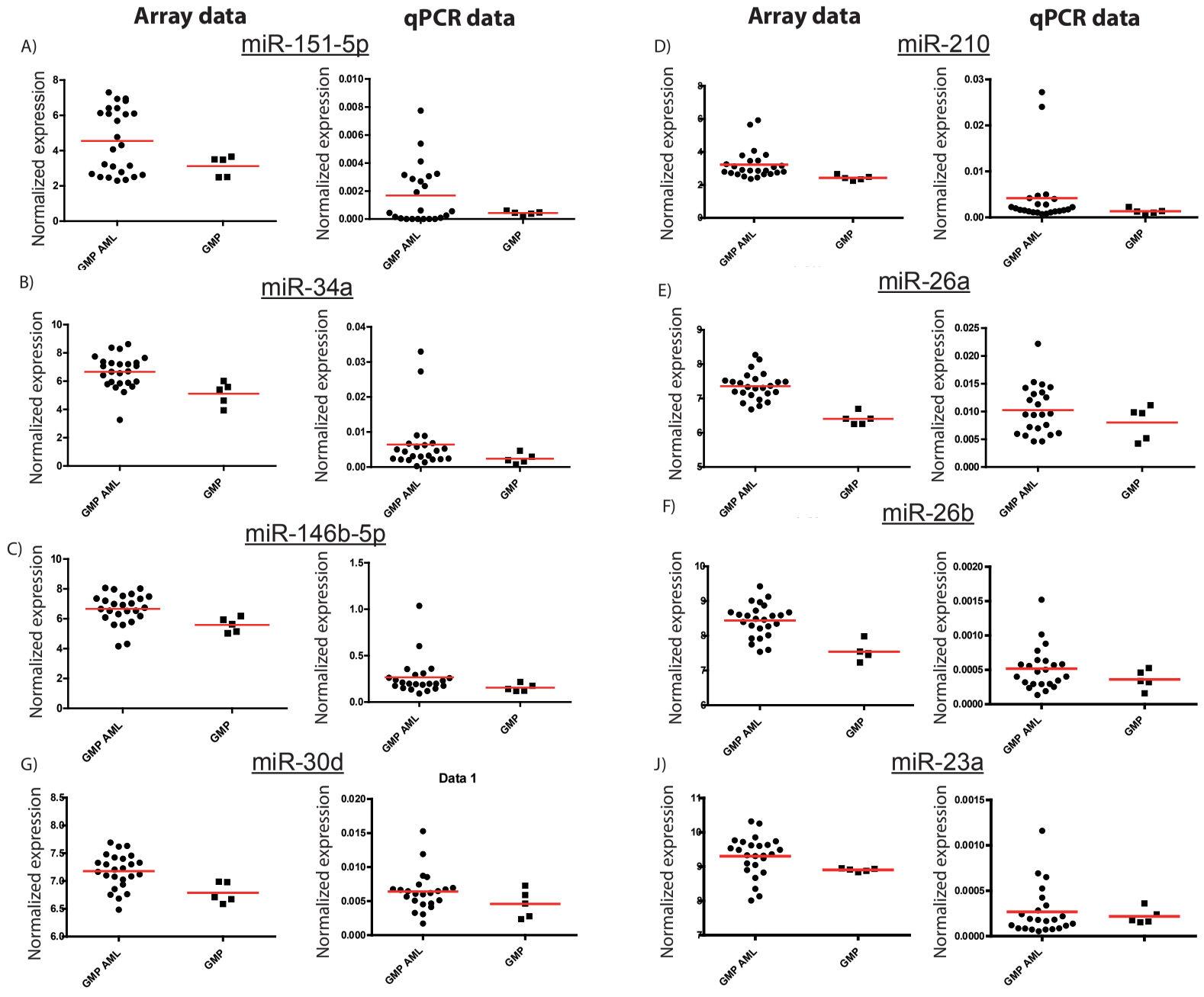


Fig 6.7 Comparing microarray data to qPCR data; miRNAs with reduced expression in GMP-like AMLs compared to GMPs.
Graph on left hand side shows microarray data and graph on right hand side shows qPCR data. Microarray data was generated from 25 GMP-like AML samples and 5 age matched normal GMP samples. qPCR data was generated from 23 GMP-like AML samples and 5 age matched normal GMP samples. The 23 GMP-like AML samples used in qPCR were also analyzed on arrays. 3 of the 5 aged matched normal GMP samples used in the qPCR were also analyzed on arrays. Normalized expression values are given; Microarray data is normalized using quantile normalization, and qPCR data is normalized to endogenous controls. Microarray values are denoted to the log2. All qPCR reactions were carried out in triplicate, plots show mean value. miRNAs within blue box are part of the 106-363 cluster. miRNAs within the red box are part of the 17-92a cluster.


Array data
qPCR data
A)
miR-151-5p
B)
miR-34a
C)
miR-146b-5p
D)
miR-210
E)
miR-26a
F)
miR-26b
G)
miR-30d
J)
miR-23a
Normalized expression
GMP AML
GMP
Data 1

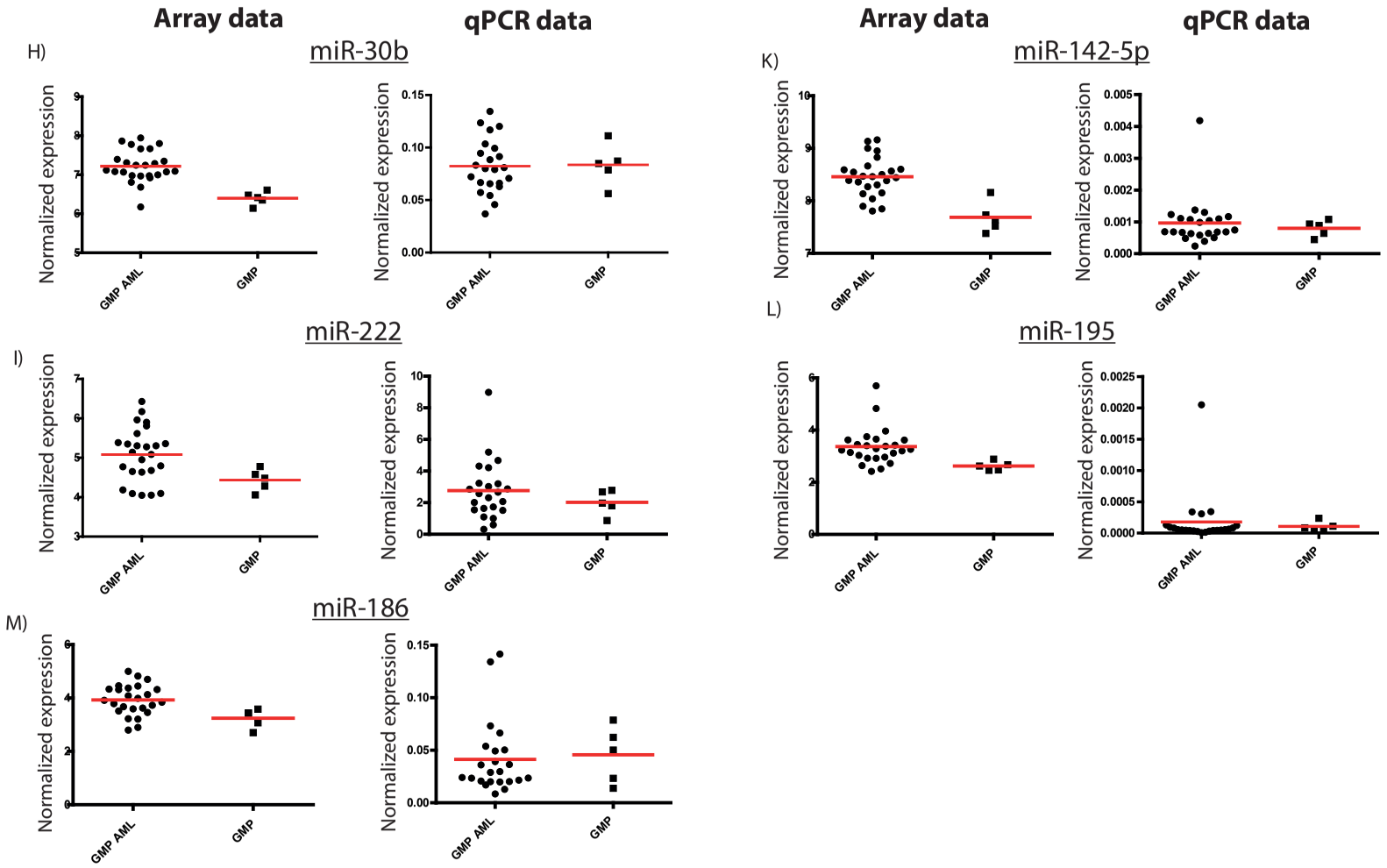


Fig 6.8 Comparing microarray date to qPCR data; MiRNAs overexpressed in GMP-like AML compared to GMP.
Graph on left hand side shows microarray data and graph on right hand side shows qPCR data. Microarray data was generated from 25 GMP-like AML samples and 5 age matched normal GMP samples. qPCR data was generated from 23 GMP-like AML samples and 5 age matched normal GMP samples. The 23 GMP-like AML samples used in the qPCR were also analyzed on arrays. 3 of the 5 aged matched normal GMP samples used in the qPCR were also analyzed on arrays. Normalized expression values are given; Microarray data is normalized using quantile normalization, and qPCR data is normalized to endogenous controls. Microarray values are denoted to the log2. All qPCR reactions were carried out in triplicate, plots show mean value.

The expression of another three was quantified similarly and significantly by both techniques (miR151-5P, Mir-34a and miR146-5p). miR-151 also had a bimodal expression pattern in the AML samples according to microarray analysis this was also seen by qPCR. The expression of four miRNAs was quantified similarly by both techniques but not significantly by qPCR (miR30d, miR-210, miR-26a and miR-26b). Six miRNAs were expressed differentially according to array data but showed equivalent levels of expression between normal GMP samples and leukemic GMP-like samples (miR-23a, miR-30b, miR222, miR-186, miR-142-5p, miR195).

Some but not all of the miRNAs could be validated by qPCR, all of the miRNAs that were found to be less expressed in AML by the array were validated, but only a few of the miRNA found to be overexpressed by the array were validated. Some of the miRNAs showed bimodal expression patterns, this may have been an artefact of the microarrays but this is unlikely as the bimodal expression patterns were confirmed by qPCR. Where the techniques disagree, it is not possible to know which technique is correct. However when qPCR results agree with the microarray results, we can conclude with more confidence that these miRNAs are aberrantly expressed in AML.

6.6 Discussion IV

The aim of this final section of work was to validate the expression of miRNAs found differentially expressed between GMP-like AML cells and normal GMP cells. Taqman miRNA assays were chosen to do this with the Fluidigm system, as this used the lowest amount of RNA and reagents and was the most time efficient method.

Exact protocols on how to use the Fluidigm system with Taqman microRNA assays were not available from the manufactures, on how we needed to conduct the experiments. However Jang et al had published a method on this [126]. They showed

that using their method they had good correlation using different amounts of RNA down to 10ng. As well as good correlation comparing multiplex to singleplex reverse transcription reactions. Therefore a protocol from their laboratory was adapted to be used in our experiment. However to ensure our system was working efficiently a validation experiment was first undertaken. The data from this showed that we saw good correlation between qPCR replicates and preamplification replicates. Demonstrating that the system was working well.

However the extension step used in the pre-amplification reaction was 4 minutes long. This is an unusually long extension step for such small templates. The Taqman manufactures were contacted about this. Their response was that a longer extension step was included in the protocol, as the annealing stage also occurs during the extension step, and that the increased time was there to allow the primers to bind their targets. As primers are used at limiting concentrations during preamplification. They however suggested that we did a trial using a one minute extension step. I repeated the initial validation experiment using a one minute extension step during preamplification, then compared the results to what had been seen previously with a 4 minute extension step. The correlation between replicates was lower when a one minute extension step was used; indicating that the extra time was needed for the primers to bind their targets. Therefore we decided to conduct the experiment using a four minute extension step.

Of the 25 miRNAs to be analysed, 3 of the assays had to be eliminated as there was poor amplification. It is likely that there was an insufficient amount of template available for the assays. A potential solution for further qPCR analysis of these genes would be to use larger amounts of RNA. Three of the assay showed inconsistent amplification across replicates; however insufficient template is unlikely to be the cause of inconsistent amplification in these assays. It is not clear what the issue with these

assays. However issues with assays such as this implicate the importance of testing assays prior to use.

Of the 19 miRNAs analysed 8 were significantly differential expressed according to qPCR data. Five showed the same pattern of expression by qPCR as the array but the difference was not deemed significant by qPCR, and 6 showed no differential expression at all according to qPCR data. Some of the microRNAs showed a bimodal expression pattern in AML in the array data, this expression pattern was also seen by qPCR. This indicates that this is not an artefact of the microarray but that in AML a segregation of expression occurs, what the functional significance of this is still needs to be determined. Although not all of the miRNAs could be validated by qPCR this not an uncommon phenomenon and has been observed by others [134,143]. Reasons for the differences between qPCR and array data could be for several reasons such as false positives in the array data, false negatives in the qPCR data, probes/primers binding to nonspecific targets (as many miRNAs exhibit only 1 or 2 base differences), as well as the efficiency of both technologies being lower for less expressed targets and finally samples were stored for a lot longer prior to qPCR analysis so sample may have been lost due to it adhering to sample tube walls. .

microRNA assay	RT:1 PA:1	RT:1 PA:2	RT:1 PA:3	RT:2 PA:1	RT:2 PA:2	RT:2 PA:3	RT:3 PA:1	RT:3 PA:2	RT:3 PA:3	RT:4 PA:1	RT:4 PA:2	RT:4 PA:3	RT:5 PA:1	RT:5 PA:2	RT:5 PA:3
17	16.042	16.043	15.943	15.390	15.313	15.328	12.129	12.028	12.295	14.762	15.126	14.667	15.667	15.758	15.660
19a	22.619	22.620	22.242	21.110	20.013	20.865	18.045	18.000	18.860	20.838	21.191	20.678	21.689	22.004	21.710
20b	17.428	17.392	17.285	16.535	16.509	16.366	13.403	13.255	13.660	16.178	16.718	15.976	16.943	16.958	16.880
363	NR	NR	NR	NR	NR	NR	27.310	27.344	28.397	NR	NR	NR	NR	NR	NR
720	17.396	17.404	17.029	16.778	16.709	16.582	13.465	13.336	13.739	15.599	15.857	15.516	16.531	16.573	16.448
92a	19.138	19.057	19.066	18.573	18.568	18.458	15.217	15.225	15.402	17.862	18.071	17.720	19.211	19.183	19.224
142-5p	NR	NR	NR	NR	NR	NR	28.057	27.003	26.992	26.242	25.878	25.536	NR	NR	NR
146b-5p	18.901	18.861	18.884	18.462	18.404	18.047	14.565	14.412	14.950	17.930	18.329	17.933	18.163	18.146	17.959
151-5p	NR	NR	NR	NR	NR	NR	27.197	26.576	26.764	NR	NR	NR	NR	NR	NR
186	20.719	20.540	20.865	19.933	19.925	19.512	16.272	16.049	16.794	17.823	18.576	17.767	20.064	19.944	19.709
195	NR	NR	NR	NR	NR	NR	NR	NR	NR	26.798	29.042	28.091	NR	NR	NR
210	NR	NR	NR	NR	NR	NR	29.896	31.713	29.241	29.658	31.154	29.733	29.315	30.483	30.067
222	15.546	15.487	15.432	14.786	14.750	14.651	11.505	11.419	11.694	11.771	12.091	11.744	14.546	14.549	14.428
23a	NR	NR	NR	NR	NR	NR	25.134	27.096	27.960	27.481	26.876	27.227	NR	NR	NR
26a	22.944	22.928	22.973	21.823	21.730	21.425	18.515	18.359	18.976	19.962	20.340	20.077	22.582	22.188	22.103
26b	25.960	26.579	26.191	25.420	25.164	25.714	21.690	21.783	22.901	22.516	23.085	22.770	25.260	25.996	24.875
30b	20.707	20.952	20.855	19.925	19.936	19.781	16.774	16.563	16.919	17.143	17.507	17.181	20.558	20.410	20.149
30d	24.466	23.916	24.158	23.188	23.079	23.221	20.024	19.866	20.539	20.046	20.316	19.783	23.581	23.483	23.562
34a	NR	NR	NR	NR	NR	NR	NR	NR	NR	23.541	23.915	23.652	NR	NR	NR
rnu24	21.064	21.346	21.417	20.463	20.549	20.515	17.439	17.276	17.570	17.909	18.162	17.782	20.919	20.809	20.729
rnu24	21.365	21.200	21.287	20.545	20.561	20.526	17.412	17.292	17.482	17.889	18.168	17.703	20.956	20.801	20.819
rnu44	20.506	20.501	20.406	19.692	19.836	19.586	16.407	16.283	16.696	16.743	17.149	16.514	19.887	19.986	19.879
rnu44	20.621	20.587	20.397	19.748	19.883	19.605	16.649	16.425	16.834	16.885	17.261	16.634	20.085	20.076	20.092
rnu48	14.586	14.603	14.600	14.420	14.483	14.314	10.996	10.904	11.063	12.318	12.494	12.147	14.261	14.289	14.278
rnu48	14.829	14.805	14.757	14.641	14.606	14.582	11.129	11.052	11.245	12.492	12.732	12.348	14.523	14.589	14.472

Appendix 6-Table 1 Mean CT values from qPCR replicates for each assay.
Table relates to figure 6.3, a 4 minute preamlification extension step was used. Threshold to determine Ct value was manually set for each assay. (NR: when amplification failed).

microRNA assay	RT:1 PA:1	RT:1 PA:2	RT:1 PA:3	RT:2 PA:1	RT:2 PA:2	RT:2 PA:3	RT:3 PA:1	RT:3 PA:2	RT:3 PA:3	RT:4 PA:1	RT:4 PA:2	RT:4 PA:3	RT:5 PA:1	RT:5 PA:2	RT:5 PA:3	min	max	average
17	0.162	0.038	0.108	0.052	0.083	0.092	0.103	0.163	0.083	0.136	0.406	0.090	0.049	0.125	0.166	**0.038**	**0.406**	**0.124**
19a	0.447	0.227	0.241	0.096	0.191	0.194	0.059	0.208	0.166	0.400	0.596	0.099	0.100	0.208	0.174	**0.059**	**0.596**	**0.227**
20b	0.248	0.104	0.193	0.071	0.267	0.148	0.124	0.236	0.087	0.161	0.544	0.132	0.080	0.103	0.248	**0.071**	**0.544**	**0.183**
363	NR	NR	NR	NR	NR	NR	0.200	0.444	0.390	NR	NR	NR	NR	NR	NR	**0.200**	**0.444**	**0.345**
720	0.164	0.111	0.136	0.078	0.142	0.059	0.086	0.166	0.073	0.143	0.345	0.064	0.065	0.042	0.114	**0.042**	**0.345**	**0.119**
92a	0.116	0.138	0.099	0.077	0.029	0.125	0.108	0.189	0.140	0.028	0.368	0.050	0.110	0.183	0.149	**0.028**	**0.368**	**0.127**
142-5p	NR	NR	NR	NR	NR	NR	1.130	0.393	0.535	0.359	0.633	0.085	NR	NR	NR	**0.085**	**1.130**	**0.523**
146b-5p	0.216	0.060	0.101	0.085	0.211	0.060	0.064	0.218	0.052	0.304	0.648	0.030	0.107	0.231	0.246	**0.030**	**0.648**	**0.176**
151-5p	NR	NR	NR	NR	NR	NR	0.070	0.301	0.475	NR	NR	NR	NR	NR	NR	**0.070**	**0.475**	**0.282**
186	0.125	0.115	0.133	0.138	0.216	0.020	0.099	0.327	0.136	0.281	0.921	0.060	0.155	0.405	0.332	**0.020**	**0.921**	**0.231**
195	NR	NR	NR	NR	NR	NR	NR	NR	NR	0.547	1.602	0.373	NR	NR	NR	**0.373**	**1.602**	**0.841**
210	NR	NR	NR	NR	NR	NR	0.958	0.858	0.438	0.771	2.111	0.909	0.390	0.135	0.508	**0.135**	**2.111**	**0.786**
222	0.095	0.020	0.074	0.015	0.088	0.017	0.049	0.124	0.058	0.121	0.401	0.024	0.073	0.113	0.109	**0.015**	**0.401**	**0.092**
23a	NR	NR	NR	NR	NR	NR	0.361	0.829	0.141	1.453	0.646	1.029	NR	NR	NR	**0.141**	**1.453**	**0.743**
26a	0.112	0.159	0.103	0.100	0.216	0.052	0.103	0.250	0.103	0.381	0.500	0.162	0.254	0.487	0.495	**0.052**	**0.500**	**0.232**
26b	1.150	0.633	0.085	0.470	0.524	1.117	0.158	0.195	0.190	0.291	0.498	0.326	0.622	1.079	0.284	**0.085**	**1.150**	**0.508**
30b	0.076	0.173	0.109	0.097	0.217	0.068	0.141	0.181	0.129	0.217	0.423	0.058	0.128	0.189	0.179	**0.058**	**0.423**	**0.159**
30d	0.895	0.262	0.178	0.143	0.350	0.220	0.076	0.229	0.029	0.076	0.523	0.067	0.287	0.830	0.672	**0.029**	**0.895**	**0.323**
34a	NR	NR	NR	NR	NR	NR	NR	NR	NR	0.438	0.463	0.175	NR	NR	NR	**0.175**	**0.463**	**0.359**
rnu24	0.331	0.252	0.292	0.104	0.162	0.189	0.057	0.246	0.079	0.065	0.394	0.149	0.188	0.167	0.132	**0.057**	**0.394**	**0.187**
rnu24	0.256	0.100	0.127	0.087	0.149	0.095	0.085	0.173	0.075	0.071	0.302	0.027	0.091	0.212	0.250	**0.027**	**0.302**	**0.140**
rnu44	0.294	0.149	0.206	0.010	0.223	0.101	0.176	0.216	0.188	0.103	0.483	0.097	0.053	0.189	0.356	**0.010**	**0.483**	**0.190**
rnu44	0.392	0.102	0.189	0.054	0.207	0.175	0.112	0.369	0.105	0.074	0.526	0.082	0.152	0.258	0.441	**0.054**	**0.526**	**0.216**
rnu48	0.200	0.052	0.105	0.023	0.142	0.028	0.083	0.177	0.103	0.073	0.234	0.046	0.037	0.125	0.220	**0.023**	**0.234**	**0.110**
rnu48	0.114	0.067	0.106	0.006	0.119	0.081	0.064	0.186	0.112	0.070	0.346	0.079	0.048	0.106	0.195	**0.006**	**0.346**	**0.113**

Appendix 6- Table 2 Standard deviation of CT values from qPCR replicates for each assay.
Table relates to figure 6.3, a 4 minute preamplification extension step was used. Threshold to determine Ct value was manually set for each assay. (NR: when amplification failed).

microRNA assay	RT:1	RT:2	RT:3	RT:4	RT:5
17	16.009	15.344	12.151	14.852	15.695
19a	22.494	20.991	18.302	20.902	21.820
20b	17.368	16.470	13.440	16.291	16.929
363	NR	NR	27.684	NR	NR
720	17.385	16.690	13.513	15.657	16.518
92a	19.087	18.533	15.281	17.884	19.206
142-5p	NR	NR	27.467	25.886	NR
146b-5p	18.882	18.305	14.643	18.064	18.090
151-5p	NR	NR	26.846	NR	NR
186	20.708	19.790	16.372	18.055	19.906
195	NR	NR	NR	28.172	NR
210	NR	NR	30.284	30.181	29.955
222	15.488	14.729	11.539	11.869	14.507
23a	NR	NR	26.730	27.195	NR
26a	22.948	21.659	18.617	20.126	22.291
26b	26.246	25.433	22.125	22.790	25.377
30b	20.838	19.881	16.752	17.277	20.372
30d	24.180	23.163	20.143	20.048	23.542
34a	NR	NR	NR	23.703	NR
rnu24	21.372	20.509	17.428	17.951	20.819
rnu24	21.284	20.544	17.396	17.920	20.858
rnu44	20.471	19.705	16.462	16.802	19.917
rnu44	20.535	19.746	16.636	16.927	20.084
rnu48	14.596	14.406	10.987	12.320	14.276
rnu48	14.79709	14.61937	11.14175	12.52387	14.52785

Appendix-Table 3 Mean CT values of pre-amplification replicates for each assay. Table relates to figure 6.3, a 4 minute preamplification extension step was used. Threshold to determine Ct value was manually set for each assay. (NR: when amplification failed).

microRNA assay	RT:1	RT:2	RT:3	RT:4	RT:5
17	0.047	0.033	0.110	0.198	0.045
19a	0.178	0.100	0.395	0.214	0.134
20b	0.061	0.074	0.167	0.313	0.031
363	NR	NR	0.505	NR	NR
720	0.022	0.081	0.168	0.145	0.052
92a	0.036	0.053	0.085	0.144	0.017
142-5p	NR	NR	0.442	0.288	NR
146b-5p	0.016	0.184	0.226	0.188	0.093
151-5p	NR	NR	0.260	NR	NR
186	0.133	0.197	0.312	0.369	0.148
195	NR	NR	NR	0.983	NR
210	NR	NR	1.045	0.688	0.483
222	0.047	0.058	0.115	0.158	0.056
23a	NR	NR	1.182	0.248	NR
26a	0.018	0.170	0.262	0.158	0.209
26b	0.253	0.225	0.550	0.233	0.465
30b	0.101	0.070	0.146	0.163	0.169
30d	0.225	0.061	0.288	0.218	0.042
34a	NR	NR	NR	0.157	NR
rnu24	0.032	0.035	0.120	0.158	0.078
rnu24	0.067	0.014	0.079	0.191	0.069
rnu44	0.046	0.103	0.173	0.262	0.049
rnu44	0.099	0.113	0.167	0.258	0.006
rnu48	0.007	0.070	0.065	0.142	0.012
rnu48	0.030	0.027	0.079	0.158	0.048

Appendix 6-Table 4 Standard deviation of CT values of pre-amplification replicates for each assay.
Table relates to figure 6.3, a 4 minute preamplification extension step was used. Threshold to determine Ct value was manually set for each assay. (NR: when amplification failed).

Chapter 7: Final Discussion

The aim of my thesis was to better understand the cell biology of acute myeloid leukemia. I focused on GMP-like and LMPP-like expanded CD34+ AML, which accounted for around 80% of CD34+ AML samples.

The aim of the first part of my thesis was to improve the Flow cytometry purification of the GMP-like and LMPP-like populations in AML. I did not manage to further purify populations based on my approach of looking at mRNA genes, which were differentially expressed between the two populations. However lessons were learned from this approach that could be useful in improving future studies of this kind. A high level of patient heterogeneity was seen in AML samples used and it was observed that this can greatly impact on data analysis, when comparing differences between leuekmic populations. The heterogeneity can mask potential differences and can introduce false positives. It is therefore important to take steps to address this when comparing populations within AML, such as using paired sample analysis. We also found that the expression of cell surface proteins often differs from what is expected based on mRNA gene expression profiles. Therefore when looking for alterations on the cell surface, employing strategies that directly assay expression may be more useful, such as FACS screen. We also observed that cell surface proteins tested appeared on some patient samples but not others. Additionally CCR8 varied on which cell population it was expressed on between different patient samples. This highlights that differences are found in the proteins expressed on the surface of AML cells. However what the functional relevance of this is remains a question to be answered.

The second part of my thesis looked at the molecular side of AML cells; by exploring miRNA expression in AML. miRNA expression profiles on highly purified AML LSC

and normal populations were generated using microarrays. The data was then explored and analysed using bioinformatic analyses.

The bioinformatic analysis revealed that leukemic stem cell populations had distinct miRNA signatures, and that the normal populations they more closely resembled were more mature progenitor populations rather than stem cell populations. Data generated from this study and previous studies in our laboratory demonstrated that on a molecular and phenotypic level the GMP-like AML cells were closest to normal GMP cells. However it was found that the GMP-like LSCs possessed an aberrant miRNA stem cell like signature. This further increased our understanding of what is happening on a molecular level in AML LSCs. The aberrantly expressed stem cell-like miRNA signature may be part of these cells leuekmic transformation and may play a role in AML cells stem cell like behaviour. Prior to beginning my study nobody had profiled Human AML and normal populations in this way, and was what many in the field agreed needed to be done.[120, 146]

My study showed that it is possible to study cells in this way and allowed for a much greater understanding of how the LSC resembles normal hematopoiesis. Therefore future studies should carefully consider what populations are profiled, when exploring gene expression. My study also revealed differences between the AML and normal populations, that would have been masked if bulk populations were profiled, highlighting the importance to separate these populations when profiling them. Another important consideration to take into account when thinking about which populations to profile, are potential therapeutic applications. Altering aberrantly expressed miRNAs may prove to be useful when treating AML. However when targeting these miRNAs we must also consider the impact on the remaining normal cells. In any treatment strategy we would want to reduce our chances of damaging the HSC compartment. However damaging progenitor populations would not be so much of an issue as long as there were HSCs remaining to replace these cells.

Therefore understanding which miRNAs are expressed in the different normals populations will be pivotal in deciding which miRNAs can be targeted as AML therapies.

As well as learning how miRNAs in AML LSCs relate to normal populations, a group of miRNAs was identified that was differentially expressed between GMP and GMP-like cells. These altered miRNAs would make good candidates to be targeted for therapy, as well as targets for further investigation. qPCR analysis was carried on the differentially expressed miRNAs in normal GMP and leuekmic GMP-like cells. This validated some but not all of the differences seen by microarray. There are many reasons that some of the miRNAs expression could not be validated by qPCR as outline in the discussion at the end of chapter 6. However this can be a useful first step in reducing the list of miRNAs to be taken forward for further analysis. For the validated miRNAs it may be interesting to look at the expression of these miRNAs in further AML samples, to see if their expression patterns can be linked to common chromosomal abnormalities or gene mutations associated with AML. To gain greater understand of the underlining networks and pathways occurring AML.

An interesting set of miRNAs was identified as downregulated in AML, which were also validated by qPCR. Seven of the nine down regulated miRNAs in AML belonged to either the mir-17-92 cluster or its paralogous cluster mir-106-363 [144, 147]. It is unsurprising that miRNAs within the same cluster are similarly expressed, as miRNAs within a cluster are transcribed together as part of the same primary miRNA. Therefore observing the down regulation of multiple members of a cluster is a good validation of the array data. Other miRNAs that are part of these clusters that were not identified as differentially expressed, were also examined and found to either not have probes present on the microarray, or no expression was detected on the microarray. The mir-17-92 cluster is found on a separate chromosome to mir-106-363 indicating the down regulation is not due to a single genomic alteration, so potentially

a common factor may be controlling these clusters. The two clusters may also share a common mechanism of action. As it is likely the clusters have many of the same mRNA targets; as several of the miRNAs from the miR106a-363 have identical seed sequences to the miRNAs in the mir-17-92 cluster (e.g. miR-92 and miR-363 have the same seed sequence). The mir-17-92 cluster is commonly thought of as oncogenic and has been found overexpressed in many cancers, contradictory to was seen in AML in this study [148-151]. However the role of these miRNAs as just oncogenes has come into question recently, with examples of the cluster potentially acting as a tumour suppressor now being observed [152, 153]. Indicating the role of this cluster may not be as straightforward as originally anticipated and that tissue type and its exact level of expression may play a role in what function it plays.

The mir-17-92 cluster has also been shown to play a role in B-cell development and is found to be over expressed in some lymphoid malignancies [148, 149, 154]. It may be that this cluster or parts of it are important in a myeloid/lymphoid switch, as miRNAs are often involved in sharpening cellular decisions. The LMPP is lymphoid primed progenitor yet it has myeloid potential and appears to play a key role in some AMLs. It could be speculated that in an LMPP/GMP background low a levels of this cluster could lead to myeloid malignancies, whereas high levels in a lymphoid background could lead to lymphoid malignancies. This raises a note of a caution when considering the mir-17-92 cluster as a potential therapeutic target. As altering the levels for treatment of a lymphoid malignancy could increase the chances of a patient developing a myeloid malignancy and vice versa.

To understand what the consequences are of the aberrantly expressed miRNAs in AML (e.g. are they passenger or driver mutations).We must look at how the aberrantly expressed miRNAs fit together with what else is happening in the diseased cells. miRNAs exert their function via binding to their target mRNAs and reducing their translation. So knowing what mRNAs these miRNAs target, is a key bit of

information to understanding what role the aberrantly expressed miRNAs are playing in AML. We initially tried to identify mRNA targets of our list of miRNAs, using a bioinformatic approach combining target prediction software and a previously obtained AML mRNA expression data set. However this approach proved unhelpful, in doing this.

A more effective approach to identify mRNA targets would be to use experimental methods, there currently several different approaches to do this. Firstly gene expression profiling of cells where miRNA of interest is overexpressed or knocked down compared to control cells. This would reveal changes in the levels of different mRNAs resulting from alteration of miRNA level. So if an mRNAs goes up when the miRNA is knocked down or vice versa, it could be a potential target [155]. There are however several issues with this approach; firstly it relies on the miRNA-mRNA target interaction resulting in mRNA degradation. miRNAs may cause a block in translation of their targets or increased degradation of their targets, the exact mechanism of action is still under debate, although there is growing evidence that both mechanism occur. The approach above would only highlight targets that are degraded and so therefore may miss some targets. It is also not possible to determine the primary effects from secondary effects using this methology, making it more difficult to identify real targets. Another method used to experimentally determine mRNA targets is HTS-CLIP (High Throughput Sequencing Cross-Linking Immunoprecipitation) [156]. miRNAs bind to their target mRNAs via the RISC complex, this method exploits this interaction. UV light is used to crosslink miRNAs/mRNAs to proteins, the RNA is then co-immunopreciptated with a component of the RISC complex such as Ago. High through put sequencing is then used to identify which RNAs were interacting with the RISC complex. Combining HTS-CLIP with knock-down or overexpression experiments provides a much more powerful method to identify targets than the previous method and overcomes many of its issues. Another method to identify

targets which also overcomes the problems of the first technique is miR-TRAP [157]. For this method miRNA of interest is labelled with a biotin and psoralen (a highly photo-reactive probe) tag. Cells are subject to UVA radiation which causes the uridine in mRNAs to react with the psoralen on the tagged miRNA. Streptavidin-biotin affinity is exploited to pull down targets pairs and the mRNAs can then be identified via qPCR or sequencing. Using one of the latter two methods to identify mRNA targets of the differentially expressed miRNAs would be a useful future study to further understand miRNAs role in AML.

Understanding how the miRNAs themselves are regulated and where their genes are found in the genome, may reveal why these miRNAs are aberrantly expressed in AML. miRs-30b/d were overexpressed in our AML cells, both of these miRNAs are found on chromosome 8, close to a common translocation site in AML [158]. This may offer a potential cause for the irregularities seen in the expression level of these miRNAs. miRNAs are most commonly found at intragenic and intronic regions of the genome. miRNAs found introns are often regulated by the same mechanisms that regulates the gene whose intron they are located in. Another miRNA we found overexpressed in AML is miR-155 which is located in an intron of Protein tyrosine kinase 2 (PTK2) and is often expressed with PTK2. PTK2 is found to be overexpressed in around 40% of AMLs and is a direct target gene of the common AML translocation protein AML-ETO[159-161]. This may provide an explanation to why this miRNA is overexpressed in around a third of AML samples. Linking the genomic location and regulation of the aberrantly expressed miRNAs to common abnormalities AML may help us understand more about the complex pathways involved in AML.

Functional studies on differentially expressed miRNAs would also be logical direction for future work. As it would improve our understanding of what roles these miRNA have in leukemia and normal hematopoiesis. Functional studies could be performed

by either overexpressing or knocking down miRNAs in normal CD34+ cells or in leukemic LSCs. However performing experiments in normal cells would be a simpler approach; as it is difficult to culture primary AML cells in vitro. The two main methods available to overexpress miRNAs are miRNA mimics or overexpression vectors. miRNA mimics are usually small single stranded chemically modified RNA molecules, that contain miRNA sequence of interest, and behave like endogenous miRNA[162]. Transfection is a common method to get cells to take up mimics, transfection however may be difficult to achieve in primary CD34+ cells. An alternative method that may be more applicable to primary cells is to use lentiviruses to infect cells with overexpression vectors. The vector is integrated into cells genome and miRNA is continuously expressed [163]. However overexpression vectors are much less efficient method than mimics, as many steps are required to generate vector and lentivirus. Caution needs to be taken when overexpressing miRNAs as if they are expressed at level much greater than what is seen endogenously, many off target effects may occur, that would not be seen endogenously.

Knockdown experiments could also be carried out; methods to knockdown miRNAs include miRNA inhibitors and miRNA sponges. Inhibitors are small chemically modified RNAs, that irreversibly bind target miRNAs and therefore block miRNA function [164]. Like mimics these are usually transfected making use with primary cells potentially difficult. An alternative method that can be used with lentiviruses are miRNA sponges. Sponges RNAs contain multiple binding sites for miRNA of interest and therefore mop up miRNAs present in cells, reducing the amount of miRNA to bind to endogenous targets[165, 166]. Recent advances in transfection may allow mimics and inhibitors to be used in CD34+ cells. However if constant knockdown or overexpression is required, such as if cells are to be used in long term experiments (e.g. xenograft experiments) overexpression vectors and sponges would be the best option.

Since I carried out this study many advances have been made in technology and if I was to perform this study again, with the current advances in technology and the hindsight learnt from this study, I would make changes in how it was performed. I would still focus on a wide range of different CD34+ AML samples, but I would ensure that extensive information on the cytogenetic and mutational status of each the samples were known. As this may then help understand what is causing effects, such as the bimodal expression patterns seen for some miRNAs. It would also be useful to investigate if any of the differentially expressed miRNAs were associated with particular phenotypes in AML. I would purify samples in the same way and would critically still use highly purified populations, but I would include more samples from the lymphoid branch, to understand how this fits in the miRNA expression patterns we observed. If I was to repeat the study I would use a different method to isolate RNA, as there have been advances in kits that can extract total RNAs including small RNAs, such as the Norgen biotek kit. This kit can efficiently extract all RNAs including small miRNAs without the use of a phenol step. The reduced efficiency seen in the miRNeasy kit is due to losses occurring at the phenol extraction stage, therefore eradicating this step will increase RNA yield. This would be advantageous as getting enough primary cells needed for RNA yields, particularly for LMPPs, was a major issue in this study. Technologies available to assay miRNA expression have also improved. Advances in qPCR technology from life technology means that profiling miRNAs in large numbers of samples is more efficient. However if I was choose a technology to profile miRNA expression at this current time I would select the Nanostring nCounter system. This technology requires a small amount of RNA 100ng the same as the microarray but the time required in preparing and running samples is less than the microarrays. The Nanostring technology uses probes to bind miRNAs like for microarrays, however each miRNA probe contains a specific barcode, the nCounter then counts how many of each barcode is present. This means that the

read out of this system is a digital read which is the major advantage of this technology over microarrays.

Our primary goal in studying AML is to generate knowledge that may be useful in developing new treatments for AML. My study highlighted a group of miRNAs that could potentially be targeted for therapy. Targeting miRNAs for therapy has become an attractive strategy; with various companies now developing programmes to do this, some of which are now entering clinical trial stages. miR-122 is a miRNA found in the liver, it was discovered that this miRNA is needed for hepatitis C viral replication. An LNA antisense miR-122 inhibitor has been developed has a treatment against hepatitis C and experiments have shown this to be an effective therapy in animal models[167]. The miR-122 inhibitor has now entered clinical trials and was the first miRNA targeted treatment to do so. Using miRNA mimics as treatment to replace down regulated miRNAs is also under development. Let-7 is a tumour suppressor and commonly down regulated in many types of cancer. Studies have shown that let-7 mimics are able to reduce lung tumour growth and are currently under development as a potential cancer therapeutic [168].

Summarizing points:

- miRNAs are aberrantly expressed in AML and therefore may be exploited as a potential new targeted therapy.
- GMP-like AML cells are most similar on an miRNA level to their immunophenotypic normal counterparts GMPs.
- Cytogenetic and patient specific alterations impact on miRNA profiles of leuekmic populations.
- miRNA expression patterns in the assayed populations showed similar patterns to what is seen at an mRNA level.

- miRNAs in AML may be affected by chromosome alterations at a primary or secondary level and are potentially targeted by common AML translocation proteins such as AML-ETO.
- GMP-like LSCs displayed an aberrant miRNA stem cell- like signature.
- Future work on differentially expressed miRNAs in AML should focus on their mRNA targets, how they are being regulated and finally on overexpression and knockdown studies.

References

1. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages.* Nature, 2000. **404**(6774): p. 193-7.
2. Morrison, S.J., et al., *Identification of a lineage of multipotent hematopoietic progenitors.* Development, 1997. **124**(10): p. 1929-39.
3. Reya, T., et al., *Stem cells, cancer, and cancer stem cells.* Nature, 2001. **414**(6859): p. 105-11.
4. Adolfsson, J., et al., *Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity.* Immunity, 2001. **15**(4): p. 659-69.
5. Mansson, R., et al., *Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors.* Immunity, 2007. **26**(4): p. 407-19.
6. Adolfsson, J., et al., *Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment.* Cell, 2005. **121**(2): p. 295-306.
7. Civin, C.I., et al., *Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells.* J Immunol, 1984. **133**(1): p. 157-65.
8. Baum, C.M., et al., *Isolation of a candidate human hematopoietic stem-cell population.* Proceedings of the National Academy of Sciences, 1992. **89**(7): p. 2804-2808.
9. Craig, W., et al., *Expression of Thy-1 on human hematopoietic progenitor cells.* J Exp Med, 1993. **177**(5): p. 1331-42.
10. Murray, L., et al., *Enrichment of human hematopoietic stem cell activity in the CD34+Thy-1+Lin- subpopulation from mobilized peripheral blood.* Blood, 1995. **85**(2): p. 368-78.
11. Bhatia, M., et al., *Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice.* Proceedings of the National Academy of Sciences, 1997. **94**(10): p. 5320-5325.
12. Conneally, E., et al., *Expansion in vitro of transplantable human cord blood stem cells demonstrated using a quantitative assay of their lympho-myeloid repopulating activity in nonobese diabetic-scid/scid mice.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9836-41.
13. Hogan, C.J., E.J. Shpall, and G. Keller, *Differential long-term and multilineage engraftment potential from subfractions of human CD34+ cord blood cells transplanted into NOD/SCID mice.* Proceedings of the National Academy of Sciences, 2002. **99**(1): p. 413-418.
14. Majeti, R., C.Y. Park, and I.L. Weissman, *Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood.* Cell Stem Cell, 2007. **1**(6): p. 635-45.
15. Goardon, N., et al., *Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia.* Cancer Cell, 2011. **19**(1): p. 138-152.
16. Doulatov, S., et al., *Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development.* Nat Immunol, 2010. **11**(7): p. 585-593.
17. Manz, M.G., et al., *Prospective isolation of human clonogenic common myeloid progenitors.* Proceedings of the National Academy of Sciences, 2002. **99**(18): p. 11872-11877.

18. Edvardsson, L., J. Dykes, and T. Olofsson, *Isolation and characterization of human myeloid progenitor populations--TpoR as discriminator between common myeloid and megakaryocyte/erythroid progenitors.* Exp Hematol, 2006. **34**(5): p. 599-609.
19. Dohner, H., et al., *Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet.* Blood, 2010. **115**(3): p. 453-74.
20. Estey, E. and H. Döhner, *Acute myeloid leukaemia.* The Lancet, 2006. **368**(9550): p. 1894-1907.
21. Shipley, J.L. and J.N. Butera, *Acute myelogenous leukemia.* Exp Hematol, 2009. **37**(6): p. 649-58.
22. Dick, J.E., *Looking ahead in cancer stem cell research.* Nat Biotechnol, 2009. **27**(1): p. 44-6.
23. Ishikawa, F., et al., *Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region.* Nat Biotech, 2007. **25**(11): p. 1315-1321.
24. Bennett, J.M., et al., *Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group.* Br J Haematol, 1976. **33**(4): p. 451-8.
25. Seiter, K.J., E Harris (May 20, 2011). "Acute Myeloid Leukemia Staging". Retrieved 26 August 2011.
26. Vardiman, J.W., et al., *The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes.* Blood, 2009. **114**(5): p. 937-51.
27. Betz, B.L. and J.L. Hess, *Acute myeloid leukemia diagnosis in the 21st century.* Arch Pathol Lab Med, 2010. **134**(10): p. 1427-33.
28. Mrozek, K., N.A. Heerema, and C.D. Bloomfield, *Cytogenetics in acute leukemia.* Blood Rev, 2004. **18**(2): p. 115-36.
29. Grimwade, D., et al., *The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial.* Blood, 2001. **98**(5): p. 1312-20.
30. Slovak, M.L., et al., *Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study.* Blood, 2000. **96**(13): p. 4075-83.
31. Grimwade, D., et al., *The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties.* Blood, 1998. **92**(7): p. 2322-33.
32. Byrd, J.C., et al., *Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461).* Blood, 2002. **100**(13): p. 4325-36.
33. Marcucci, G., T. Haferlach, and H. Dohner, *Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications.* J Clin Oncol, 2011. **29**(5): p. 475-86.
34. Mrozek, K., et al., *Molecular signatures in acute myeloid leukemia.* Curr Opin Hematol, 2009. **16**(2): p. 64-9.
35. Gilliland, D.G., C.T. Jordan, and C.A. Felix, *The molecular basis of leukemia.* Hematology Am Soc Hematol Educ Program, 2004: p. 80-97.
36. Jan, M. and R. Majeti, *Clonal evolution of acute leukemia genomes.* Oncogene, 2012. **20**(10): p. 48.
37. Graubert, T.A. and E.R. Mardis, *Genomics of acute myeloid leukemia.* Cancer J, 2011. **17**(6): p. 487-91.

38. Taussig, D.C., et al., *Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction.* Blood, 2010. **115**(10): p. 1976-84.
39. Clarke, M.F., et al., *Cancer Stem Cells—Perspectives on Current Status and Future Directions: AACR Workshop on Cancer Stem Cells.* Cancer Research, 2006. **66**(19): p. 9339-9344.
40. O'Brien, C.A., A. Kreso, and C.H.M. Jamieson, *Cancer Stem Cells and Self-renewal.* Clinical Cancer Research, 2010. **16**(12): p. 3113-3120.
41. Al-Hajj, M., et al., *Prospective identification of tumorigenic breast cancer cells.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3983-8.
42. Singh, S.K., et al., *Identification of human brain tumour initiating cells.* Nature, 2004. **432**(7015): p. 396-401.
43. Ricci-Vitiani, L., et al., *Identification and expansion of human colon-cancer-initiating cells.* Nature, 2007. **445**(7123): p. 111-5.
44. Quintana, E., et al., *Efficient tumour formation by single human melanoma cells.* Nature, 2008. **456**(7222): p. 593-598.
45. Shackleton, M., et al., *Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution.* Cell, 2009. **138**(5): p. 822-829.
46. Clarke, M.F., et al., *Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells.* Cancer Res, 2006. **66**(19): p. 9339-44.
47. Dick, J.E., *Stem cell concepts renew cancer research.* Blood, 2008. **112**(13): p. 4793-807.
48. Hope, K.J., L. Jin, and J.E. Dick, *Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity.* Nat Immunol, 2004. **5**(7): p. 738-43.
49. Horton, S.J. and B.J. Huntly, *Recent advances in acute myeloid leukemia stem cell biology.* Haematologica, 2012. **97**(7): p. 966-74.
50. Krause, D.S. and R.A. Van Etten, *Right on target: eradicating leukemic stem cells.* Trends Mol Med, 2007. **13**(11): p. 470-81.
51. Huntly, B.J. and D.G. Gilliland, *Leukaemia stem cells and the evolution of cancer-stem-cell research.* Nat Rev Cancer, 2005. **5**(4): p. 311-21.
52. Saito, Y., et al., *Identification of therapeutic targets for quiescent, chemotherapy-resistant human leukemia stem cells.* Sci Transl Med, 2010. **2**(17): p. 3000349.
53. Lapidot, T., et al., *A cell initiating human acute myeloid leukaemia after transplantation into SCID mice.* Nature, 1994. **367**(6464): p. 645-8.
54. 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.
55. Cozzio, A., et al., *Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors.* Genes Dev, 2003. **17**(24): p. 3029-35.
56. Huntly, B.J.P., et al., *MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors.* Cancer Cell, 2004. **6**(6): p. 587-596.
57. Krivtsov, A.V., et al., *Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9.* Nature, 2006. **442**(7104): p. 818-22.
58. Taussig, D.C., et al., *Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells.* Blood, 2008. **112**(3): p. 568-75.
59. Eppert, K., et al., *Stem cell gene expression programs influence clinical outcome in human leukemia.* Nat Med, 2011. **17**(9): p. 1086-1093.
60. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.

61. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.
62. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.
63. Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II.* EMBO J, 2004. **23**(20): p. 4051-60.
64. Cai, X., C.H. Hagedorn, and B.R. Cullen, *Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs.* RNA, 2004. **10**(12): p. 1957-66.
65. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-419.
66. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex.* Nature, 2004. **432**(7014): p. 231-235.
67. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs.* Nature, 2004. **432**(7014): p. 235-240.
68. BOHNSACK, M.T., K. KEVIN CZAPLINSKI, and D. GÖRLICH, *Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs.* RNA., 2004. **10**: p. 185–191.
69. Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* EMBO J, 2002. **21**(17): p. 4663-4670.
70. Ketting, R.F., et al., *Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans.* Genes & Development, 2001. **15**(20): p. 2654-2659.
71. Miyoshi, K., et al., *Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila.* RNA, 2010. **16**(3): p. 506-15.
72. Ouellet, D.L., et al., *MicroRNAs in gene regulation: when the smallest governs it all.* J Biomed Biotechnol, 2006. **2006**(4): p. 69616.
73. Gu, S. and M. Kay, *How do miRNAs mediate translational repression?* Silence, 2010. **1**(1): p. 11.
74. Griffiths-Jones, S., *The microRNA Registry.* Nucleic Acids Research, 2004. **32**(suppl 1): p. D109-D111.
75. ; Available from: http://www.mirbase.org/.
76. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets.* Cell, 2005. **120**(1): p. 15-20.
77. Xie, X., et al., *Systematic discovery of regulatory motifs in human promoters and 3[prime] UTRs by comparison of several mammals.* Nature, 2005. **434**(7031): p. 338-345.
78. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic Acids Research, 2006. **34**(suppl 1): p. D140-D144.
79. Buza-Vidas, N., et al., *Dicer is selectively important for the earliest stages of erythroid development.* Blood, 2012. **120**(12): p. 2412-6.
80. Cobb, B.S., et al., *A role for Dicer in immune regulation.* J Exp Med, 2006. **203**(11): p. 2519-27.
81. O'Carroll, D., et al., *A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway.* Genes Dev, 2007. **21**(16): p. 1999-2004.
82. Guo, S., et al., *MicroRNA miR-125a controls hematopoietic stem cell number.* Proceedings of the National Academy of Sciences, 2010. **107**(32): p. 14229-14234.
83. Bissels, U., A. Bosio, and W. Wagner, *MicroRNAs are shaping the hematopoietic landscape.* Haematologica, 2012. **97**(2): p. 160-167.

84. Ooi, A.G., et al., *MicroRNA-125b expands hematopoietic stem cells and enriches for the lymphoid-balanced and lymphoid-biased subsets.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21505-10.
85. Chaudhuri, A.A., et al., *Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A.* Proceedings of the National Academy of Sciences, 2012. **109**(11): p. 4233-4238.
86. Fazi, F., et al., *A Minicircuitry Comprised of MicroRNA-223 and Transcription Factors NFI-A and C/EBPα Regulates Human Granulopoiesis.* Cell, 2005. **123**(5): p. 819-831.
87. Johnnidis, J.B., et al., *Regulation of progenitor cell proliferation and granulocyte function by microRNA-223.* Nature, 2008. **451**(7182): p. 1125-9.
88. Boldin, M.P., et al., *miR-146a is a significant brake on autoimmunity, myeloproliferation, and cancer in mice.* J Exp Med, 2011. **208**(6): p. 1189-201.
89. Hou, J., et al., *MicroRNA-146a Feedback Inhibits RIG-I-Dependent Type I IFN Production in Macrophages by Targeting TRAF6, IRAK1, and IRAK2.* The Journal of Immunology, 2009. **183**(3): p. 2150-2158.
90. Nahid, M.A., et al., *miR-146a Is Critical for Endotoxin-induced Tolerance: IMPLICATION IN INNATE IMMUNITY.* Journal of Biological Chemistry, 2009. **284**(50): p. 34590-34599.
91. Taganov, K.D., et al., *NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses.* Proc Natl Acad Sci U S A, 2006. **103**(33): p. 12481-6.
92. Felli, N., et al., *MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation.* Proc Natl Acad Sci U S A, 2005. **102**(50): p. 18081-6.
93. Lu, J., et al., *MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors.* Dev Cell, 2008. **14**(6): p. 843-53.
94. O'Connell, R.M., J.L. Zhao, and D.S. Rao, *MicroRNA function in myeloid biology.* Blood, 2011. **118**(11): p. 2960-2969.
95. Ventura, A., et al., *Targeted Deletion Reveals Essential and Overlapping Functions of the miR-17<92 Family of miRNA Clusters.* Cell, 2008. **132**(5): p. 875-886.
96. Xiao, C., et al., *MiR-150 Controls B Cell Differentiation by Targeting the Transcription Factor c-Myb.* Cell, 2007. **131**(1): p. 146-159.
97. Bissels, U., A. Bosio, and W. Wagner, *MicroRNAs are shaping the hematopoietic landscape.* Haematologica, 2012. **97**(2): p. 160-7.
98. Visone, R. and C.M. Croce, *MiRNAs and cancer.* Am J Pathol, 2009. **174**(4): p. 1131-8.
99. Sassen, S., E. Miska, and C. Caldas, *MicroRNA—implications for cancer.* Virchows Archiv, 2008. **452**(1): p. 1-10.
100. Calin, G.A., et al., *Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.* Proceedings of the National Academy of Sciences, 2002. **99**(24): p. 15524-15529.
101. Boyerinas, B., et al., *The role of let-7 in cell differentiation and cancer.* Endocr Relat Cancer, 2010. **17**(1): p. F19-36.
102. Johnson, S.M., et al., *RAS Is Regulated by the let-7 MicroRNA Family.* cell, 2005. **120**(5): p. 635-647.
103. He, L., et al., *A microRNA polycistron as a potential human oncogene.* Nature, 2005. **435**(7043): p. 828-33.
104. Mi, S., et al., *MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia.* Proceedings of the National Academy of Sciences, 2007. **104**(50): p. 19971-19976.

105. 10 Garzon, R., et al., *Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin.* Proceedings of the National Academy of Sciences, 2008. **105**(10): p. 3945-3950.
106. Marcucci, G., et al., *IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study.* J Clin Oncol, 2010. **28**(14): p. 2348-55.
107. 40 Whitman, S.P., et al., *FLT3 internal tandem duplication associates with adverse outcome and gene- and microRNA-expression signatures in patients 60 years of age or older with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study.* Blood, 2010. **116**(18): p. 3622-3626.
108. 33 Metzeler, K.H., et al., *TET2 Mutations Improve the New European LeukemiaNet Risk Classification of Acute Myeloid Leukemia: A Cancer and Leukemia Group B Study.* Journal of Clinical Oncology, 2011. **29**(10): p. 1373-1381.
109. 39 Schwind, S., et al., *BAALC and ERG expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study.* Blood, 2010. **116**(25): p. 5660-5669.
110. 9 Li, Z., et al., *Distinct microRNA expression profiles in acute myeloid leukemia with common translocations.* Proceedings of the National Academy of Sciences, 2008. **105**(40): p. 15535-15540.
111. Jongen-Lavrencic, M., et al., *MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia.* Blood, 2008. **111**(10): p. 5078-85.
112. Dixon-McIver, A., et al., *Distinctive patterns of microRNA expression associated with karyotype in acute myeloid leukaemia.* PLoS One, 2008. **3**(5): p. 0002141.
113. Garzon, R., et al., *MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia.* Blood, 2008. **111**(6): p. 3183-3189.
114. Garzon, R., et al., *MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia.* Blood, 2008. **111**(6): p. 3183-9.
115. 4 Han, Y.-C., et al., *microRNA-29a induces aberrant self-renewal capacity in hematopoietic progenitors, biased myeloid development, and acute myeloid leukemia.* The Journal of Experimental Medicine, 2010. **207**(3): p. 475-489.
116. Popovic, R., et al., *Regulation of mir-196b by MLL and its overexpression by MLL fusions contributes to immortalization.* Blood, 2009. **113**(14): p. 3314-22.
117. O'Connell, R.M., et al., *Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder.* J Exp Med, 2008. **205**(3): p. 585-94.
118. 49 Garzon, R., et al., *MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia.* Blood, 2008. **111**(6): p. 3183-3189.
119. Cammarata, G., et al., *Differential expression of specific microRNA and their targets in acute myeloid leukemia.* Am J Hematol, 2010. **85**(5): p. 331-9.
120. Chung, S.S., W. Hu, and C.Y. Park, *The role of microRNAs in hematopoietic stem cell and leukemic stem cell function.* Therapeutic Advances in Hematology, 2011. **2**(5): p. 317-334.
121. 53 Gao, X.-n., et al., *MicroRNA-193b regulates c-Kit proto-oncogene and represses cell proliferation in acute myeloid leukemia.* Leukemia Research, 2011. **35**(9): p. 1226-1232.
122. 36 Vázquez, I., et al., *Silencing of hsa-miR-124 by EVI1 in cell lines and patients with acute myeloid leukemia.* Proceedings of the National Academy of Sciences, 2010. **107**(44): p. E167-E168.
123. Ihaka, R. and R. Gentleman, *R: A Language for Data Analysis and Graphics.* Journal of Computational and Graphical Statistics, 1996. **5**(3): p. 299-314.

124. Lopez-Romero, P., *Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library.* BMC Genomics, 2011. **12**(64): p. 1471-2164.
125. Saeed, A.I., et al., *TM4 microarray software suite.* Methods Enzymol, 2006. **411**: p. 134-93.
126. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15545-15550.
127. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.
128. Jang, J.S., et al., *Quantitative miRNA expression analysis using fluidigm microfluidics dynamic arrays.* BMC Genomics, 2011. **12**(144): p. 1471-2164.
129. Jan, M., et al., *Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker.* Proceedings of the National Academy of Sciences, 2011.
130. Bakker, A.B., et al., *C-type lectin-like molecule-1: a novel myeloid cell surface marker associated with acute myeloid leukemia.* Cancer Res, 2004. **64**(22): p. 8443-50.
131. Schendel, D.J., *Is it time to abandon RHAMM/HMMR as a candidate antigen for immunotherapy of acute myeloid leukemia?* Haematologica, 2012. **97**(10): p. 1454-1455.
132. Greiner, J., et al., *mRNA expression of leukemia-associated antigens in patients with acute myeloid leukemia for the development of specific immunotherapies.* Int J Cancer, 2004. **108**(5): p. 704-11.
133. Greiner, J., et al., *Receptor for hyaluronan acid-mediated motility (RHAMM) is a new immunogenic leukemia-associated antigen in acute and chronic myeloid leukemia.* Exp Hematol, 2002. **30**(9): p. 1029-35.
134. Anguille, S., V.F. Van Tendeloo, and Z.N. Berneman, *Leukemia-associated antigens and their relevance to the immunotherapy of acute myeloid leukemia.* Leukemia, 2012. **26**(10): p. 2186-96.
135. Snauwaert, S., et al., *RHAMM/HMMR (CD168) is not an ideal target antigen for immunotherapy of acute myeloid leukemia.* Haematologica, 2012. **97**(10): p. 1539-47.
136. Ach, R., H. Wang, and B. Curry, *Measuring microRNAs: Comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods.* BMC Biotechnology, 2008. **8**(1): p. 69.
137. Ng, Y.Y., et al., *Gene-expression profiling of CD34+ cells from various hematopoietic stem-cell sources reveals functional differences in stem-cell activity.* J Leukoc Biol, 2004. **75**(2): p. 314-23.
138. Alicia Báez1*, B.M.-A., Ph.D.2*, Concepción Prats-Martín, M.D1*, Isabel Álvarez-Laderas1*, María Victoria Barbado, Ph.D1*, Teresa Caballero-Velázquez1*, Jose Ignacio Piruat, Ph.D1*, Magdalena Carmona, MD1*, Jose Antonio Pérez-Simón Sr., MD, PhD1* and Alvaro Urbano-Ispizua, MD/Ph.D, *Mirnas and Gene Expression Profiles in CD34+ Cells Are Dependent On the Source of Progenitor Cells Employed in Transplantation* in *ASH*2012.
139. 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.
140. Liu, B., et al., *MiR-26a enhances metastasis potential of lung cancer cells via AKT pathway by targeting PTEN.* Biochim Biophys Acta, 2012. **11**(704): p. 4.
141. Huse, J.T., et al., *The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo.* Genes Dev, 2009. **23**(11): p. 1327-37.

142. Lu, J., et al., *MiR-26a Inhibits Cell Growth and Tumorigenesis of Nasopharyngeal Carcinoma through Repression of EZH2.* Cancer Research, 2011. **71**(1): p. 225-233.
143. Liu, X.X., et al., *MicroRNA-26b is underexpressed in human breast cancer and induces cell apoptosis by targeting SLC7A11.* FEBS Lett, 2011. **585**(9): p. 1363-7.
144. Olive, V., I. Jiang, and L. He, *mir-17-92, a cluster of miRNAs in the midst of the cancer network.* Int J Biochem Cell Biol, 2010. **42**(8): p. 1348-54.
145. Git, A., et al., *Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression.* RNA, 2010. **16**(5): p. 991-1006.
146. Marcucci, G., et al., *The prognostic and functional role of microRNAs in acute myeloid leukemia.* Blood, 2011. **117**(4): p. 1121-1129.
147. Tanzer, A. and P.F. Stadler, *Molecular evolution of a microRNA cluster.* Journal of Molecular Biology, 2004. **339**(2): p. 327-335.
148. Ota, A., et al., *Identification and Characterization of a Novel Gene, C13orf25, as a Target for 13q31-q32 Amplification in Malignant Lymphoma.* Cancer Research, 2004. **64**(9): p. 3087-3095.
149. Jin, H.Y., et al., *MicroRNA-17[sim]92 plays a causative role in lymphomagenesis by coordinating multiple oncogenic pathways.* EMBO J, 2013. **advance online publication**.
150. Hayashita, Y., et al., *A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation.* Cancer Res, 2005. **65**(21): p. 9628-32.
151. Li, H., et al., *miR-17-5p promotes human breast cancer cell migration and invasion through suppression of HBP1.* Breast Cancer Res Treat, 2011. **126**(3): p. 565-75.
152. Yu, Z., et al., *A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation.* J Cell Biol, 2008. **182**(3): p. 509-17.
153. Zhang, L., et al., *microRNAs exhibit high frequency genomic alterations in human cancer.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 9136-41.
154. Ventura, A., et al., *Targeted Deletion Reveals Essential and Overlapping Functions of the miR-17~92 Family of miRNA Clusters.* Cell, 2008. **132**(5): p. 875-886.
155. Lim, L.P., et al., *Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.* Nature, 2005. **433**(7027): p. 769-73.
156. Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.* Nature, 2009. **460**(7254): p. 479-86.
157. Baigude, H., et al., *miR-TRAP: a benchtop chemical biology strategy to identify microRNA targets.* Angew Chem Int Ed Engl, 2012. **51**(24): p. 5880-3.
158. 2013; Available from: http://www.ensembl.org/Homo_sapiens/Location/View?l=8:135810763-135814850;r=8:135810763-135814850.
159. Ding, J., et al., *Gain of miR-151 on chromosome 8q24.3 facilitates tumour cell migration and spreading through downregulating RhoGDIA.* Nat Cell Biol, 2010. **12**(4): p. 390-399.
160. Recher, C., et al., *Expression of focal adhesion kinase in acute myeloid leukemia is associated with enhanced blast migration, increased cellularity, and poor prognosis.* Cancer Res, 2004. **64**(9): p. 3191-7.
161. Prall, S., et al., *Focal adhesion kinase (FAK) in t(8;21) rearranged acute myeloid leukaemia (AML).* Klin Padiatr, 2012. **224**(03): p. A16.
162. Chorn, G., et al., *Single-stranded microRNA mimics.* RNA, 2012. **18**(10): p. 1796-804.
163. Bousquet, M., et al., *MicroRNA miR-125b causes leukemia.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21558-63.

164. Krutzfeldt, J., et al., *Silencing of microRNAs in vivo with /`antagomirs/'.* Nature, 2005. **438**(7068): p. 685-689.
165. Ebert, M.S., J.R. Neilson, and P.A. Sharp, *MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells.* Nat Methods, 2007. **4**(9): p. 721-6.
166. Loya, C.M., et al., *Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms.* Nat Methods, 2009. **6**(12): p. 897-903.
167. Lanford, R.E., et al., *Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection.* Science, 2010. **327**(5962): p. 198-201.
168. Esquela-Kerscher, A., et al., *The let-7 microRNA reduces tumor growth in mouse models of lung cancer.* Cell Cycle, 2008. **7**(6): p. 759-64.