Characterizing the functional role of argonaute protein PIWIL4 in acute myeloid leukemia

Thesis submitted to obtain a doctoral degree in Human Biology at the University of Ulm, Germany

Submitted by

Shiva Bamezai

New Delhi, India

2015
Dean – Prof Dr. Thomas Wirth

1st Supervisor – Prof. Dr. Christian Buske

2nd Supervisor – Prof. Hartmut Geiger

Date of examination – 27.02.2015
For Ramesh and Gita
# Index

## Contents

<table>
<thead>
<tr>
<th>List of abbreviations</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>III</td>
</tr>
</tbody>
</table>

## 1. Introduction

1.1 Hematopoiesis

1.1.1 Self Renewal

1.1.2 The hematopoietic hierarchy

1.1.3 Transcription factors and hematopoietic cell fate restriction

1.1.4 Epigenetics and Hematopoiesis

1.2 Acute Myeloid Leukemia

1.2.1 Genetics of AML

1.2.2 Epigenetics and AML

1.3 PIWI proteins

1.3.1 The Piwi sub family

1.3.2 Piwi proteins structure

1.3.3 PIWI and PIWI interacting RNA

1.3.4 PIWI-piRNA axis – TE silencing and beyond

1.3.5 PIWI proteins are crucial for self-renewal

1.3.6 PIWI proteins as epigenetic modifiers

1.3.7 PIWIL4 and normal hematopoiesis

1.3.8 PIWIL4 and cancer

1.4 Aim of study

## 2. Materials and Methods

## 3. Results

3.1 Piwi genes are expressed in human bone marrow

3.2 PIWIL4 is highly and ubiquitously expressed in human hematopoietic stem-progenitors and differentiated cells
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3 PIWIL2 expression is not significantly deregulated in AML</td>
<td>49</td>
</tr>
<tr>
<td>3.4 PIWIL4 expression is significantly deregulated in AML</td>
<td>50</td>
</tr>
<tr>
<td>3.5 PIWIL4 expression is highly deregulated in MLL translocation</td>
<td>52</td>
</tr>
<tr>
<td>harboring patients</td>
<td></td>
</tr>
<tr>
<td>3.6 PIWIL4 expression is associated with MLL-AF9 AML</td>
<td>53</td>
</tr>
<tr>
<td>3.7 PIWIL4 shows a broad range of expression in different MLL leukemia</td>
<td>54</td>
</tr>
<tr>
<td>3.8 Knockdown of PIWIL4 results in loss of PIWIL4 protein from nuclear</td>
<td>56</td>
</tr>
<tr>
<td>and cytoplasmic compartments</td>
<td></td>
</tr>
<tr>
<td>3.9 Depletion of PIWIL4 drastically reduces the growth of MLL</td>
<td>57</td>
</tr>
<tr>
<td>rearrangement harboring AML cell lines in vitro</td>
<td></td>
</tr>
<tr>
<td>3.10 PIWIL4 knockdown adversely effects growth of AML cells in vivo</td>
<td>59</td>
</tr>
<tr>
<td>3.11 PIWIL4 knockdown effects cell cycle but not apoptosis</td>
<td>60</td>
</tr>
<tr>
<td>3.12 Depletion of PIWIL4 in human cord blood HSPCs does not impact</td>
<td>61</td>
</tr>
<tr>
<td>in vitro growth</td>
<td></td>
</tr>
<tr>
<td>3.13 Depletion of PIWIL4 in primary MLL-AF9 patient BM drastically</td>
<td>63</td>
</tr>
<tr>
<td>effects colony forming ability in vitro</td>
<td></td>
</tr>
<tr>
<td>3.14 Depletion of PIWIL4 leads to global loss of H3K9me3 and localized</td>
<td>64</td>
</tr>
<tr>
<td>gain in H3K4me3</td>
<td></td>
</tr>
<tr>
<td>3.15 PAZ domain lacking mutant of PIWIL4 is unable to rescue the loss</td>
<td>67</td>
</tr>
<tr>
<td>of proliferative potential and H3K9me3 marks</td>
<td></td>
</tr>
<tr>
<td>3.16 Depletion of PIWIL4 leads to differential gene expression</td>
<td>69</td>
</tr>
<tr>
<td>3.17 Changes in histone marks in promoter/regulatory regions show</td>
<td>72</td>
</tr>
<tr>
<td>low correlation with gene expression changes</td>
<td></td>
</tr>
<tr>
<td>3.18 Over-expression of PIWIL4 in THP-1 and RAJI cell lines leads</td>
<td>74</td>
</tr>
<tr>
<td>to increased proliferation in vitro</td>
<td></td>
</tr>
<tr>
<td>3.19 piRNA like RNA is expressed in AML cells - THP-1 and MV4-11</td>
<td>75</td>
</tr>
<tr>
<td>3.20 PIWIL4 associates with epigenetically active proteins and</td>
<td>81</td>
</tr>
<tr>
<td>proteins involved in piRNA biogenesis in AML cells</td>
<td></td>
</tr>
</tbody>
</table>

4. Discussion                                                      83
5. Summary                                                          91
6. References                                                       93
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM</td>
<td>bone marrow</td>
</tr>
<tr>
<td>ALL</td>
<td>acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>AML</td>
<td>acute myeloid leukemia</td>
</tr>
<tr>
<td>AA</td>
<td>amino acid</td>
</tr>
<tr>
<td>bp</td>
<td>basepairs</td>
</tr>
<tr>
<td>BMCD34*</td>
<td>Bone marrow CD34 positive cells</td>
</tr>
<tr>
<td>BMNC</td>
<td>Bone marrow mononuclear cells</td>
</tr>
<tr>
<td>ChIP</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>chromatin immunoprecipitation coupled with high-throughput DNA sequencing</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CB</td>
<td>Cord blood</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>Ct</td>
<td>Threshold cycle</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosin-guanine dinucleotide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic-acid</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethylpyrocarbonate</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle Medium</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>GMP</td>
<td>granulo-monocyte progenitor</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>Histone 3 tri methyl lysine 9</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Histone 3 tri methyl lysine 4</td>
</tr>
<tr>
<td>H4K20me3</td>
<td>Histone 4 tri methyl lysine 20</td>
</tr>
</tbody>
</table>
HSC  hematopoietic stem cell
IP   Immunoprecipitation
Lin  Lineage
LC/MS Liquid chromatography/mass spectrometry
LT-HSC long term HSC
MEP  megakaryocyte erythroid progenitor cells
MNC  mono nuclear cells
MLP  multilymphoid progenitor
MPP  multipotent progenitor
PB   peripheral blood
PBS  phosphate buffered saline
PCR  polymerase chain reaction
p    probability of obtaining a test statistic equal to the actually observed one
PIWI P-element induced wimpy testis
piRNA Piwi interacting RNA
q    quantitative
RT-PCR reverse-transcription PCR
RNA  ribonucleic-acid
RNAi RNA interference
RPMI Roswell Park Memorial Institute medium
WBC  white blood cells
1. **INTRODUCTION**

Blood is a highly regenerative connective tissue where approximately $10^{12}$ cells arise daily from the bone marrow of an adult human. The generation of the various cellular components of blood is known as hematopoiesis.

### 1.1 Hematopoiesis

Hematopoiesis, the life-long production, multiplication and specialization of blood cells, gives rise to the functionally mature cells which perform various functions required for human survival. Red blood cells (RBCs) or erythrocytes carry oxygen to all organs through the circulatory system. The white blood cells (WBCs) or Leukocytes can be further divided into lymphocytes and myelocytes. Lymphocytes, consisting of B and T cells, are the cornerstone of the adaptive immune system. Myelocytes are involved in diverse roles such as innate and adaptive immune system and blood clotting. These specialized cells are all derived from primitive cells that exist in a hierarchy with hematopoietic stem cells (HSCs) lying at the very top of the hierarchical system. In other words, the onus for the lifelong production of the various cellular components of blood lies with hematopoietic stem cells or HSCs, rare cells residing in the medulla of bones. The existence of the hematopoietic hierarchy (as shown later in figure1), relies on two unique properties of HSCs i.e. self-renewal and differentiation.

#### 1.1.1 Self-Renewal and differentiation

Differentiation of HSCs is the process by which stem cells become committed and restricted to a functional fate. Since HSCs have to produce functionally mature cells for the entire lifetime of an organism, to prevent exhaustion of the number of HSCs in an individual, differentiation into functionally mature cells co-
exists in a balance with the maintenance and expansion of HSCs. The maintenance and expansion of HSCs is performed through a unique property of stem cells known as self-renewal which allows HSCs to either maintain their cell numbers through asymmetric divisions leading to generation of a daughter stem cell and a short term (ST)-HSC with limited self-renewal potential (maintenance) or through symmetric divisions that give rise to two daughter stem cells (expansion) (Weissman 2000).

1.1.2 The hematopoietic hierarchy

Surface marker identification and clonal analysis have played a crucial role in identifying and understanding HSC characteristics and function. On one hand, simultaneous detection of several independent cell surface markers is required for purification of human HSCs. For example, CD34, expressed on less than 5% of all blood cells is in fact present on 99% of all HSCs (Doulatov, Notta et al. 2012). Human HSCs are also positive for the markers CD93, CD90 and integrin CD49f while being negative for the marker CD38 and CD45RA (figure 1). On the other hand, clonal analysis using transplantation assays in conditioned hosts such as mice, functionally tests HSC property of self-renewal and differentiation into all cell types that compose the hematopoietic tissue (Doulatov, Notta et al. 2012). Both these tools have been used to further dissect HSCs in order to identify a hierarchy within the HSC pool. While most human stem cell and progenitor activity is associated with expression of the surface marker CD34, recent studies have pointed to the existence of self-renewing Lin⁻ CD34⁻ CD38⁻ CD93<sup>hi</sup> cells that represent an immature and quiescent human HSC population which could be placed above CD34<sup>+</sup> HSCs in the hematopoietic hierarchy (Anjos-Afonso, Currie et al. 2013; Anjos-Afonso and Bonnet 2014).
Below HSCs in the hematopoietic hierarchy lie the multipotent intermediates or MPP (multi potent progenitor) which in turn gives rise to the first lineage-committed progenitors, which lack the ability to self-renew. The lineage committed progenitors i.e. the CMP (common myeloid progenitor) and MLP (multi lymphoid progenitor), through several further differentiated intermediates, progressively undergo fate restriction to finally develop into mature blood cells.
that lie at the bottom of the hierarchical system, such as - myeloid cells (granulocytes, monocytes), lymphoid cells (B-cells T-cells and NK cells), erythrocytes and thrombocytes (Morrison, Uchida et al. 1995; Doulatov, Notta et al. 2012).

1.1.3 Transcription factors and hematopoietic cell fate restriction

HSCs, progenitors and mature hematopoietic cells differ from each other in their respective gene expression profiles. The determination of hematopoietic cell fate is determined by tightly regulated transcription factors that govern cell fate through changes in gene expression (Domen and Weissman 1999; Krause 2002; Leeanansaksiri and Dechsukhum 2006).

Transcription factors such as members of the HOX gene family, a family of homeodomain-containing transcription factors, are expressed in HSC enriched subpopulations and immature progenitor compartments and downregulated during differentiation in both mice and humans. (Giampaolo, Sterpetti et al. 1994; Moretti, Simmons et al. 1994; Sauvageau, Lansdorp et al. 1994; Kawagoe, Humphries et al. 1999; Pineault, Helgason et al. 2002). Extensive investigation in mouse models through engineered overexpression from retroviral vectors, and transgenic knockout and knock-in studies has led to the assertion that Hox genes such as Hoxb4 (Antonchuk, Sauvageau et al. 2002), Hoxa9 (Lawrence, Helgason et al. 1997) and others (Shen, Detmer et al. 1992; Fuller, McAdara et al. 1999; Shimamoto, Tang et al. 1999; Bjornsson, Larsson et al. 2003), are essential for normal hematopoiesis. Moreover, Hox co-factors such Meis1 and Pbx1 are also essential for hematopoiesis. Pbx1 null mice die during the embryonic stage as a result of severe hematopoietic defects (DiMartino, Selleri et al. 2001) and Meis1 deficient mice fail to generate megakaryocytes, exhibit severe hemorrhaging and die during the embryonic stage (Hisa, Spence et al. 2004).

Another transcription factor family necessary for normal hematopoiesis is the GATA family of zinc-finger transcription factors. GATA-2 plays a critical role in maintaining the pool of multipotent progenitors and HSC, both during embryogenesis and in the adult (Tsai, Keller et al. 1994). GATA-1 and its
transcriptional cofactor called Friend of GATA-1 (FOG-1) have been found to be essential for erythroid and megakaryocytic differentiation (Pevny, Simon et al. 1991; Tsang, Fujiwara et al. 1998).

PU.1, a member of the Ets family of transcription factors, is essential in the development of cells of the monocytic, granulocytic and lymphoid lineages (Scott, Simon et al. 1994).

1.1.4 Epigenetics and Hematopoiesis

The regulation of gene expression through transcription factors is also tightly coordinated. The binding of transcription factors to DNA is governed in part by the accessibility of the DNA to the transcription factor, while transcription relies on initiation and elongation. Epigenetic changes have a major influence on transcription by way of regulating accessibility to DNA through condensation of chromatin, through regulation of initiation, and posing as an obstacle to elongation and preventing cryptic transcription. In other words, epigenetic mechanisms control gene transcription (Briggs, Bryk et al. 2001; Heard, Rougeulle et al. 2001; Jenuwein and Allis 2001; Ma, Baumann et al. 2001; Nakayama, Rice et al. 2001; Rice and Allis 2001; Smith, Allis et al. 2001; Wang, Huang et al. 2001; Boggs, Cheung et al. 2002; Irvine, Lin et al. 2002). In stem cells, self-renewal and/or differentiation are accompanied by, and are often preceded by, epigenetic changes in gene regulatory regions marking them as active, silent, or poised (addressed in detail in section 1.4.1.3) (Bernstein, Mikkelsen et al. 2006; Meissner, Mikkelsen et al. 2008; Ji, Ehrlich et al. 2010; Kim, Doi et al. 2010). In hematopoiesis, epigenetic mechanisms help drive changes in gene expression that accompany the transition from hematopoietic stem cells to terminally differentiated blood cells (Ji, Ehrlich et al. 2010; Kim, Doi et al. 2010; Prasad, Ronnerblad et al. 2014; Ronnerblad, Andersson et al. 2014).

1.1.4.1 DNA Methylation – writers and erasers
The transfer of a methyl moiety from an S-adenosylmethionine (SAM) to a cytosine, or less commonly adenine, is known DNA methylation. This reaction is catalyzed by 'writers' of this modification - DNA methyl transferases or DNMTs. In humans, 4 DNMTS are known: DNMT1, DNMT2, DNMT3A and DNMT3B (Holliday and Pugh 1975). DNMT1 is classically known to be a maintenance methylation enzyme, copying the methylation profile of parent cells to its daughter cells while DNMT3A and 3B are known to be de novo methyltransferases. In addition, non-enzymatic proteins required for the function of these enzymes have also been identified in humans – these are DNMT3L and UHRF1.

DNA methylation in mammals occurs mainly at CpG dinucleotides and is conventionally associated with regulation of gene expression and genomic stability through repression of chromatin. In light of this, it comes as no surprise that most of the CpG methylation in the genome is found in repetitive regions, rich in dormant transposable elements (TEs) and endogenous retroviruses.

Conversely, DNA demethylation is a process by which the methyl moiety is passively or actively ‘erased’ from cytosine bases. Passively, DNA is demethylated in the absence of maintenance methylation during several rounds of replication. But recently, a phenomenon of active demethylation by enzymes such as the Ten eleven translocation (TET) oxidases – TET1, TET2, TET3 and Thymine DNA glycosylase (TDG) has been identified. The discovery that TET family enzymes can oxidize 5-methylcytosine has greatly advanced our understanding of DNA demethylation (Tahiliani, Koh et al. 2009; Ito, D'Alessio et al. 2010; Ito, Shen et al. 2011). TET oxidases act as both writers and erasers of epigenetic marks by catalyzing the conversion of 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), thereby erasing one mark and establishing another. 5-hmC is an important intermediate in the process of demethylation that can either be passively depleted through DNA replication or actively reverted to cytosine through iterative oxidation and thymine DNA glycosylase (TDG)-mediated base excision repair. This discovery has revealed that the processes of DNA methylation and demethylation are dynamic (Kohli and Zhang 2013).
Table 1. Epigenetic enzymes - writers and erasers, adapted and modified from Butler et al. (Butler and Dent 2013)

A table depicting enzymes that establish/erase DNA modifications (writers/erasers) and their classical effect on gene transcription. TBD (to be determined)

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Writer/Eraser</th>
<th>DNA modification</th>
<th>Transcriptional outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNMT1, 2, 3A, 3B</td>
<td>Writer</td>
<td>5mC</td>
<td>Repression</td>
</tr>
<tr>
<td>TET 1, 2, 3</td>
<td>writer/eraser</td>
<td>5hmC, 5caC, 5fC</td>
<td>Activation</td>
</tr>
<tr>
<td>AID</td>
<td>Eraser</td>
<td>5mC</td>
<td>TBD</td>
</tr>
<tr>
<td>TDG</td>
<td>Eraser</td>
<td>5mC</td>
<td>TBD</td>
</tr>
</tbody>
</table>

1.4.1.2 DNA methylation and hematopoiesis

The role of DNMT3A and DNMT1 in hematopoiesis has been investigated using mouse models (Broske, Vockentanz et al. 2009; Trowbridge, Snow et al. 2009; Challen, Sun et al. 2012). Using a tissue specific cre-lox knockout system of Dnmt3a in hematopoietic cells, Challen et al. could demonstrate that Dnmt3a is necessary for differentiation of HSCs, but not for lineage choice (Challen, Sun et al. 2012). In contrast, two independent studies could confirm that Dnmt1 is required for maintenance of HSCs and highly important for lineage specification (Broske, Vockentanz et al. 2009; Trowbridge, Snow et al. 2009). Since conditional knockout of Dnmt1 in hematopoietic cells lead to rapid death of mice due to bone marrow failure (Broske, Vockentanz et al. 2009), an inducible hypomorphic Dnmt1 system was used for repopulation assays. Studies concluded that HSCs with hypomorphic Dnmt1 had reduced self-renewal capacity. Furthermore, myeloerythroid transcription factors were upregulated in Dnmt1 hypomorphic mice, whereas lymphoid regulators were downregulated. These results clearly indicate that Dnmt1, and maintenance of methylation, is important for hematopoietic lineage choice (Broske, Vockentanz et al. 2009; Trowbridge, Snow et al. 2009; Vockentanz, Broske et al. 2010).

The aforementioned evidences give ample proof that global DNA methylation plays a regulatory role in hematopoietic lineage specification. DNA methylation changes are often associated with genes important for hematopoietic control,
such as transcription factors and genes involved in functions of mature hematopoietic cells (Ji, Ehrlich et al. 2010; Bocker, Hellwig et al. 2011; Bock, Beerman et al. 2012). Many studies have shown changes in methylation patterns where lineage specific genes become unmethylated and transcriptionally active, while genes associated with other lineages remain methylated and transcriptionally silent. Methylation changes have also been described in binding sites for hematopoietic transcription factors and putative enhancers (Schmidl, Klug et al. 2009; Hodges, Molaro et al. 2011; Bock, Beerman et al. 2012; Hogart, Lichtenberg et al. 2012; Lee, Xiao et al. 2012).

Experimental evidence seems to support the notion that maintenance methylation is required for lymphoid, but not for myeloid commitment. Several studies have described loss of methylation in myeloid and erythroid cells during differentiation (Ji, Ehrlich et al. 2010; Bocker, Hellwig et al. 2011; Hodges, Molaro et al. 2011; Bock, Beerman et al. 2012; Hogart, Lichtenberg et al. 2012). By contrast, lymphoid cells appear to show a net gain in methylation (Hodges, Molaro et al. 2011).

1.4.1.3 Histone modifications – writers and erasers

The organization of DNA into higher order structures, or nucleosomes, is a central component to epigenetic gene regulation. Each nucleosome, which represents the basic repeating unit of chromatin, consists of 147 bp of DNA wrapped around a core of eight histones including two molecules each of H2A, H2B, H3 and H4, where histone H3 and H4 have long tails protruding from the nucleosome (Luger, Mader et al. 1997). The concept of a ‘histone code’ was proposed following the discovery of specific post-translational covalent modifications (PTMs) of these histone tails by acetylation, methylation, phosphorylation, glycosylation, SUMOylation and ubiquitylation. Such modifications act in a concerted manner to induce structural changes in the chromatin fiber and to regulate the accessibility of transcription factors to gene regulatory sequences, thereby regulating gene expression (Jenuwein and Allis 2001). Enzymes that erase PTMs, in turn, are also vital determinants of gene expression and cell fate.
The following table enlist the various writers of commonly investigated histone methylation modifications, the enzymes that are known to establish them and the effect of the modifications on chromatin structure and in turn gene expression –

Table 2. Histone methylation marks and their association with chromatin state (Khare, Habib et al. 2012)

<table>
<thead>
<tr>
<th>Histone methylation marks</th>
<th>Enzymes</th>
<th>Associated with</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K4me1/2/3</td>
<td>MLL1,2,3,4,5 ; SMYD3; ASH1L; SETD1A; SETD1B</td>
<td>Euchromatin</td>
</tr>
<tr>
<td>H3K79me1/2/3</td>
<td>DOT1L</td>
<td>Euchromatin</td>
</tr>
<tr>
<td>H3K361/2/3</td>
<td>NSD1,2; SETD2, SMYD2</td>
<td>Euchromatin</td>
</tr>
<tr>
<td>H3K27me1/2/3</td>
<td>EZH1,2; NSD3</td>
<td>Heterochromatin</td>
</tr>
<tr>
<td>H3K9me1/2/3</td>
<td>EHMT1,2; PRDM2; SETD1B,2; SUV39H1,2</td>
<td>Heterochromatin</td>
</tr>
<tr>
<td>H3K20me1/2/3</td>
<td>SETD8 ; NSD1,2; SUV20H1,2</td>
<td>Heterochromatin</td>
</tr>
</tbody>
</table>

According to Zhou et al., similar to DNA methylation, histone modifications may serve as ‘dials’ or ‘switches’ for cell type specificity (Zhou, Goren et al. 2011). Some promoters show patterns of overlapping histone modifications such as H3K4me3 (associated with active transcription) and H3K27me3 (associated with repression), commonly referred to as bivalent domains (Bernstein, Mikkelsen et al. 2006). Bivalent promoters are often associated with genes involved in cell fate determination and differentiation, and are believed to keep genes in a state that is transcriptionally silent, but poised for activation.

As shown in table 2, histone modification can affect gene transcription in varying fashion depending on their localization. At promoters, they can contribute to activation, repression or a poised state. At gene bodies, active genes have higher nucleosome occupancy in their exons while repressed genes do not. At distal sites, histone marks correlate with levels of enhancer activity. Importantly, on a global scale, they may confer repression of varying stabilities and be associated with different genomic features. For example, modifications such as H3K9me3, H4K20me3 and H3K27me3 are associated with gene repression and often associated with heterochromatin, while H3K4me3 and H3K79me2 are associated with euchromatin.
1.4.1.4 Histone modifications and Hematopoiesis

Several studies have shown that, like DNA methylation, histone modifications may be involved in regulating self-renewal and differentiation. Bivalent domains appear to participate in regulating genes related to hematopoietic control. In HSCs bivalent domains are present at lineage specifying genes (Adli, Zhu et al. 2010). Many HSC and progenitor-specific genes such as Lef1, Pax5, Ebf1 and Gata6 which are regulators of hematopoietic fate are associated with bivalent domains, containing both the H3K4me3 and H3K27me3 marks on their promoter regions. The presence of bivalent chromatin is associated with an increased likelihood of transcriptional induction. The ‘transcriptional priming’ of these genes eventually sees them lose repressive H3K27me3 marks in early blood cells resulting in their expression and differentiation of the cells (Adli, Zhu et al. 2010; Abraham, Cui et al. 2013).

Recent studies have also indicated that heterochromatin formation through H3K9me2/3 and H4K20me3 modifications could play a role in the epigenetic regulation of gene expression in hematopoiesis. Centromeres in human peripheral blood cells aggregate into distinct "myeloid" and "lymphoid" spatial patterns, suggesting that the three-dimensional organization of centromeric heterochromatin may be determined during hematopoietic differentiation (Alcobia, Quina et al. 2003). Composition of these heterochromatic structures may contribute to the dynamic relocation of genes in different nuclear compartments during cell differentiation, which might have functional implications for cell-stage-specific gene expression (Alcobia, Quina et al. 2003). Interestingly, during normal human hematopoietic differentiation, changes in gene subnuclear location relative to pericentromeric heterochromatin appear to be dictated by whether the gene will be permanently silenced or activated (Guillemin, Maleszewska et al. 2009).

Although there's a huge scope for further studies to better understand the role of histone modifications in hematopoiesis, murine knockout models have been used to understand the functional relevance of enzymes that establish these marks. Mll1, a histone methyltransferase that establishes the H3K4me3 modification was identified as essential for normal HSC self-renewal and critical
for primitive hematopoiesis (Artinger, Mishra et al. 2013). Mice with induced Mll deletion in the bone marrow die after three weeks from bone marrow failure and HSC depletion (Jude, Climer et al. 2007). In contrast, committed progenitors do not require MLL (Jude, Climer et al. 2007).

Similar studies have been performed for Ezh2, the enzymatic part of the polycomb group (PcG) PRC2 complex that establishes H3K27me3 marks which repress gene transcription. Overexpression of Ezh2 increases HSC self-renewal (Kamminga, Bystrykh et al. 2006). In addition, Ezh2 overexpression also leads to abnormal myeloid expansion in mice (Herrera-Merchan, Arranz et al. 2012), while Ezh2 mutation causes deficiencies in early B cell development (Su, Basavaraj et al. 2003).

Lsd1 a H3K4me2 demethylasewhen knocked-down in mice causes expansion of myeloerythroid progenitors by enhancing proliferation, while inhibiting terminal differentiation of granulocytes, megakaryocytes and erythrocytes (Sprussel, Schulte et al. 2012).

1.2 Acute Myeloid Leukemia (AML)

Leukemia is defined as a chronic or acute disease that involves the blood system, characterized by atypical proliferation and lack of differentiation capacity of white blood cells and is classified according to the lineage of the dominating WBC. Leukemia can be generally subdivided into the following four categories based on blood cell type affected and how quickly the disease arises: acute myeloid leukemia (AML), ALL (acute lymphoblastic leukemia), CML (chronic myeloid leukemia) and CLL (chronic lymphoblastic leukemia). In spite of this heterogeneity, leukemia shares characteristics with the majority of known neoplasms i.e. increase in proliferation, block in differentiation, overcoming of cell cycle check points, inhibition of apoptosis and increased telomere maintenance (Weissman 2000; Warner, Wang et al. 2004). These characteristics are acquired over a period of time through a process where accumulated somatic mutations (or in some cases germ line mutations) act as
a pre-leukemic base for additional genetic/epigenetic events which lead to full blown leukemia (Jan, Snyder et al. 2012; Jan and Majeti 2013; Kronke, Bullinger et al. 2013).

1.2.1 Genetics of AML

The study of the molecular pathogenesis of AML through the use of cytogenetic analysis, massive parallel sequencing and next generation sequencing has established that 50% of AML patients harbor various cytogenetic abnormalities, while nearly 50% are cytogenetically normal but harbor various gene mutations. Such variability in genetic and molecular basis of pathology makes AML a very heterogeneous group in terms of clinical course and response to therapy. In the 1970s, a FAB classification was created to divide AML into subtypes, M0 through M7, based on the maturation of the leukemic cells, when observed microscopically. The FAB classification system was used to group AML into subtypes but it did not take into account many of the factors that are known to impact prognosis. The World Health Organization (WHO) published a newer system in 2001 (Harris, Jaffe et al. 1999), which was later updated in 2008 (Tefferi and Vardiman 2008), that includes some of these factors to help better classify cases of AML based on a patient’s prognosis.

In general, both prognosis and treatment choice for AML patients are based on the presence or absence of specific genetic alterations, which determine AML classification in three risk based-categories: favorable, intermediate, and unfavorable.

AML with a favorable prognosis includes patients with inv(16) (that generates the CBFB–MYH11 fusion protein), t(15;17) (that generates the PML–RARA fusion protein), or t(8;21) (that generates the AML1–ETO fusion protein). The 5-year overall survival (OS) rate of patients in this category is 55%.
Table 3. Commonly occurring gene mutations in AML and their biological and clinical features, adapted from Dohner et al. 2008 (Dohner and Dohner 2008)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Biological and Clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NPM1</strong></td>
<td>Encodes a phosphoprotein with pleiotropic functions. Mutations found in 25-35% of all adult AML, 45-64% of CN-AML NPM1 mutations associated with FLT3-ITD (~40%) and FLT3 TKD mutations Associated with better response to induction chemotherapy genotype &quot;mutant NPM1 without FLT3-ITD&quot; is associated with favorable OS</td>
</tr>
<tr>
<td><strong>CEBPA</strong></td>
<td>Encodes a master regulatory transcription factor in hematopoiesis. Mutations found in 10-18% of CN-AML Associated with better OS.</td>
</tr>
<tr>
<td><strong>RUNX1</strong></td>
<td>Encodes a transcription factor involved in normal hematopoietic differentiation. RUNX1 mutations found in about 10% of CN-AML</td>
</tr>
<tr>
<td><strong>FLT3-ITD</strong></td>
<td>Member of the class III receptor tyrosine kinase family, important for HSPC survival ITD FLT3-ITD found in about 20% of all AML, and in 28-34% of CN-AML. Association with inferior outcome</td>
</tr>
<tr>
<td><strong>FLT3-TKD</strong></td>
<td>FLT3 TKD point mutations found in 5-10% of all AML, in 11-14% of CN-AML</td>
</tr>
<tr>
<td><strong>KIT</strong></td>
<td>Member of the class III receptor tyrosine kinase family; Associated with stem cell factor; key role in survival, proliferation and differentiation of HSPCs KIT mutations found in about 30% of CBF AML, and in rare cases of other AML types. Mutations, in particular in exon 17, associated with inferior outcome in many studies</td>
</tr>
<tr>
<td><strong>RAS</strong></td>
<td>Signaling molecule regulating mechanism of proliferation, differentiation and apoptosis. NRAS mutations found in 9-14% of CN-AML</td>
</tr>
<tr>
<td><strong>WT1</strong></td>
<td>Transcription factor implicated in regulation of HSPC apoptosis, proliferation and differentiation Mutations found in about 10% of CN-AML</td>
</tr>
</tbody>
</table>

The unfavorable subgroup includes patients with monosomy 5, monosomy 7, 11q23 (that generates MLL-highly variable breakpoints on the partner fusion protein), or complex cytogenetics, and the 5-year OS rate is reduced to 11%. The intermediate subgroup includes cytogenetically normal (NK) AML patients. Patients belonging to this group have a 5-year OS rate ranging between 24 and 42% (Riva, Luzi et al. 2012). Besides cytogenetic abnormalities, gene mutations also have an enormous clinical significance in AML. Commonly occurring mutation include- mutations in genes affecting cell proliferation (FLT3, KIT, N-
RAS), myeloid transcription (AML1/RUNX1, CEBPA), tumor suppressor genes (P53, WT1) and apoptosis-related genes (NPM1) (see table 3).

Recent advancements in genome sequencing have revealed a new set of mutations such as DNMT3A (Yan, Xu et al. 2011), ASXL1 (Abdel-Wahab, Adli et al. 2012; Abdel-Wahab, Tefferi et al. 2012), EZH2 (Lund, Adams et al. 2014), IDH1 (Figueroa, Abdel-Wahab et al. 2010), IDH2 (Figueroa, Abdel-Wahab et al. 2010), TET2 (Delhommeau, Dupont et al. 2009; Tahiliani, Koh et al. 2009; Abdel-Wahab, Tefferi et al. 2012) that have broadened our understanding of AML as not only a genetic but also an epigenetic disease (Shih, Abdel-Wahab et al. 2012).

Interestingly, some of the abnormalities belonging to this new class seem to be associated with worse patient outcome and more frequently observed in older patients (Ley, Ding et al. 2010; Paschka, Schlenk et al. 2010; Green, Evans et al. 2011; Yan, Xu et al. 2011).

1.2.2 Epigenetics and AML

1.2.2.1 DNA Methylation

Traditionally, a genome-wide loss of methylation with an accompanying hypermethylation of promoter-associated CpG-islands has been considered a hallmark of cancer (Feinberg and Tycko 2004). Recent genome-scale studies suggest that aberrant DNA methylation in cancer occurs at defined genomic locations, termed cancer-specific differentially methylated regions (DMRs) (Hansen, Timp et al. 2011). Several studies have evaluated genome-wide methylation in AML. It was found that a small common set of genes showed consistent aberrant methylation across several hundred cases of AML (Figueroa, Abdel-Wahab et al. 2010; 2013). Increased methylation variation in AML compared with healthy tissues was also described (Schoofs, Rohde et al. 2013). Most interestingly, well-defined genetic aberrations of AML showed highly distinct methylation profiles which suggests that aberrant DNA methylation patterns—rather than evolving homogeneously across cancer types—could in fact reflect underlying pathogenetic mechanisms.
Unsurprisingly, genes associated with DNA methylation are found to be mutated in high number of AML cases. *DNMT3A* mutations rank among the most frequent mutations in AML. They occur in about 20% of AML patients while gain-of-function *IDH1* and *IDH2* mutations occur 7 to 8% respectively. Mutations in *TET2* itself have also been described in AML with a frequency ranging from 7–23% (Patel, Gonen et al. 2012).

### 1.2.2.2 Histone modifications

In leukemia, loss of function mutations in *EZH2* contributes to leukemogenesis while the human *MLL1* gene is found to be translocated in 10 to 15 % of all AML (major translocations being *MLL-AF9* (43%) and *MLL-AF10* (13%)) (Shih, Abdel-Wahab et al. 2012). These translocations act as gain of function mutations that lead to increased expression of target genes such as Hoxa9 through interaction with other epigenetic enzymes such as Dot1l which establish enhanced levels of H3K79me2, an active gene expression mark, at fusion target genes (Krivtsov, Feng et al. 2008; Bernt, Zhu et al. 2011; Daigle, Olhava et al. 2011). Another such example is histone demethylase JARID1A which contains a C-terminal PHD finger that binds its substrate, H3K4me2/3. Translocations occurring in AML create a JARID1A-NUP98 fusion protein that uses this PHD finger to aberrantly bind PcG-silenced Hox genes and maintain them in a transcriptionally active state, resulting in an arrest of hematopoietic differentiation (Wang, Song et al. 2009).

HDACs are notoriously linked with hematologic malignancies through their association with leukemogenic fusion proteins. In addition, they are also known to mediate aberrant transcriptional repression of genes required for hematopoietic differentiation. For instance, overexpression of Hdcac1 was shown to inhibit myeloid differentiation in murine models. This has led to small-molecule HDAC inhibitors being used for leukemia therapy in combination with other cytotoxic agents (Glozak and Seto 2007; Wada, Kikuchi et al. 2009; Quintas-Cardama, Santos et al. 2011; Shih, Abdel-Wahab et al. 2012).
All of this adds to the growing notion that epigenetic modifiers have a vital role in leukemia and further investigation is necessary to understand their role in normal hematopoiesis and leukemogenesis.

**1.3 PIWI proteins**

Recent work has identified the germ cell-associated Piwi (P-element induced wimpy testis) proteins as being crucial players in self-renewal of germ stem cells (GSCs). Furthermore, Piwi proteins, in association with non-coding RNA- which exclusively interact with Piwi proteins-known as Piwi-interacting RNA (piRNA), are critically involved in the DNA methylation and histone modification machinery for repressing transposable elements (Aravin, Hannon et al. 2007; Aravin, Sachidanandam et al. 2007; Brennecke, Aravin et al. 2007; Aravin, Sachidanandam et al. 2008; Brennecke, Malone et al. 2008). Interestingly, *PIWI* genes have been found to be deregulated in many cancers (Lee, Schutte et al. 2006; Liu, Sun et al. 2006; Grochola, Greither et al. 2008; Liu, Shen et al. 2010; Zeng, Qu et al. 2011; Lu, Zhang et al. 2012).

**1.3.1 The Piwi Sub-family**

Piwi genes belong to a highly conserved family of genes encoding Argonaute proteins. They were originally identified in Drosophila as proteins which are essential for GSC self-renewal and maintenance (Cox, Chao et al. 1998; Cox, Chao et al. 2000). Piwi proteins are highly conserved across evolutionary lineages and are present in both plants and animals. Although the majority of eukaryotic organisms possess more than one Piwi-like (Piwil) protein, the functions of individual members of the family seem to be non-redundant. Drosophila possesses three Piwi encoding genes—piwi, aubergine (aub), and argonaute3 (ago3). Mice possess three Piwi encoding genes (Piwil1, Piwil2, and Piwil4), while humans possess four members
Introduction

(PIWIL1–4) (Cox, Chao et al. 1998; Cox, Chao et al. 2000; Deng and Lin 2002; Qiao, Zeeman et al. 2002).

### Table 4. Human PIWI like proteins

Table depicting human PIWIL proteins, their chromosomal location, coding sequence (CDS) and protein size

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alternate name</th>
<th>Organism</th>
<th>Location</th>
<th>Size of mRNA/CDS (kb)</th>
<th>Size of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIWIL1</td>
<td>HWI</td>
<td>Human</td>
<td>chr 12 (q24.33)</td>
<td>3.5; 2.5</td>
<td>852 aa</td>
</tr>
<tr>
<td>PIWIL2</td>
<td>HULI</td>
<td>Human</td>
<td>chr 8 (q21.3)</td>
<td>3.9; 2.9</td>
<td>873 aa</td>
</tr>
<tr>
<td>PIWIL3</td>
<td>HIWI2</td>
<td>Human</td>
<td>chr 22 (q11.23)</td>
<td>3.5; 2.6</td>
<td>882 aa</td>
</tr>
<tr>
<td>PIWIL4</td>
<td>HULI</td>
<td>Human</td>
<td>chr 11 (q21)</td>
<td>3.2; 2.5</td>
<td>847 aa</td>
</tr>
<tr>
<td>Piwil1</td>
<td>Miwi</td>
<td>Murine</td>
<td>chr 5 (72.0 cm)</td>
<td>3.9; 2.5</td>
<td>862 aa</td>
</tr>
<tr>
<td>Piwil2</td>
<td>Miwi2</td>
<td>Murine</td>
<td>chr 14 (D2)</td>
<td>4.9; 2.9</td>
<td>971 aa</td>
</tr>
<tr>
<td>Piwil4</td>
<td>Miwi2</td>
<td>Murine</td>
<td>chr 9 (A2)</td>
<td>3.3; 2.6</td>
<td>878 aa</td>
</tr>
</tbody>
</table>

aa, amino acids; CDS, coding sequence.

1.3.2 PIWI protein structure

The Piwi protein, like their argonaute counterparts, contains three main structural features (figure 2): the PAZ (Piwi Argonaut and Zwille) domain, the MID (middle) domain, and the C-terminal PIWI domain. In addition to Piwi proteins, the PAZ domain is also found in Dicer, which plays a key role in RNA interference mechanisms (Cerutti, Mian et al. 2000; Parker, Roe et al. 2004; Tahbaz, Kolb et al. 2004; Hutvagner and Simard 2008). Crystal structures of PAZ domains from different organisms (Lingel, Simon et al. 2003; Yang, Chen et al. 2010; Schwalbe 2011) have revealed a specific binding pocket that anchors the characteristic two nucleotide 3’ overhang that results from digestion of RNAs by RNases. The PIWI domain is a catalytic domain containing an RNase-H-like fold with a degenerate catalytic center (Asp-Asp-Asp/Glu/His/Lys) and requires the binding of a divalent cation for its activity (Tolia and Joshua-Tor 2007). A third functionally crucial domain lies between the PAZ and PIWI domain—the MID domain, discovered upon structural analysis of the Argonaute protein Af-Piwi in the archaeon *Archaeoglobus fulgidus* (Parker, Roe et al. 2004). This domain binds to the 5’ phosphates of small RNAs and thus anchors small RNAs onto the Piwi protein (Parker, Roe et al. 2005). In addition, Piwi proteins also contain a highly conserved motif in the MID domain that is similar to the 7-methylguanine (m7G) cap-binding motif of eukaryotic translational initiation factor 4E (eIF4E). Some Argonautes are known to inhibit translation (Parker, Roe et al. 2005; Kiriakidou, Tan et al. 2007; Parker, Parizotto et al.
2009) probably through their m7G-cap binding capability which may prevent eIF4E binding and therefore represses translation. Interestingly, in murine germ cells Piwil1 has been shown to associate with the translational machinery (Grivna, Pyhtila et al. 2006). Another interesting aspect of the MID domain is its possible role in protein–protein interactions. A recent study demonstrated that Piwi proteins in *Drosophila*, mice, and *Xenopus* contain symmetric dimethylarginines (sDMA), a post-translational modification established by the arginine methyltransferase PRMT5 (Kirino, Kim et al. 2009). These methylated arginine residues on the MID domain seem to act as a molecular scaffold for binding of other proteins such as the Tudor-domain containing proteins (Siomi, Mannen et al. 2010; Huang, Houwing et al. 2011).

![Figure 2. Structure of Piwi proteins.](image)

**Figure 2. Structure of Piwi proteins.** (A): General structure of Argonaute proteins depicting the PAZ, MID, and PIWI domains (connected through linker domains L1 and L2) and their known functions. (B): Protein alignment of individual human PIWI proteins indicating domains and size of proteins. Information for protein domain sizes was obtained from Nextprot (www.nextprot.org) and NCBI (http://www.ncbi.nlm.nih.gov). As the size of the MID domain is not clearly defined for all human Piwi proteins, it is only shown schematically in the figure. Abbreviations: MID, middle; PAZ, Piwi Argonaut and Zwille; PIWIL, Piwi-like.

### 1.3.3 PIWI and PIWI interacting RNA

An important clue to the molecular function of Piwi proteins lies in their small RNA partners. The recently discovered piRNAs (Piwi-interacting RNA) are
noncoding RNAs that associate with Piwi proteins. First identified in germ cells (Aravin, Gaidatzis et al. 2006; Girard, Sachidanandam et al. 2006; Kim 2006) and more recently in somatic cells (Cichocki, Lenvik et al. 2010; Zeng, Qu et al. 2011), piRNAs differ greatly from micro RNA (miRNA) and small-interfering RNA in their structure and origin (figure 3A). piRNAs are considerably longer than mature miRNA and are derived from single stranded precursors that originate largely from repetitive sequences in the genome. piRNAs are believed to act as sequence-specific guides for Piwi protein complexes to mediate what seems to be their primary function: transcriptional gene silencing of retrotransposons and other genetic elements in GSC. Furthermore, Piwi proteins seem to be essential for the biogenesis and/or the stability of piRNA (in Drosophila and mice) (Aravin, Gaidatzis et al. 2006; Aravin, Hannon et al. 2007; Aravin, Sachidanandam et al. 2007; Aravin, Sachidanandam et al. 2008). Extensive analyses of piRNAs associated with Piwi proteins in flies and mice have identified the genomic origins of piRNAs and led to the proposal of two biogenesis pathways: the primary processing pathway and the ping-pong amplification loop (figure 3B). Both mechanisms are important for mounting an effective defense against TEs. First, the primary piRNA biogenesis pathway results in production of primary piRNA. Next, the ping-pong cycle further shapes the piRNA population by using the initial pool of piRNAs, which associate with specific Piwi proteins, to target and cleave multiple TE transcripts thereby further amplifying production of piRNAs.
Figure 3. The PIWI-piRNA axis (A): Basic structure of a piRNA— with a 5’ monophosphate, 5’ uridine residue, and a 3’ methyl modification. (B): Schematic presentation of the primary biogenesis and the “Ping Pong cycle” (secondary biogenesis) in mouse germ
cells. **(C)**: Distribution of sequences to which known piRNA map in postnatal day 10 mice (figure adapted and modified from Aravin et al. (Aravin, Hannon et al. 2007)). Abbreviations: ncRNA, non-coding RNA; piRNA, piwi-interacting RNA; PIWIL, Piwi-like; rRNA, ribosomal RNA; TE, transposable element.

### 1.3.4 PIWI-piRNA axis: TE silencing and beyond

Although Piwi proteins have more than one essential function, it is likely the primary function is the maintenance of the self-renewing property of GSCs. The specific germline function of the Piwi-piRNA pathway is most likely the repression of TEs (Aravin, Hannon et al. 2007; Brennecke, Malone et al. 2008), a process that is thought to involve heterochromatin formation and transcriptional and post-transcriptional silencing. This premise is bolstered by the fact that majority of piRNAs are antisense to TE sequences, suggesting that TEs are the primary target. In mammals, it appears that the activity of piRNAs in TE silencing is most important during embryo development (Aravin, Sachidanandam et al. 2008). Piwi-associated TE silencing depends on several other proteins. Importantly, post-translational modification of Piwi proteins renders them capable of binding to other proteins, as they contain conserved sDMA that provide a binding platform. Tudor-domain containing proteins were shown to form complexes with Piwi proteins and have been identified as necessary players in TE silencing in the mouse male germline. In mice, Tdrd9 (Tudor-domain containing 9) and Tdrd1 form a complex with Piwil4 and Piwil2, respectively, and cooperate non-redundantly in the Piwi-piRNA pathway. Also, the Tdrd9-Piwil4 complex has been shown to be essential for retrotransposon silencing (Shoji, Tanaka et al. 2009).

Although limiting TE activity, which imposes a constant threat to the genomic integrity of GSCs, is a major task of the Piwi-piRNA pathway, there are hints that this complex functions beyond TE silencing and might affect protein coding regions in GSCs. There is a class of piRNA, namely the pachytene piRNA (piRNA expressed during prophase of meiosis), which are associated with non-repetitive and possibly protein coding elements, indicating that they could well be targeting protein coding regions (figure 3C). Moreover, protein-coding genes that are embedded in heterochromatin regions of the genome often contain TE
remnants in their introns, which probably enhance their chances of being targeted by piRNAs.

In fact, one of the first piRNAs described in *D. melanogaster* was found to target repetitive but protein-coding *Stellate* genes, which encode a functional homolog of the β-subunit of protein kinase CK2 (Aravin, van der Heijden et al. 2009; Ishizu, Nagao et al. 2011). A plethora of studies now indicate that a specific population of *D. melanogaster* piRNAs might also be involved in the silencing of nonrepetitive, protein-coding genes (Li, Alls et al. 2003; Rouget, Papin et al. 2010; Qi, Watanabe et al. 2011). An example of this is the *FasIII* gene of *Drosophila*. *fasIII* encodes an immunoglobulin-like cell adhesion molecule that could be a potential target of the Piwi-piRNA complex in follicle cells of fly ovaries as some *Drosophila* piRNAs show strong complementarity to the *fasIII* primary transcript (Saito, Inagaki et al. 2009). Another example is *nanos*, a developmentally crucial protein required for posterior patterning of the *Drosophila* embryo. piRNAs induce degradation of maternally deposited mRNAs of *nanos* during the maternal–zygotic transition in the *D. melanogaster* embryo (Rouget, Papin et al. 2010).

### 1.3.5 Piwi proteins are crucial for self renewal

The first study to establish the role of Piwi proteins in GSC self-renewal was conducted by Cox et al. in *Drosophila* (Cox, Chao et al. 1998) who showed that Piwi proteins are essential for asymmetric division of germ stem cells (GSCs) but not for their differentiation. Piwi was shown to provide an essential maternal contribution for embryogenesis as eggs produced from homozygous *piwi* mutant GSCs failed to undergo ordered embryonic development and could not be rescued by the paternal *piwi*+ gene. The group could also show that the somatic expression of Piwi modulated the number of GSCs and the rate of their division, while its germline expression contributed to promoting stem cell division in a cell-autonomous manner (Cox, Chao et al. 2000). To date, Piwi proteins have been shown to function in GSCs in mice, *Drosophila*, *Caenorhabditis elegans*, and certain plant species (Cox, Chao...
et al. 1998; Cox, Chao et al. 2000) and are also found to be essential for stem cell function during regeneration in planaria (Reddien, Oviedo et al. 2005).

Various studies in mice have shown that the three Piwi genes, *Piwil1*, *Piwil2*, and *Piwil4*, are expressed in different cellular compartments (*Piwil1* and *Piwil2*—cytoplasmic; *Piwil4*—nuclear) and different time windows (*Piwil1*—14 days postpartum (dpp) to 27 dpp; *Piwil2*—12.5 days postcoitum (dpc) to 20 dpp; *Piwil4*—15.5 dpc to 3 dpp (Thomson and Lin 2009)) during GSC differentiation, indicating important nonredundant functions for each protein. In mammals, *Piwil1*, *Piwil2*, and *Piwil4* are essential for spermatogenesis. *Piwil1* mutant/knockout mice are sterile due to spermatogenic arrest at the beginning of the round spermatid stage (Deng and Lin 2002). *Piwil2* and *Piwil4* male knockout mice are sterile because of impaired spermatogenesis caused by meiotic arrest at the pachytene stage and early prophase of meiosis-I, respectively (Kuramochi-Miyagawa, Kimura et al. 2004; Carmell, Girard et al. 2007; Unhavaithaya, Hao et al. 2009). Of note, the deficiency of *Piwil1*, *Piwil2*, or *Piwil4* in the male germline leads to the activation of TEs such as long interspersed nuclear element and long terminal repeat retrotransposons, caused by a decrease in the methylation patterns in the promoter regions of these TEs (Kuramochi-Miyagawa, Kimura et al. 2004; Carmell, Girard et al. 2007; Aravin, Sachidanandam et al. 2008; Reuter, Berninger et al. 2011).

Interestingly, knockout *Piwil2* or *Piwil4* mice display a phenotype remarkably similar to that of *Dnmt3L*-deficient mice that also display meiotic and epigenetic defects in spermatogenesis along with an increase in transposition of certain retrotransposons (Webster, O'Bryan et al. 2005; Aravin, Hannon et al. 2007; Aravin, Sachidanandam et al. 2008; Aravin, van der Heijden et al. 2009). In addition, deficiency of *Piwil2* or *Piwil4* causes the profile of prepachytene repeat-associated piRNA (a population of piRNA expressed before entry into meiosis) to be markedly reduced, while deficiency of *Piwil1* affects the profile of nonrepeat-associated pachytene piRNA (Grivna, Pyhtila et al. 2006; Aravin, Sachidanandam et al. 2008). These data led to the hypothesis that in GSCs, the Piwi-piRNA complexes might serve as sequence-specific guides that direct the de novo DNA methylation machinery to TEs.
Expression of a Piwi homolog, namely PIWIL1 (or HIWI) has been described in human bone marrow CD34+ hematopoietic stem cells (HSCs) and progenitors. Interestingly, expression of PIWIL1 becomes silenced upon differentiation of bone marrow CD34+ cells, indicating a stem cell- and progenitor-associated expression pattern in the human hematopoietic compartment. However, in this report it was not tested whether PIWIL1 expression is important for normal HSC function (Sharma, Nelson et al. 2001).

1.3.6 PIWI proteins as epigenetic modifiers

In *Drosophila*, multiple lines of evidence point toward a role of Piwi in epigenetic regulation in somatic cells. First, piwi is a typical suppressor of position effect variegation in somatic cells—a common feature it shares with other key epigenetic factors such as heterochromatin protein 1 (HP1) and HP2 (Pal-Bhadra, Leibovitch et al. 2004). Second, piwi deficiency results in loss of methylation of histone 3 at lysine 9 (H3K9me) and the delocalization of HP1 and HP2 from polytene chromosomes in the soma of piwi and aubergine mutants (Pal-Bhadra, Leibovitch et al. 2004). In addition, piwi directly interacts with Rhino, the homolog of HP1 in *Drosophila* (Brower-Toland, Findley et al. 2007; Klattenhoff, Xi et al. 2009). piwi has been generally implicated in heterochromatin formation and transcriptional silencing in *Drosophila* (Pal-Bhadra, Leibovitch et al. 2004; Brower-Toland, Findley et al. 2007) but is also known to promote euchromatic histone modifications and piRNA transcription in subtelomeric heterochromatin or telomere-associated sequences (TAS) (Yin and Lin 2007). piwi binds with TAS on chromosome 3 (3R-TAS) and to piRNA uniquely mapped to 3R-TAS (3R-TAS1 piRNA). In *piwi* mutants, euchromatic histone modifications at 3R-TAS are lost while heterochromatic histone modifications accumulate. Furthermore, expression of both the 3R-TAS1 piRNA and 3R-TAS is suppressed in *piwi* mutants. This indicates that piwi maintains the euchromatic character of a gene through euchromatic histone modifications contrary to its known role of epigenetic silencing.

The most substantial evidences that implicate Piwi-piRNA complexes in epigenetic change, namely, H3K9me2/3 are studies conducted by Huang et al.
in 2013 and Pezic et al in 2014. The study conducted by Huang et al. confirmed that in Drosophila germ cells, Piwi-piRNA complex bind to numerous piRNA-complementary sequences throughout the genome, implicating piRNAs as a major mechanism that guides Piwi and Piwi-associated epigenetic factors to program the genome. They could demonstrate that inserting piRNA-complementary sequences to an ectopic site leads to Piwi, HP1α, and Su(var)3-9 recruitment to the site as well as H3K9me2/3 enrichment and reduced RNA polymerase II association, indicating that piRNA is both necessary and sufficient to recruit Piwi and epigenetic factors to specific genomic sites (figure 4).

Figure 4. PIWI-piRNA complex directs epigenetic machinery to loci in Drosophila, figure adapted and modified from Huang et al. (Huang, Yin et al. 2013) pg. 513

According to Huang et al. piRNA are transcribed from a single piRNA gene but can have complete sequence complementarity to several targets. Once bound to these targets through the help of piRNA, HP-1α and HMTs (histone methyltransferase) interact with Piwi directly or through its associated proteins leading to an epigenetic mark being established at the loci. Adapted from A Major Epigenetic Programming Mechanism Guided by piRNA, Vol.24 Issue.5, Xiao A Huang, Hang Yi, Sarah Sweeney, Debashish Raha, Michael Snyder and Haifan Lin. Figure 8, Copyright (2012), with permission from Elsevier.

The authors could also show that Piwi deficiency drastically changed the epigenetic landscape and polymerase II profile throughout the genome, revealing the Piwi-piRNA mechanism as a major epigenetic programming mechanism in Drosophila germ cells. (Huang, Yin et al. 2013) Similarly, Pezic et al. could show in germ cells, and for the first time in somatic cells of knockout mice that Piwi4/Miwi2 was required to maintain a high level of the repressive
H3K9me3 histone modification almost exclusively on long interspersed nuclear elements (LINEs) in a piRNA dependent mechanism (Pezic, Manakov et al. 2014).

1.3.7 PIWIL4 and normal hematopoiesis

To date, the role of PIWIL genes in human hematopoiesis has not been understood. Recent work has shown that Piwil genes are expressed in hematopoietic cells while Piwi4 or Miwi2 is the only Piwil gene expressed in primitive hematopoietic cell types within the bone marrow. Experiments have shown that Piwil4 knockout mice are able to maintain long-term hematopoiesis with no observable effect on the homeostatic HSC compartment in adult mice. Furthermore, Piwil4-deficient hematopoietic cells are capable of normal lineage reconstitution after competitive transplantation giving a strong indication that Piwil4 could be dispensable for normal adult hematopoiesis (Nolde, Cheng et al. 2013).

In a recent study performed by Jacobs et al. a knockout model was used in which a targeting vector was inserted between exons 1 and 3 of the murine Piwil4 gene, causing the deletion of exon 2. As subtle defects in hematopoiesis might easily go unnoticed, particularly if it did not cause overt anemia or immunodeficiency, these studies were designed to detect slight defects in hematopoietic lineages. Despite this, no significant effects on hemoglobin levels, WBC count, WBC count during ageing or difference in hematopoietic recovery following sub-lethal irradiation were observed (Jacobs, Wagner et al. 2013).

Although such studies have not yet been extended to human hematopoietic cells, the aforementioned studies do provide a strong hint that despite its high expression, Piwil4 does not have an obvious impact on self-renewal of HSCs.

1.3.8 PIWIL4 and cancer
To date, a select few studies have attributed a proto-oncogenic role to PIWIL4 in soft tissue sarcomas (Greither, Koser et al. 2012), cervical (Su, Ren et al. 2012) and epithelial ovarian cancers (Chen, Liu et al. 2013). In other cases, such as breast and colon cancer (Li, Yu et al. 2010), PIWIL4 has been suspected to play a role in tumorigenesis and exhibits a ubiquitous expression pattern.

In cervical carcinoma tissues, PIWIL4 expression is found to be deregulated in comparison to adjacent normal cervical tissues. Furthermore, PIWIL4 seems to promote cell growth, cell invasion and inhibit cell apoptosis through the p14ARF/p53 pathway (Su, Ren et al. 2012). In soft tissue sarcomas, PIWIL4 expression was significantly associated with a worse prognosis while low expression was associated with a 2.58-fold increased risk of tumor-related death (Greither, Koser et al. 2012).

A recent study conducted by Keam et al. (Keam, Young et al. 2014) in breast cancer cells describes the high expression of PIWIL4 in breast cancer while also showing for the first time the presence of PIWIL4 bound piRNA in human somatic cells. Furthermore, the study could also show that PIWIL4 had similar interacting partners in breast cancer cells as in germ cells of lower organisms, such as heat shock proteins and RNA helicases. Amongst others, structural proteins such as myosin and tubulin were also shown to be interacting partners of PIWIL4 (Keam, Young et al. 2014).

Although these studies indicate that PIWIL4 could be a proto-oncogene, no line of evidence has shown its relation to an epigenetic mechanism in cancer. Moreover, studies to understand the role of PIWIL4 in malignant hematopoiesis are lacking.
1.4 Aim of the study

Acute myeloid leukemia (AML) is a stem cell disease characterized by mutations and deregulated expression of epigenetic factors. The understanding of the mechanism behind the influence of epigenetic aberrations on AML pathogenesis is limited. PIWIL proteins are stem cell self-renewal associated proteins known to mediate epigenetic silencing in lower organisms. The deregulation of one such PIWI protein family member, *PIWIL4* in different cancers makes it an interesting candidate for an experimental study directed at improving the understanding of epigenetic mechanisms in malignant hematopoiesis, particularly AML.

The objective of this study was to first, profile normal and malignant human hematopoietic tissues for the expression of *PIWIL4* and subsequently understand its functional role in normal hematopoietic cells and in AML by means of depletion and over-expression studies. A lentiviral transduction model provided an excellent tool for this purpose. We sought to understand the effect of deregulated *PIWIL4* expression in AML cells by assaying its effect on phenotypic indices such as growth – *in vitro* and *in vivo* by way of liquid culture and colony forming assays (*in vitro*) and immune-compromised mouse models (*in vivo*). To understand the molecular mechanisms behind any phenotypic change in AML cells we assayed gene expression and epigenetic changes. We focused primarily on epigenetic changes in the form of histone modifications that are associated with PIWI proteins in lower organism. In a bid to better understand the epigenetic mechanism behind PIWIL4 function in malignant cells we generated a *PIWIL4* mutant and assayed its epigenetic function versus the wild type protein. We further gained mechanistic insights by performing a search for PIWI interacting RNA (piRNA) through small RNA deep sequencing and piRNA specific microarray, and used mass spectrometry to analyze the protein binding partners of PIWIL4 in AML cells. With the aforementioned assays we intend to understand the novel epigenetic role of *PIWIL4* in human AML cells.
2. Materials and Methods

2.1 Eukaryotic cells

2.1.1 Primary cells

<table>
<thead>
<tr>
<th>Bone Marrow Mono Cells, cryoamp</th>
<th>Lonza Cologne GmbH, Cologne, Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone marrow CD34+ cells</td>
<td>Lonza Cologne GmbH, Cologne, Germany</td>
</tr>
<tr>
<td>Cord blood cells</td>
<td>University Hospital, Ulm</td>
</tr>
<tr>
<td>MLL-AF9 patient cells</td>
<td>University Hospital, Ulm</td>
</tr>
</tbody>
</table>

The following leukemic cell lines and cell line information were obtained from DSMZ, Braunschweig, Germany

2.1.2 Human AML cell lines

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Mutations</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>EOL-1</td>
<td>MLL partial tandem duplication described to carry the fusion FIP1L1-PDGFRA</td>
<td>acute myeloid (eosinophilic) leukemia established in 1984 at diagnosis from the peripheral blood of a 33-year-old man</td>
</tr>
<tr>
<td>HL-60</td>
<td>AML FAB M2</td>
<td>established from the peripheral blood of a 35-year-old woman with acute myeloid leukemia (AML FAB M2) in 1976</td>
</tr>
<tr>
<td>MONO-MAC-6</td>
<td>AML FAB M5 t(9;11)(p22;q23) leading to MLL-AF9 fusion gene</td>
<td>established from the peripheral blood of a 64-year-old man with relapsed acute monocytic leukemia (AML FAB M5)</td>
</tr>
<tr>
<td>Cell Line</td>
<td>Cell Type</td>
<td>Origin</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>MV4-11</td>
<td>AML Fab M5 FLT3 internal tandem duplication</td>
<td>established from a 10-year-old boy with acute monocytic leukemia (AML FAB M5) at diagnosis</td>
</tr>
<tr>
<td>NB-4</td>
<td>cells carry the t(15;17) PML-RARA fusion gene</td>
<td>established from the bone marrow of a 23-year-old woman with acute promyelocytic leukemia (APL = AML FAB M3) in second relapse in 1989</td>
</tr>
<tr>
<td>OCI-AML3</td>
<td>AML FAB M4 cells carry an NPM gene mutation</td>
<td>established from the peripheral blood of a 57-year-old man with acute myeloid leukemia (AML FAB M4) at diagnosis in 1987</td>
</tr>
<tr>
<td>OCI-AML5</td>
<td>AML FAB M4</td>
<td>established from the peripheral blood of a 77-year-old man with acute myeloid leukemia (AML M4) in relapse in 1990</td>
</tr>
<tr>
<td>THP-1</td>
<td>Acute Monocytic Leukemia carries t(9;11)(p21;q23) leading to MLL-AF9 fusion gene</td>
<td>established from the peripheral blood of a 1-year-old boy with acute monocytic leukemia (AML) at relapse in 1978</td>
</tr>
</tbody>
</table>

### 2.1.3 Human ALL cell lines

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Cell Type</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAJI</td>
<td>B-cell ALL (human Burkitt lymphoma) cells are described to carry t(8;14) leading to MYC-IGH@ (MYC-IGH) fusion gene</td>
<td>established from the left maxilla of a 12-year-old African boy with Burkitt lymphoma in 1963</td>
</tr>
</tbody>
</table>

### 2.1.4 Human packaging cell line
2.2 Prokaryotic cells

Max Efficiency DH5 alpha Competent Cells

F- φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mk+) phoA supE44 λ- thi-1 gyrA96 relA1 Invitrogen, Carlsbad, CA, USA

2.3 Antibodies

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Company</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIWIL4</td>
<td>Promab Biotech</td>
<td>Western, Intracellular staining, Immunoprecipitation</td>
</tr>
<tr>
<td>CD34</td>
<td>BD biosciences</td>
<td>FACS</td>
</tr>
<tr>
<td>CD45</td>
<td>BD biosciences</td>
<td>FACS</td>
</tr>
<tr>
<td>CD38</td>
<td>BD biosciences</td>
<td>FACS</td>
</tr>
<tr>
<td>CD19</td>
<td>BD biosciences</td>
<td>FACS</td>
</tr>
<tr>
<td>CD33</td>
<td>BD biosciences</td>
<td>FACS</td>
</tr>
<tr>
<td>CD3</td>
<td>BD biosciences</td>
<td>FACS</td>
</tr>
<tr>
<td>CD15</td>
<td>BD biosciences</td>
<td>FACS</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>Abcam</td>
<td>Western, ChIP-seq</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Abcam</td>
<td>Western, ChIP-seq</td>
</tr>
<tr>
<td>Histone H3</td>
<td>Abcam</td>
<td>Western</td>
</tr>
</tbody>
</table>

2.4 Taq Man Real time assays

TaqMan® Real Time primers were acquired from Applied Biosystems, Foster City, CA, USA.

<table>
<thead>
<tr>
<th>Name</th>
<th>ID</th>
<th>Exon boundary</th>
</tr>
</thead>
</table>
**Materials and Methods**

### 2.5 Solutions and Media

<table>
<thead>
<tr>
<th>Material</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMEM High Glucose (4.5 g/l)</td>
<td>PAA Laboratories GmbH, Pasching, Austria</td>
</tr>
<tr>
<td>Dulbecco’s PBS</td>
<td>PAA Laboratories GmbH, Pasching, Austria</td>
</tr>
<tr>
<td>RPMI 1640</td>
<td>PAA Laboratories GmbH, Pasching, Austria</td>
</tr>
<tr>
<td>IMDM</td>
<td>PAA Laboratories GmbH, Pasching, Austria</td>
</tr>
<tr>
<td>BIT 9500</td>
<td>Stem cell technologies</td>
</tr>
<tr>
<td>Penicillin/Streptomycin</td>
<td>PAA Laboratories GmbH, Pasching, Austria</td>
</tr>
<tr>
<td>Foetal bovine serum (FBS)</td>
<td>PAN Biotech, Aidenbach, Germany</td>
</tr>
<tr>
<td>Human Serum AB</td>
<td>Lonza Cologne GmbH, Cologne, Germany</td>
</tr>
<tr>
<td>FACS Buffer</td>
<td>PBS containing 1 μM SYTOX® Blue dead cell stain and 3% FBS, stored in dark at 4°C</td>
</tr>
<tr>
<td>MACS buffer</td>
<td>PBS containing 3% FBS, 2mM EDTA and and 0.08% Ciproflxacin; sterile filtered</td>
</tr>
</tbody>
</table>

### 2.6 Reagents and Kits

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloroform</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>Ulm University, Ulm, Germany</td>
</tr>
<tr>
<td>DEPC-treated Water</td>
<td>Applied Biosystems/Ambion, Austin, TX, USA</td>
</tr>
<tr>
<td>DMSO</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>Material/Kit Name</td>
<td>Manufacturer/Location</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Isopropanol (99.5% purity)</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>Ethanol (99.5% purity)</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>Formaldehyde (37.5%)</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>Glacial Acetic acid</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>Coomassie blue</td>
<td>Applichem, Germany</td>
</tr>
<tr>
<td>TransIT Transfection Reagent</td>
<td>Mirus Bio LLC, USA</td>
</tr>
<tr>
<td>Polybrene</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>Puromycin dihydrochloride</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>SYTOX® Blue Dead Cell Stain</td>
<td>Invitrogen, Carlsbad, CA, USA</td>
</tr>
<tr>
<td>TaqMan® Universal PCR Master Mix, No AmpErase®</td>
<td>Applied Biosystems, Foster City, CA, USA</td>
</tr>
<tr>
<td>TRizol® Reagent</td>
<td>Invitrogen, Carlsbad, CA, USA</td>
</tr>
<tr>
<td>Trypan blue</td>
<td>Invitrogen, Carlsbad, CA, USA</td>
</tr>
<tr>
<td>Platinum® Taq DNA Polymerase Kit</td>
<td>Invitrogen, Carlsbad, CA, USA</td>
</tr>
<tr>
<td>QIAGEN Plasmid Maxi Kit (25)</td>
<td>Qiagen, Hilden, Germany</td>
</tr>
<tr>
<td>QIAprep Spin Miniprep Kit (250)</td>
<td>Qiagen, Hilden, Germany</td>
</tr>
<tr>
<td>Thermoscript RT-PCR System</td>
<td>Invitrogen, Carlsbad, CA, USA</td>
</tr>
<tr>
<td>FailSafe™ PCR System with PreMix Choice</td>
<td>Epicentre Biotechnologies, Madison, USA</td>
</tr>
<tr>
<td>illustra GFX PCR DNA and Gel Band Purification Kit</td>
<td>GE Healthcare, Little Chalfont, UK</td>
</tr>
<tr>
<td>Agarose</td>
<td>Sigma-Aldrich, Taufkirchen, Germany</td>
</tr>
<tr>
<td>Gel Loading Dye, Blue (6x)</td>
<td>New England BioLabs, Ipswich, MA, USA</td>
</tr>
<tr>
<td>GelRed™ 10000x</td>
<td>VWR, Darmstadt, Germany</td>
</tr>
<tr>
<td>EcoRI, XbaI, BamH1 restriction enzymes</td>
<td>New England BioLabs, Ipswich, MA, USA</td>
</tr>
</tbody>
</table>
### Materials and Methods

<table>
<thead>
<tr>
<th>Cytokines – SCF, FLT-3, IL-6, IL-3, TPO</th>
<th>Miltenyi biotec GmbH, Bergisch Gladbach, Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACS CD34 microbead kit</td>
<td>Miltenyi biotec GmbH, Bergisch Gladbach, Germany</td>
</tr>
<tr>
<td>RetroNectin</td>
<td>Takara/Clontech Inc., Saint Germain en Laye, France</td>
</tr>
<tr>
<td>BD Pharmingen™ - APC BrdU Flow Kit</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>BD Pharmingen™ - Annexin V</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>Classic Magnetic IP kit</td>
<td>Pierce Thermo Fischer scientific, IL, USA</td>
</tr>
</tbody>
</table>

### 2.7 Plasmids

<table>
<thead>
<tr>
<th>pGreen puro</th>
<th>Lentiviral shRNA vector, Systen biosciences, Mountain View, CA, USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCDH-MSCV-MCS-EF1-GFP-T2A-Puro (pCDH)</td>
<td>Lentiviral expression vector Systen biosciences, Mountain View, CA, USA</td>
</tr>
<tr>
<td>pMD2.G</td>
<td>Contains VSV-G envelope protein, Addgene</td>
</tr>
<tr>
<td>psPAX2</td>
<td>Contains Gag, cPPT, and HIV_Rev_NES, Addgene</td>
</tr>
</tbody>
</table>

### 2.8 Markers

<table>
<thead>
<tr>
<th>Quick-Load® 1 kb DNA Ladder</th>
<th>New England BioLabs, Ipswich, MA, USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR Protein Ladder</td>
<td>New England BioLabs, Ipswich, MA, USA</td>
</tr>
</tbody>
</table>

### 2.9 Technical Equipment

<table>
<thead>
<tr>
<th>7900HT Fast real-time PCR System</th>
<th>Applied Biosystems, Foster City, CA, USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD FACS Aria™ III</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
</tbody>
</table>

34
## Materials and Methods

### Consumable Supplies

<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD LSRFortessa™</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>Cytospin™ Cytocentrifuge</td>
<td>Thermo Fisher Scientific, Waltham, WA, USA</td>
</tr>
<tr>
<td>Eppendorff 5415R</td>
<td>Eppendorff, Hamburg, Germany</td>
</tr>
<tr>
<td>Eppendorff 5810R</td>
<td>Eppendorff, Hamburg, Germany</td>
</tr>
<tr>
<td>Galaxy 170S Incubator</td>
<td>New Brunswick Scientific, Edison, NJ, USA</td>
</tr>
<tr>
<td>Innova 44 Incubator Shaker</td>
<td>New Brunswick Scientific, Edison, NJ, USA</td>
</tr>
<tr>
<td>NanoDrop® ND-1000 Spectrophotometer</td>
<td>Thermo Fisher Scientific, Waltham, WA, USA</td>
</tr>
<tr>
<td>peqSTAR 96 Universal Gradient Thermocycler</td>
<td>PEQLAB Biotechnology GmbH, Erlangen, Germany</td>
</tr>
<tr>
<td>Vortex-Genie 2</td>
<td>Scientific Industries, Bohemia, NY, USA</td>
</tr>
<tr>
<td>Zeiss Axiovert 40 C</td>
<td>Carl Zeiss Microl Imaging GmbH, Göttingen, Germany</td>
</tr>
</tbody>
</table>

### 2.10 Consumable Supplies

<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.45 µm filters for 50 mL Falcon tubes</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>1.5 &amp; 2 mL Eppendorff Tubes</td>
<td>Eppendorff, Hamburg, Germany</td>
</tr>
<tr>
<td>10 cm Cell + tissue culture dish</td>
<td>Sarstedt, Nümbrecht, Germany</td>
</tr>
<tr>
<td>5 mL Polystyrene round-bottom tube with cell-strainer cap (5 mL FACS tube)</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>BD 3 mL syringe</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>BD 50 mL Polypropylene conical tube</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>BD Microlance™ 3</td>
<td>BD Biosciences, Heidelberg, Germany</td>
</tr>
<tr>
<td>Corning Cell culture dish (35mm &amp; 100 mm)</td>
<td>Corning Incorporated, Corning, NY, USA</td>
</tr>
<tr>
<td>Costar® stripette (5 mL, 10 mL, 25 mL)</td>
<td>Corning Incorporated, Corning, NY, USA</td>
</tr>
</tbody>
</table>
2.11 Cell Culture- thawing and culturing

**Cell lines**

Cell lines were thawed at 37 °C in a water bath and washed with PBS to expel remnants of DMSO. Suspension cells (MV4-11, THP-1, NB-4, OCIAML3, OCIAML5 and RAJI) were cultured in RPMI + 10% FBS + 1% Penicillin streptomycin (P/S), while adherent cells (Lenti-X™ 293T) were cultured in Dulbecco’s Modified Eagle medium (DMEM) + 10% FBS + 1% P/S on 10 cm cell culture plates. Culture conditions for all cells were - 37 °C with 5% carbon dioxide (CO₂). Adherent cell lines were split upon reaching 80% confluence.
**Patient samples**

The diagnosis of AML was performed according to the French-American-British criteria (Bennett, Catovsky et al. 1976) and the WHO classification (Harris, Jaffe et al. 1999). Human studies abided by the tenets of the revised World Medical Association Declaration of Helsinki (http://www.wma.net/e/policy/b3.htm).

Patient samples and cord blood were thawed using Lonza’s protocol for ‘thawing Poietics cells’ (http://bio.lonza.com). Cells were maintained in IMDM media containing 10% BIT (Stem cell technologies), $10^{-5}$ M beta mercaptoethanol and 1% P/S, supplemented with cytokines – SCF (50ng/ml), FLT-3, IL-3, IL-6 and TPO (all 10ng/ml) (Bruserud, Gjertsen et al. 2000).

**Human bone marrow cells**

Bone marrow cells were thawed using Lonza’s protocol for ‘thawing Poietics cells’.

**Human bulk cord blood**

Human cord blood was thawed using Lonza’s protocol for ‘thawing Poietics cells’. Cells were suspended in 7 mL of IMDM and layered over an equal amount of Pancoll (density 1,077 g/mL) in a 50mL falcon tube and centrifuged at 630 g for 30 min without acceleration and brake. After centrifugation, the middle fraction containing leukocytes (‘buffy coat’) was transferred to a new tube, mixed with MACS buffer (PBS + 3% FBS + 2mM EDTA and 0.08% Ciprofloxacin; sterile filtered) to 50 mL and centrifuged at 300 x g for 5min at room temperature. The cell pellet was resuspended in 300ul of MACS buffer. MACS CD34 Micro Bead Kit was used to purify CD34$^+$ cells from bulk according to manufacturer’s protocol. Cells were initially maintained in IMDM media containing 10% BIT (Stem cell technologies), $10^{-5}$ M beta mercaptoethanol and 1% P/S, supplemented with cytokines – SCF (100ng/ml), FLT-3 (100ng/ml), IL-3, IL-6 and TPO (all 10ng/ml). After 48hours or for the
purpose of assays, cytokine concentration was reduced to 10ng/ml for all the aforementioned cytokines (Gemelli, Montanari et al. 2006).

### 2.12 Sorting bone marrow subpopulations

Out of the thawed, bulk CD34+ bone marrow (BM) and bone marrow mononuclear cells (BMNC) (purchased from Lonza), 2x10^5 cells were used as single color controls. Bulk and single color control cells, were then pelleted down at 300 x g for 10 minutes at room temperature and the supernatant discarded.

Depending on cell number, a suitable amount of Human Serum type AB (Fc block) was mixed with the pelleted cells and single color controls and blocked for 10 minutes on ice. A sufficient amount of antibody against the specific cell surface marker (CD34, CD38, CD33, CD19, CD3 and CD15) was added to the cell suspension and stained for 20 minutes at 4°C on a rocking station in the dark. Afterwards, the cells were washed with 3 mL PBS and spun down in facs tubes at 300 g for 10 minutes. Single color control cells were then resuspended in 150 µL PBS and the sytox single color control in FACS buffer. Bulk cells were taken up in FACS buffer to a cell density of 1x10^7 cells per mL.

### 2.13 Quantitative Real Time PCR

Using the 7900HT Applied Biosystems Real Time PCR machine, TaqMan probes-set for the gene PIWIL4 was used to detect expression levels, while TATA-box binding protein (TBP) was used as an endogenous control (#4333769F).

### 2.14 Cloning shRNA and cDNA into lentiviral vectors

*Cloning of shRNA against PIWIL4 in pGreenPuro vector*

Two shRNA targeting the sequence – CCCATGCTTGTTAGTCTGTTA- on the 9th exon and CCGACCATATGCAGAGACTTA on the 19th exon of human
PIWIL4 transcript (shPIWIL4) were obtained from Sigma Aldrich (TRCN0000141359 and TRCN0000139278). Since the plasmid was cloned in pLKO.1 lentiviral vector containing only a puromycin resistance marker, the shRNA was cloned into the pGreen Puro lentiviral vector (System Biosciences) which contains both the puromycin resistance marker as well as GFP. Both the top and the bottom complimentary strand of the shRNA were designed with overhangs for the enzymes BamH1 and EcoR1, according to the protocol provided by System Biosciences (Cat no. #SI505A-1) and ordered as oligos. The top and the bottom strands were annealed at 95 degrees for 2 mins with Tris-HCl (pH 8). The annealed strands containing overhangs were used for cloning into pGreen Puro vector. Cloning and confirmation of positive clones was performed according to manufacturer’s protocol. Figure 5 shows the pGreenPuro lentiviral vector, with BamH1, EcoR1 restriction sites and H1 promoter which drives the expression of the cloned shRNA.

**Figure 5.** pGreenPuro vector from system biosciences with a H1 promoter. Figure adopted from the vector manufacturer- System biosciences ([http://www.systembio.com/](http://www.systembio.com/))

**Cloning of wild type PIWIL4 in pCDH vector**

PIWIL4 transcript (NM_152431.2) was amplified using the FailSafe™ PCR System, using primers sequences containing a BamH1 site at its 5’end and a Not1 site on its 3’end. Amplified fragment was run on a gel, excised, gel purified and digested with BamH1 and Not1 and re-purified. The pCDH MSCV EF1 copGFP Puro T2A vector (as shown in Figure 6A) was digested with the
same enzymes, purified and used for cloning using the T4 DNA ligase at 16 °C overnight. Cloning was confirmed with enzyme digestion and sequencing.

**Materials and Methods**

---

**Cloning of mutants in pCDH vector**

Two mutants for PIWIL4 were generated by gene synthesis. Using the enzymes *BamH1* and *Not1*, the mutants were cloned into pCDH vector by Genscript, NJ, USA.

For the first mutant, PIWIL4 coding sequence (CDS) was modified such that the shPIWIL4-A targeting the sequence – `CCCATGCTTGTTAGTCTGTTA` on exon 9 does not bind with it. This was called the shRNA mutant or S-mutant. The mutations were made such that the amino acid sequence remains unchanged, as shown in figure 6B below.

**Figure 6A.** pCDH vector from system biosciences with a MSCV promoter. Figure adopted from the vector manufacturer - System biosciences ([http://www.systembio.com/](http://www.systembio.com/))

**Figure 6B.** S-mutant -Modified CDS on exon 9 of shRNA resistant PIWIL4mutant

Original nucleotide sequence: `CCCATGCTTGTTAGTCTGTTA`

Original aa sequence: `PMLLVSLL`

Modified nucleotide sequence: `CCAATGTTGTTAGTTTGTTA`

(Un)Modified aa sequence: `PMLLVSLL`
Another mutant, known as the piRNA mutant or P-mutant, was generated by making two sets of mutations. Firstly, like the S-mutant, the P-mutant was also mutated at exon 9 to render it resistant to shPIWIL4-A. Next, the PAZ domain of PIWIL4 was removed to inhibit the potential interaction between human PIWIL4 and non-coding RNA. For this, 350 bp stretching from nucleotides 1019 to 1369 was removed from PIWIL4 sequence. As shown in Figure 6C, ORF finder was used (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to ensure that the amino acid sequence of P-mutant was still within frame and would theoretically code for a protein.

![Wild type PIWIL4](image1)

![PAZ-deleted mutant PIWIL4](image2)

Figure 6C. P-mutant - Modified CDS of PIWIL4 lacking the piRNA binding PAZ domain

### 2.15 Lentivirus generation

Up to two days before the planned transfection, 1 x 10^6 Lenti-X™ 293T cells were plated out in a 10 cm dish. The transfection solution TransIT®-LT1 Transfection Reagent (Mirus) was used to transfect helper plasmids – pSPAX2 and pMD2.G along with either scrambled control (PGPSCR), shRNA (PGPshPIWIL4), empty vector (pCDH) or overexpression (pCDH PIWIL4) containing plasmids. The transfection was performed according to the manufacturer's protocol. Media was changed next day to 3ml of DMEM with 30% FBS and 1% P/S. Virus containing media (VCM) was collected in at two time points - after 48 and after 72hrs.
2.16 Transduction of human cells

**Cell lines**

The virus from pGreen Puro and pCDH MSCV lentiviral vectors i.e. Scrambled control (PGPSCR), shRNA (PGPshPIWIL4), empty vector (pCDH) and overexpression (pCDH PIWIL4) were used to transduce AML cell lines. Sorting for GFP+ cells was performed 48hrs after day of infection.

**Patient samples**

5X10⁷ AML cells were thawed (as described in previous sections) and kept in culture overnight. Meanwhile a 6-well suspension dish was coated with retronectin overnight at 4 °C overnight. Coated dish was blocked using 2% BSA and washed with PBS. 500ul of virus was loaded onto the wells and centrifuged at 4 °C for 45 mins at 500 g. Virus media was discarded and 500ul of fresh virus was added followed by cells such that approx. 1 to 1.25X 10⁷ cells were present in each well. Patient cells were kept in a volume of 2ml (including virus medium in IMDM media, complete with all cytokines as mentioned in section 3.1.2) for the next 3 hours before supplementing with additional 3ml of media complete with all cytokines. Cells were harvested after 48hrs using cell disassociation buffer, washed and sorted for GFP positive cells. Assays were set up immediately after sort.

**CD34⁺ cord blood**

Bulk cord blood cells that were thawed and purified for CD34⁺ cells (as described in section 3.1.4) were cycled for 24hrs. 6-well dish loaded with virus was prepared as described above. 2x10⁵ CB cells per well were used for transduction in 2ml volume of media (including virus) supplemented with necessary cytokines (as described in section 3.1.4). 2 ml of additional media was added after 3hrs. Cells were harvested after 48hrs using cell disassociation buffer. Cells were subsequently washed, blocked for 10 mins at
ice using human serum albumin, stained with CD34 (CD34PE – BD biosciences) antibody for 20 mins on ice, washed and sorted for GFP and CD34 (CD34PE – BD biosciences) double positive cells.

2.17 Proliferation and CFC assays

Cell lines
For a Proliferation Assay, 50000 cells were taken from each control and knockdown after sorting of GFP+ cells. The required amount of medium containing the cells was seeded into a 6-well-plate and added up with medium to a total amount of 5 ml. Cells were incubated at 37 °C for 6 days and counted every second day using a cell counter (Beckman Counter ViCell-XR). For CFC assay (colony forming cell assay) 300 GFP+ sorted cells per dish were seeded into two dishes containing methylcellulose (MethoCult® H 4330 medium). The colonies were scored after 14 days.

Patient samples
From each control and experimental arm, 48hrs after sorting of GFP+ cells, 20,000 cells were seeded per CFC dish into two dishes. The colonies were scored after 14 days as blast colony, blast cluster, CFU-GM, CFU-G/CFU-M, BFU-E or CFU-E.

Human CD34+ cord blood
From each control and experimental arm, 48hrs after sorting of GFP+ cells, 300 cells were seeded per CFC dish into two dishes. The colonies were scored after 14 days as CFU-GM, CFU-G/CFU-M, BFU-E or CFU-E.

2.18 In-vivo xenograft study
For this assay, 8-12 week old female NSG mice (NOD.Cg-Prkdc<sup>Scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ) were used. NSG mice, received from Jackson Laboratory, were
bred, maintained and observed 6 times a week for signs of sickness. THP-1 and MV4-11 cell lines were transduced and sorted for GFP positivity 48hrs hence. Mice were subsequently injected intravenously with 1x10^6 MV4-11 and THP-1 cells transduced with either scrambled control or shRNA immediately after sorting. Mice were bled every four weeks to check for engraftment in peripheral blood using flow cytometry.

2.19 Cell cycle and Apoptosis assay

Equal number of THP-1 cells treated with scrambled and shRNA were sorted and starved with 0.1% FBS RPMI for 16hours. Post starvation, cells were maintained in normal growth media (10% FBS) for 48 hours before being harvested for the purpose of cell cycle and apoptosis assay. The Cell Cycle Assay was performed by using BD Pharmingen™ - APC BrdU Flow Kit. According to the Manufacturer’s Manual (BD Pharmingen™ - BrdU Flow Kits Instruction Manual, 2008) Apoptosis Assay was performed by using BD Pharmingen™ - FITC/APC Annexin V Apoptosis Detection Kit I according to the Manufacturer’s Technical Data Sheets (550475 Rev. 8 and 556547 Rev. 5)

2.20 Intracellular staining

THP-1 cells were harvested 96 hours post-transduction. Cell were fixed in 4% PFA for 10 min at room temperature and thereafter treated with 0.1% Triton X-100 for 5 min at room temperature. Blocking was performed in 10% BSA + 5% serum (from the host species of the secondary antibody) for 1 h at room temperature.

Cells were then incubated with primary antibody against PIWIL4 (Promab) at room temperature for 2h. Isotype control sections were incubated with respective IgG primary antibodies. Incubation with the secondary antibodies labeled with Alexa Flouro 488 (invitrogen) was performed for 45 min at room temperature and the Nuclei were stained with DAPI. Tissue sections were then mounted with anti-fade fluorescent or aqueous mounting medium (Invitrogen) and protein signals were visualized under microscope (Carl Zeiss).
2.21 Western Blotting

PIWIL4 knockdown

Lysate for AML cells was prepared using RIPA buffer supplemented with phosphotase and protease inhibitors. Cells were snap frozen in liquid nitrogen and thawed on ice. This process was repeated 3 times in total. The lysate was centrifuged at 1600 \( g \) for 20 mins and supernatant was aspirated and used for western blot. Western blot for AML cells was performed using 90ug of lysate. An 8% SDS PAGE gel was run for 2hrs at 100V followed by a wet transfer to a PVDF membrane at 100V for 1 hr. Anti-PIWIL4 mouse monoclonal antibody from Promab, a secondary mouse IgG HRP antibody followed by an ECL kit from GE was used to detect PIWIL4.

Histone methylation western blot

Lysate from AML cells was prepared in the same fashion as mentioned above. A 15% SDS PAGE gel was used. 50ug of lysate was loaded for each sample and antibodies for H3K9me2, H3K9me3 and H3 (all from Abcam) were used. An anti-rabbit secondary antibody was used for detection.

2.22 Chromatin immunoprecipitation-sequencing

ChIP-seq was performed using the Histone ChIP kit (Diagnode). For performing ChIP, for each antibody, 1x10^6 THP-1 cells transduced with PGPSCR and PGPshPIWIL4 were taken on day 10 after transduction. The antibodies used were as following - H3K9me3 (Upstate Biotechnology) and H3K4me3 (Abcam). Protocol was followed as published in Nature Medicine in 2012 by Schnek et al. (Schenk, Chen et al. 2012).

2.23 RNA deep sequencing

In parallel with ChIP-seq, THP-1 cells were also harvested at day 10 for performing RNA seq. Total RNA was extracted using Trizol, following the
materials and Methods

manufacturer’s protocol. Samples were checked for RNA quality using the Bioanalyzer 2100 from Agilent Technologies. All samples that had a RIN value of more than 9.5 were used for sequencing. Starting with 500 ng total RNA, the library was prepared using the TruSeq RNA Sample Preparation Kit version 2, (#RS-122-2001 and RS-122-2002). Libraries were sequenced at the Genomics Core Facility of the University of Ulm using an Illumina HiSeq 2000 (Illumina, San Diego, CA). The read length of the RNA-Seq was 100 bp.

2.24 Immunoprecipitation and LC/MS

Lysate preparation and immunoprecipitation was performed according to manufacturer’s instructions, as provided with the Pierce Classic Magnetic IP kit. Lysate was prepared using THP-1 cells overexpressing PIWIL4. Cells were transduced with overexpression construct and sorted for GFP 48hrs post-transduction. 10 days post transduction, 10⁷ cells were pelleted. Cells were suspended in Pierce IP lysis buffer along with 10X protease inhibitor cocktail and lysed by snap-freezing in liquid nitrogen and thawing on ice multiple times. Lysate was centrifuged at 13,000g for 10 mins and supernatant was collected and cell debris pellet discarded.

Protein was measured using Bradford assay. 2mg of protein and 5ug of antibody was used for each immunoprecipitation i.e. for IgG control (SantaCruz) and PIWIL4 (Promab). Immunocomplex formation was performed overnight at 4°C with rotation. Lysate was mixed with 50ul of protein AG magnetic beads and rotated at room temperature for 1hr. Beads were washed thrice with IP lysis buffer and once with ultrapure water (Millipore). Protein was eluted using 100ul of low pH buffer at room temperature for 10 mins with rotation. 50ul of precipitated lysate was run on a 10% SDS gel for 2 hours. Gel was removed from cassette and fixed in a Fixing solution (50% methanol and 10% glacial acetic acid) for 1 hour and stained with Coomassie dye (0.1% Coomassie Brilliant Blue R-250, 50% methanol and 10% glacial acetic acid) for 30 mins at room temperature on a rocker. Gel was destained in Destaining solution (5% MeOH, 7.5% HoAC, 87.5% H2O) overnight on rocker at room temperature and subsequently stored in 5% glacial acetic acid.
2.25 Small RNA deep sequencing

30ug of initial RNA from MV4-11 cells transduced with scrambled or shPIWIL4 were used for purifying RNA in the size range of 25-32 nucleotides in length. RNA purification and library preparation were performed according to published protocol (Malone, Brennecke et al. 2012). piRNA were mapped according to size (28-32nt) and according to sequence conservation i.e. containing Uridine at 1st base position or Adenine at 10th base position. Scrambled arm was compared to shRNA arm for difference in percentage of reads of identified individual piRNA like RNA.

2.26 piRNA microarray

piRNA expression levels were measured using a microarray spotted with 23,000 known and putative piRNA sequences using THP-1 cells, untreated or transduced with scrambled or shRNA. Experiment was performed in collaboration with Arraystar, MD, USA. piRNA sequences spotted on array were obtained from web resource - piRNA bank (http://pirnabank.ibab.ac.in/). piRNA that showed more than 2 fold differential regulation with a p-value of less than 0.05 compared to scrambled control were selected for analysis.

2.27 Statistical analysis

All experiments were done in triplicates and for all results the standard deviation was calculated and shown in form of error bars in the figures. The significance was calculated using an unpaired Student’s T-test and the resulting p-values are indicated in the figures. Values of p < 0.05 were considered to be statistically significant.
3. RESULTS

3.1 PIWI like genes are expressed in human bone marrow

The expression of human PIWI like genes – PIWIL1, PIWIL2, PIWIL3 and PIWIL4 in human CD34\(^+\) bone marrow cells (CD34\(^+\)BM) and in bone marrow mononuclear (BMNC) cells was checked using qRT-PCR (figure 7A). Firstly, PIWIL1 expression was undetectable in CD34\(^+\)BM and BMNCs using our real time probe. This could be due to the efficiency of the probe in use or the actual lack or low levels of PIWIL1 transcripts in the aforementioned cell types. PIWIL2 was expressed in both populations, but its expression was higher in the differentiated BMNCs. PIWIL3 levels were detectable but low in both populations, while PIWIL4 was expressed, with no significant difference between the two populations assayed.

Figure 7. Expression of PIWI genes in Hematopoiesis (A) PIWIL4 is the highest expressed PIWI gene in human hematopoietic cells (B) PIWIL4 is ubiquitously expressed in various hematopoietic lineages. Values are shown in \(\Delta C_T\) therefore, a higher value implies lower expression. To show a positive relation between bar height and higher expression, y-axis has been reversed. **p<0.001, ns – not significant.

When comparisons were made amongst the PIWIL genes PIWIL2 and PIWIL4 were relatively higher expressed in both CD34\(^+\)BM and BMNCs compared to PIWIL3 and PIWIL1 (which was undetectable). Comparable to published findings in primitive hematopoietic cells in mice (Nolde, Cheng et al. 2013), PIWIL4 was the highest expressed PIWI like gene in human CD34\(^+\) BM stem progenitors, showing over a 100 fold higher expression than PIWIL2.
Results

(p<0.001). Such a difference in expression between PIWIL4 and PIWIL2 was not observed in mature bone marrow mononuclear cells (BMNC).

### 3.2 PIWIL4 is highly and ubiquitously expressed in human hematopoietic stem-progenitors and differentiated cells

Since PIWIL4 was found to be the highest expressed human PIWI like gene, we further analyzed its expression in the various human BM and peripheral blood (PB) hematopoietic subpopulations. For this, we subjected normal human BM and PB to FACS (Fluorescence activated cell sorting) and sorted CD34+ CD38- HSC enriched population, total CD34+ stem progenitors, CD34+ CD38+ cells, CD19+ B-cells, CD3+ T cells, CD33+ myeloid cells (from BM) and terminally differentiated CD15+ cells (from PB). Using quantitative real time PCR we found (Fig 9B) that CD33+ myeloid cells and CD15+ mature NK cells showed a trend towards higher expression while lymphoid cells in the form of CD3+ T cells and CD19+ B cells showed a trend towards lower expression (figure 7B). But, this difference was not statistically significant. If the sample size were to increase, maybe such a trend could reach significance. But, with the current data, it could be concluded that PIWIL4 is expressed ubiquitously in all the aforementioned hematopoietic cells. Although the importance of PIWI proteins has been widely highlighted in stem cells (Cox, Chao et al. 1998; Cox, Chao et al. 2000), no bias towards stem cells versus differentiated cells was observed in the expression pattern.

### 3.3 PIWIL2 expression is not significantly deregulated in AML

Having checked the expression of PIWIL2 and PIWIL4 in normal CD34+BM and BMNCs, we subsequently investigated their expression levels in leukemic patient samples and drew comparisons against their relative expression in normal hematopoietic tissue. PIWIL2 expression was first re-analyzed from published microarray data generated via using two different probe sets (220686_s_at and 217421_at). (data not shown) Both probe sets yielded similar results i.e. in both AML and ALL patients, PIWIL2 had a broad range of
expression but was generally found to be either lower or equally expressed compared to normal human bone marrow. This result was confirmed through quantitative real time PCR where 40 AML and 18 ALL patients were analyzed (figure 8).

![Figure 8](image)

**Figure 8. Expression of PIWIL2 is not significantly deregulated in AML and ALL patients**

Real time PCR in patient samples shows no significant deregulation of PIWIL2 compared to bone marrow MNCs.

Since all the AML patients assayed were cytogenetically normal patients, they were compared with BMNCs and not BM CD34+ cells since CN AML patients are known to be CD34-: First, PIWIL2 expression could not be detected in all patients (7 out of 40 in AML; 5 out 18 in ALL patients). Secondly, patients that did exhibit PIWIL2 expression (33 out of 40 in AML; 13 out 18 in ALL patients) showed a broad range of expression but on an average showed expression levels which were not significantly different from BMNCs.

### 3.4 PIWIL4 expression is significantly deregulated in AML

All AML (n=68) and ALL (n=18) patients analyzed were positive for PIWIL4 expression. qRT-PCR results showed that both AML and ALL patients had a broad range of PIWIL4 expression (figure 9A). Albeit some ALL patients exhibited a 15 to 25 fold increase in expression compared to healthy tissue, majority of ALL patients showed a similar expression to normal bone marrow mononuclear cells. In AML patients, the observation was more pronounced,
where *PIWIL4* was overexpressed in more than 72% of the AML patients (p<0.0001) while 28% of the patients showed no significant difference compared to normal BMNCs and CD34+BM stem progenitors. Published microarray data from online sources such as Hemaexplorer (Bagger, Rapin et al. 2012; Bagger, Rapin et al. 2013) was also re-analyzed for comparisons with our own microarray and qRT-PCR data (figure 9B).

Figure 9. **Expression of PIWIL4 is deregulated in AML** (A) qRT PCR results - *PIWIL4* is significantly deregulated in AML as opposed to ALL patients (B) Published array data when re-analyzed shows that *PIWIL4* expression is deregulated when compared to HSCs, stem-progenitors (HSPCs), polymorphonuclear cells (PMNCs) and CD14+ monocytes. ***p<0.0001

Again, *PIWIL4* expression was significantly higher in AML patients (p<0.0001) when compared to HSCs, stem-progenitors (HSPCs), polymorphonuclear cells (PMNCs) and CD14+ monocytes. These data demonstrated that *PIWIL4* mRNA levels are upregulated in AML.
3.5 *PIWIL4 expression is highly deregulated in MLL translocation harboring AML patients*

Since AML patients had a broad range of *PIWIL4* expression and a small cohort of AML patients showed a considerably higher expression compared to the majority of AML patients analyzed, and to better understand these patients and their relationship with *PIWIL4* expression we dissected the AML patients into their respective subtypes. We observed the following *PIWIL4* was minimum two-fold to several fold higher expressed in 87% of cytogenetically normal (CN)-AML patients compared to normal CD34⁺ BM cells (n=38; p<0.0001). (figure10)

![Figure 10. PIWIL4 expression in subtypes of AML patients. PIWIL4 was found to be highly deregulated in MLL-AF9 translocation harboring patients, where its expression was 8 fold higher than the average expression all other subtypes. **p<0.001](image)

Expression levels of *PIWIL4* were independent of the mutational status of the patients, however, *NPMc⁺* positive patients showed a trend towards higher *PIWIL4* expression compared to *FLT3⁺* CN-AML patients. Similar to CN-AML patients, 62% of abnormal karyotype AML patients showed two-fold higher expression compared to normal CD34⁺ BM cells. Notably, in *MLL-AF9* rearranged AML patients, which are clinically considered as poor prognosis,
Results

PIWIL4 was 64-fold higher expressed in 90% (p<0.0001) of the patients and 8-fold higher expressed compared to average expression of inv16 (CBFB-MYH11), PML-RARA and CN-AML patients (p<0.001). Furthermore, promyelocytic leukemia patients harboring the PML-RARA translocation, clinically considered as having a favorable prognosis, showed the lowest expression of PIWIL4.

3.6 PIWIL4 expression is associated with MLL-AF9 AML

To establish a relationship between high PIWIL4 expression and MLL leukemia we subsequently assayed the expression of PIWIL4 in AML cell lines (n=3; figure 11A), which would be representative of the previously assayed patients samples. Therefore, the following cell lines were chosen for analysis - THP-1 (MLL-AF9), MONO MAC-6 (MLL-AF9), OCI-AML3 (NPMc+), OCI-AML5 (E2A-PBX1), EOL-1 (FIP1L1-PDGFRA) and HL-60 (NRAS mutation). RAJI, a Burkitt lymphoma cell line was chosen as a negative control since ALL patient samples had shown low expression levels of PIWIL4. Similar to observations made in patient analysis, the majority of AML cell lines assayed showed significantly higher expression compared to BMNCs and CD34+BM stem progenitors. As surmised from the patient data, amongst all the cell lines assayed, the MLL translocation harboring cell line MONO MAC-6 showed the highest expression of PIWIL4 (16 fold higher vs BMNC, p<0.0001). Similar to observations made in promyelocytic leukemia patients, HL-60, a promyelocytic cell line showed no significant difference in PIWIL4 expression when compared to healthy tissue while RAJI, the Burkitt lymphoma cell line showed no detectable expression of PIWIL4. To further bolster the potential association between high PIWIL4 expression and MLL leukemia we analyzed published microarray data on CD34+ cord blood transduced with AML1-ETO, CBFB-MYH11 and MLL-AF9 translocations (figure 11B) (Wei, Wunderlich et al. 2008). The analysis revealed that PIWIL4 expression was highly deregulated in MLL-AF9 immortalized CD34+ cord blood cells compared to CD34+ cord blood transduced with AML1-ETO and CBFB-MYH11.
translocations that are not sufficient for immortalization of CD34+ cord blood cells.

Figure 11. PIWIL4 expression in AML cell lines and human cord blood cells transduced with AML translocations. (A) PIWIL4 was found to be deregulated in AML cell lines especially in MLL-AF9 translocation harboring THP-1 and MonoMac-6. RAJI, a Burkitt lymphoma cell line shows no detectable expression (n.e.) of PIWIL4 (B) PIWIL4 expression was deregulated in cord blood immortalized with MLL-AF9. ***p<0.0001, **p<0.001, *p<0.05

3.7 PIWIL4 shows a broad range of expression in different MLL acute myeloid leukemia

To assess if high PIWIL4 expression was prevalent in all MLL rearrangement harboring patients or if it was a MLL-AF9 restricted phenomenon, we further explored PIWIL4 expression in MLL rearrangement harboring patients using published data. Analysis of PIWIL4 expression in published microarray data of 42 MLL rearrangement harboring patients (Pigazzi, Masetti et al. 2011) revealed a uniform expression in different MLL-rearranged AML patients i.e. MLL-AF10, MLL-AF6, MLL-AF9, MLL-ENL and MLL-SEPTIN6 (figure 12A).
Figure 12. *PIWIL4* expression of *PIWIL4* in AML patients harboring various MLL translocations (A) *PIWIL4* expression is not significantly different in *MLL-AF9* patients when compared to *MLL-AF10, AF-6, ENL* and *SEPTIN6* translocations (B) *MLL-AF9* patients can be divided into high and low *PIWIL4* expressing patients (C) 353 genes are found deregulated when *PIWIL4* high expressing patients are compared to low expressing cohort

Although there were no significant differences among the various MLL-rearrangement subtypes, due to a broad range in expression, especially in the case of *MLL-AF9* patients, patients were divided according to high and low *PIWIL4* expression (two fold difference in expression) (figure12B). Analysis of the gene expression data upon dividing patients into high and low *PIWIL4* expressing patients revealed 353 genes that were differentially regulated
Results

(p<0.01, fold >1.5). Pathways such as PI3-AKT signaling, Cancer and regulation of actin cytoskeleton through Rho GTPases were the most deregulated when analyzed using KEGG array analysis. In particular, genes such as AML associated FLT3, proto-oncogene CBL and NRAS, Rho GTPase associated genes CDC42L1 (RHOU), CDC42BPA, ARHGEF6, ARHGAP27 and ARHGAP30 were some of the significantly differentially regulated genes that were upregulated in high PIWIL4 expressing patients. The aforementioned evidences lead us to devise a knockdown strategy to check the dependency of AML cells on PIWIL4 expression, especially in the case of MLL-rearrangement harboring AML cells.

3.8 Knockdown of PIWIL4 results in loss of PIWIL4 protein from nuclear and cytoplasmic compartments

In order to achieve a stable and constitutive downregulation of PIWIL4, we cloned two anti-PIWIL4 shRNA sequences (Sigma Aldrich - TRCN0000141359 or shPIWIL4-A and TRCN0000139278 or shPIWIL4-B) into the lentiviral pGreenPuro vector. shPIWIL4-A and shPIWIL4-B were used to knockdown PIWIL4 in the cell lines – THP-1 (MLL-AF9), MONO MAC-6 (MLL-AF9) and OCI AML-3 (NPMc+). RAJI (Burkitt lymphoma), which shows no expression of PIWIL4 was used as an off target control. As shown in figure 13A, loss of PIWIL4 expression could be detected at the transcript level in all the aforementioned cell lines (except RAJI which shows no detectable expression). shPIWIL4-A consistently showed a higher efficiency of knockdown in all cell lines assayed as compared to shPIWIL4-B (average knockdown among all cell lines= 67% vs 35%). At the protein level, using an anti-PIWIL4 monoclonal antibody (Promab, 10G9B11) for western blot and intracellular staining, we could show that PIWIL4 protein levels depreciated upon knockdown (Figure 13B and C). Consistent with published data on homologues of lower organism, PIWIL4 was found to be abundantly present in the nuclear compartment whilst showing comparatively lesser expression in the cytoplasmic compartment. The knockdown at the transcript level lead to a
Results

decrease in the protein levels in both the cytoplasmic and the nuclear compartments.

Figure 13. Knockdown of *PIWIL4* in AML cell lines (A) Knockdown at the transcript level measured using qRT pCR shows that shPIWIL4- gives a superior knockdown to shPIWIL4-B in AML cell lines tested. (B) A western blot using an anti-PIWIL4 antibody (Promab) also shows loss of protein levels. (C) Intracellular staining shows expression of PIWIL4 predominantly in nucleus and in cytoplasm. Knockdown results in loss of nuclear and cytoplasmic PIWIL4 expression

3.9 Depletion of PIWIL4 drastically reduces the growth of MLL rearrangement harboring AML cell lines in vitro

AML cell lines, lentivirally transduced with shRNA constructs and scrambled control were tested for their proliferative and clonogenic potential in vitro by performing liquid expansion and myeloid colony forming assay (CFC), respectively. As shown in figure 14, stable knockdown of *PIWIL4* using shPIWIL4-A in THP-1, MONO-MAC-6 and OCI-AML-3 lead to a marked depreciation in proliferative capacity in liquid culture assays (THP-1, MONOMAC-6 and OCI-AML-3 more than 65%, 65% and 70% reduction in cell number, respectively n=3). Knockdown using shPIWIL4-B, which yielded a considerably lower depletion of *PIWIL4*, expectedly lead to a lower impact on
Results

proliferative capacity of the AML cell lines tested (THP-1, MONO-MAC-6 and OCI-AML-3 - 45%, 28% and 25% reduction in cell number, respectively n= 3).

Figure 14. Proliferation of AML cells is affected upon knockdown of PIWIL4. Proliferation assay shows the adverse effect of PIWIL4 knockdown on growth rate assayed for 6 days. RAJI, the off-target control remains unaffected. ***p>0.0001, **p>0.001, *p>0.05

Colony forming capacity of PIWIL4 depleted AML cells was tested using shRNA PIWIL4-A (figure 15). THP-1, MONO MAC-6 and OCI-AML-3 showed a 55%, 65% and 52% reduction in total colony number, respectively (n=3). On the other hand, RAJI - a Burkitt lymphoma cell line that showed no detectable expression of PIWIL4 - used as negative control and off target control for this experiment, showed no decrease in cell growth or colony forming capacity in vitro indicating that high expression of PIWIL4 is important for the proliferative and clonogenic potential of the tested AML cells.
Figure 15. Colony forming potential of AML cells is affected upon knockdown of PIWIL4. Colony forming assay shows the adverse effect of PIWIL4 knockdown on colony forming potential assayed for 7 days. RAJI, the off-target control remains unaffected. ***p<0.0001, **p<0.001, *p<0.05

3.10 PIWIL4 knockdown adversely effects growth of AML cells in vivo

To test whether PIWIL4 depletion inhibits the proliferative potential of these cell lines in vivo, 1x10^6 THP-1 (Scrambled / shPIWIL4-A) were injected intravenously (IV) into sub-lethally irradiated NSG mice. NSG mice injected with scrambled-control containing cells succumbed to human leukemia with the median latency of 30 days (n=8) (figure 16A). All the mice were highly engrafted showing multiple organ infiltration and splenomegaly (Scr spleen wt. - 343± 47 mg vs shPIWIL4 spleen wt. – 205± 40 mg, p=0.04). Mice injected with shRNA containing cells showed no engraftment in PB for up to 4 weeks and an average of 5% engraftment in PB at 8 weeks. Death of mice was significantly delayed to a median latency of 62 days (average n=8, p=0.001). Experiments performed with MV4-11, another MLL translocation harboring cell line yielded similar results. Mice injected with scrambled-control containing cells succumbed to human leukemia with the median latency of 30 days, while experimental mice died with a median latency of 48 days (n=4, p=0.001), with decreased splenomegaly (figure16B).
Results

Figure 16. (A) Kaplan Meier survival curve showing the increased survival period (each drop representing one mouse) and (B) bar graph showing decreased spleen weight of mice injected with PIWIL4 depleted AML cell lines THP-1 and MV4-11. 1x10⁶ cells were injected per sample (scrambled v shRNA-A), per mouse. Mice injected with cells containing scrambled control perished with a median latency of 30 days while mice with shRNA transduced MV4-11 and THP-1 cells perished with a latency of 48 and 62 days respectively. Spleen weight was significantly reduced in mice injected with shRNA transduced cells **p<0.001, *p<0.05

3.11 PIWIL4 knockdown effects cell cycle but not apoptosis

To attribute the decreased in-vitro and in-vivo proliferation to cell cycle or cell death, a Bromodeoxyuridine (BrdU) based cell cycle assay and an Annexin V based apoptosis assay were performed on PIWIL4 depleted cells (figure 17A and B). In THP-1 cell line, at an early time point (at day 2 post-sorting or day 4 post-transduction), BrdU assay could demonstrate that PIWIL4 depletion induced a mild but significant increase in the percentage of cells in the G0/G1 phase (+12%; p>0.002) and a significant decrease in the percentage of cells in the S phase (-14%; p>0.0005) and G2M phase (-3.5%; p>0.001). (figure 17A)
3.12 Depletion of PIWIL4 in human cord blood HSPCs does not impact in vitro growth

Since depletion of PIWIL4 depreciated AML growth in vitro and in-vivo, next the importance of PIWIL4 expression for normal CD34+ HSPC growth and differentiation was assayed by knocking down PIWIL4 in CD34+ human cord blood cells. Cells were transduced with either scrambled or shPIWIL4-A (n=4) and knockdown was assessed using qRT-PCR. An average of 36% reduction in PIWIL4 transcript levels could be achieved (figure 18A). The effect on normal growth and differentiation was assayed using human myeloid colony forming assay where cells were plated in methylcellulose (#H4434) complete with necessary human cytokines. Colonies were identified as – colony forming unit granulocyte erythrocyte monocyte megakaryocyte (CFU-GEMM) which originate from primitive multipotent progenitor cells, further differentiated -

Figure 17. Cell cycle and Apoptosis assay. (A) Cell cycle assay showing a significant increase in the percentage of cells in the G0/G1 phase in PIWIL4 depleted THP-1 AML cells. (B) Apoptosis assay showing no significant changes in percentages of early, late apoptotic or dead cells.

No significant effect was observed in the percentage of dead, early apoptotic or late apoptotic cells between scrambled and shRNA transduced cells upon performing Annexin-V assay (at day 2 post-sorting) (figure 17B).
Results

colony forming unit granulocyte (CFU-G) or colony forming unit macrophage (CFU-M), primitive erythroid colonies or blast forming unit-erythroid (BFU-E) and further differentiated erythroid colonies (CFU-E).

Figure 18. PIWIL4 depletion has no adverse effect on cord blood CD34\(^+\) growth in vitro. (A) Knockdown at the mRNA level, (B) Total colony count (C) Total number of cells generated from colonies (D) Morphology of CFC colonies (40X). PIWIL4 depletion in cord blood cells does not have an impact on the number of myeloid and erythroid colonies, the total cell number or the morphology of the cells.

When scrambled colonies were compared to shPIWIL4-A at day 14, no differences in number or size of CFU-GEMM, CFU-G/M, BFU-E/CFU-E colonies could be found (figure 18B). When dishes were sacrificed and total number of cells counted, no significant difference was observed between scrambled and the shRNA arm (figure 18C). This result gave an indication that normal human CD34\(^+\) HSPC growth depend less on PIWIL4 expression in vitro than AML cells.
3.13 Depletion of PIWIL4 in primary MLL-AF9 patient BM drastically effects colony forming ability in vitro

To test this hypothesis further, depletion of PIWIL4 was performed for MLL-AF9 patient primary bone marrow cells. 25 x 10^6 Bone marrow cells from were transduced with either scrambled or shPIWIL4-A patients (figure 19).

Figure 19. PIWIL4 depletion adversely effects in vitro growth of MLL-AF9 harboring patient BM cells. Figure showing (A) Knockdown at the mRNA level, (B) Colony count (C) Total number of cells generated from colonies (D) Morphology of CFC colonies (40X) (E) Cytomorphology of the colony cells (100X). PIWIL4 depletion in MLL-AF9 patient bone marrow cells leads to loss of colony forming potential in vitro.
Sorted cells were checked for knockdown and their growth/colony forming ability was assayed using CFC. An average knockdown of 60% could be achieved in the MLL-AF9 patients versus scrambled control (figure 19A). After 14 days, colonies were scored as – blast colonies and blast clusters and erythroid BFU-E/CFU-E colonies. In the scrambled arm versus the shRNA, blast colonies and blast clusters were more than 4 and 6 times in number, respectively (figure19B). Moreover, the sizes of the colonies were larger in the scrambled arm. No difference was observed in number of normal BFU-E/CFU-E colony numbers. Upon sacrificing the CFC dishes, an average of 5x10⁵ cells were attained in the scrambled arm while only 5x10⁴ cells were found in the shPIWIL4-A arm (figure 19C). Therefore, the PIWIL4 depleted MLL-AF9 harboring patient BM cells showed a 10 times decrease in growth in vitro compared to control cells. Visual analysis of cytospins gave an impression of increased differentiation in the shRNA arm compared to scrambled control.

**3.14 Depletion of PIWIL4 leads to global loss of H3K9me3 and localized gain in H3K4me3**

For delineating the possible reason for dependence of AML cell growth on PIWIL4 and in accordance with its known epigenetic function in lower organisms, histone modifications were assayed in PIWIL4 depleted cells. As PIWIL4 is important for heterochromatin formation (in lower organisms) and its association with H3K9me3 and H3K4me3 histone modifications in germ stem cells of lower organisms is well established, chromatin immunoprecipitation-deep sequencing (ChIP-seq) for these modifications was performed. 2x10⁶ THP-1 cells transduced with scrambled or shPIWIL4 on day 10 post-transduction were used for performing ChIP for H3K9me3 and H3K4me3 followed by library generation and deep sequencing.
Figure 20. PIWIL4 depletion in AML cells leads to global loss of H3K9me3 and partial loss in H3K9me2 marks. (Upper panel) ChIP-seq analysis of H3K9me3 marks in THP-1 cells, showing enrichment of peaks in various genomic regions in scrambled vs shPIWIL4-A transduced arms. (Lower panel) Western blot showing global H3K9me3 levels in scrambled vs shPIWIL4-A transduced AML cell lines. ChIP-seq for H3K9me3 in THP-1 cells shows a reduced enrichment of H3K9me3 peaks globally in the shRNA arm as compared to scrambled control. Western blot for H3K9me3 reveals the extent of the global loss of H3K9me3 while a partial loss in H3K9me2 is also observed.

ChIP-seq analysis showed that THP-1 cells transduced with shPIWIL4-A have lower global H3K9me3 levels as compared to cells transduced with scrambled control (figure 20). This phenomenon was also observed globally as well as in regions 1000 bp upstream of transcription start site of genes (TSSup1k), indicating a differential H3K9me3 level at the regulatory elements of genes. As shown in the lower half of figure 20, a global loss of H3K9me3 was also observed in western blots for AML cells transduced with shPIWIL4-A i.e. in THP-1, MONO-MAC-6, MV4-11 and OCI-AML-3. RAJI the off-target control,
which in previous experiments showed no effect on proliferation upon PIWIL4 knockdown, did not show any effect on H3K9me3 levels. Additionally, minor loss of H3K9me2 was also observed in western blots for the aforementioned cells lines, except MV4-11.

Figure 21. PIWIL4 depletion in THP-1 cells leads to localized gain of H3K4me3. ChIP-seq for H3K4me3 in THP-1 cells shows gain of H3K4me3 mark at localized areas such as 1kb upstream of transcription start site (TSS1k).

In contrast to the findings of H3K9me3, H3K4me3 levels were generally higher in shPIWIL4 transduced cells than in scrambled control. This trend was also observed in the TSSup1k. (figure 21). As a result of this contrasting enrichment of H3K9me3 and H3K4me3, TSSup1k regions were analyzed to see if those regions that showed lower H3K9me3 levels would simultaneously show higher H3K4me3 levels.

Though this was an interesting observation at a global level, there were no TSSup1k regions that showed this alternating methylation levels. Due to higher H3K4me3 levels in the absence of PIWIL4 and the lack of observable overlapping regions between sites that had higher H3K4me3 levels and lower
H3K9me3, it is possible that alternative machinery dictates establishment of H3K4me3, especially in the absence of PIWIL4.

3.15 **PAZ domain lacking mutant of PIWIL4 is unable to rescue the loss of proliferative potential and H3K9me3 marks**

In order to establish that decrease in proliferation and global loss of H3K9me3 marks were a direct result of PIWIL4 depletion, two PIWIL4 mutants were generated. The first mutant, a shRNA resistant mutant or S-mutant, was mutated at the nucleotide level such that shPIWIL4-A does not bind with its mRNA. No change was induced in the amino acid sequence, thereby generating a functional PIWIL4 protein which cannot be knocked-down with shPIWIL4-A. The second mutant was a double mutant generated by deleting the piRNA binding PAZ domain and mutating it similar to the S-mutant rendering it resistant to shPIWIL4-A. This mutant was called the piRNA mutant or P-mutant (mentioned in detail in *Methods* section 3.4.3).

THP-1 cells were transduced with scrambled control and shPIWIL4-A respectively. Each arm was then further transduced with empty vector (pCDH) control, S-mutant and P-mutant to achieve a double transduction. A proliferation assay was performed and cells were harvested for western blot at the end of the assay. In the proliferation assay the adverse effect on growth brought about by PIWIL4 knockdown was rescued through overexpression of the shRNA resistant PIWIL4 cDNA or the S-mutant (shR S-mut) (figure 22A). Such a rescue was not observed in PIWIL4 depleted cells that were transduced with the P-mutant (shR P-mut), which grew 4 times slower (p<0.0001) compared to shRNA and S-mutant double transduced cells (shR S-mut). (figure 22B). Overall, the S-mutant transduced cells (both scrambled and shRNA) grew significantly faster than their respective empty vector and P-mutant counterparts.
Results

Figure 22. PIWIL4 overexpression rescues effect on growth and global H3K9me3 levels. (A) Proliferation assay of PIWIL4 depleted THP-1 cells rescued using wild type shRNA resistant PIWIL4 (‘S-mut’) versus empty vector (B) Proliferation assay of PIWIL4 depleted THP-1 cells rescued using PAZ deleted mutant (‘P-mut’) versus empty vector (C) Western blot showing effect on H3K9me3 in PIWIL4 depleted THP-1 cells rescued with S-mut and P-mut. Expression of S-mutant in THP-1 cells transduced with shPIWIL4-A helped the cells overcome the effect of PIWIL4 depletion on growth and global H3K9me3 levels. The P-mutant lacking the piRNA binding PAZ domain could not rescue THP-1 cells from the adverse effects of PIWIL4 depletion.

This effect was mirrored in global H3K9me3 levels observed by western blot (figure 22C). As expected, within the empty vector control arm, global H3K9me3 levels were reduced in the shRNA versus scrambled transduced THP-1 cells. But this effect was rescued by the expression of the S-mutant such that H3K9me3 levels were similar in scrambled control versus shRNA. On the other hand, similar to its effect on growth, the P-mutant could not rescue the effect on H3K9me3, where global levels were lower in the shRNA arm as compared to the scrambled.

In summary, effect of PIWIL4 depletion on growth and global H3K9me3 levels could be rescued through overexpression of PIWIL4 cDNA which was resistant
to the shRNA (S-mutant). On the other hand such a rescue was not observed when a PAZ domain deleted mutant of PIWIL4 was overexpressed.

3.16 Depletion of PIWIL4 leads to differential gene expression

To understand the impact of PIWIL4 depletion on gene expression, RNA-Seq was performed in THP-1 cells. Comparison of PIWIL4 knockdown versus scrambled control revealed 2711 differentially expressed genes at a p-value of 0.05 and a false discovery rate (FDR) q-value of 0.05. As shown in the heat map the number of upregulated and downregulated genes were nearly similar in number i.e. 47% of the genes were upregulated while the other 53% were downregulated compared to scrambled control (figure 23). The pathway analysis of the differentially regulated genes performed using KEGG analysis yielded significantly deregulated pathways, the top few of which are listed above (table, figure 23).

Figure 23. Differential gene expression and disruption of pathways upon PIWIL4 knockdown in THP-1 cells. (Left panel) Heatmap showing differentially expressed genes. (Right panel) List of top five deregulated pathways. PIWIL4 depleted THP-1 cells were analyzed using deep sequencing. A heatmap shows 2711 genes that showed changes in gene expression in shRNA arm vs scrambled control (fpkm $\geq$ 5; p$\leq$0.005, Fold $\geq$0.5), while the top five deregulated pathways are listed.

In the PI3K-Akt pathway, one of the highest ranked pathways, 90 genes were differentially expressed. It’s widely known that the PI3K-Akt pathway is a pro-survival pathway promoted by TGF-beta cytokines in tumor cells. Therefore, in PIWIL4 knockdown THP-1 cells, reduced proliferative ability could be
attributed to the down-regulation of this pathway. Indeed, genes essential for the activation of the PI3K-Akt pathway such as TGFBR1, TGFB1, JAK1, PDK1 and AKT1S1 were downregulated in the shRNA arm compared to scrambled control. Furthermore, upstream regulator of PI3K-AKT and TGF-beta signaling – such as Toll like receptors (TLR1) and PPP1R15A respectively, and downstream targets of PI3K-Akt pathway, such as the growth and proliferation associated kinase, SGK1, were also downregulated.

Figure 24. KEGG PI3K-Akt pathway showing differentially regulated genes in PIWIL4 depleted THP-1 cells.

Genes involved in the regulation of actin cytoskeleton that are downstream of the PI3K-Akt pathway, particularly related to Rho GTPase regulation of actin cytoskeleton such as the CDC42 RAC GTPase regulators (ARHGAP22, ARHGAP26 and ARHGAP31), CDC42 binding protein alpha (CDC42BPA), CDC42 like GTPase (RHOU), and downstream targets of Rho GTPase pathways such as LIMK2, ENAH, integrins, actins and myosin components showed deregulated expression. As a consequence of the downregulation of the PI3K-Akt pathway, the MAP kinase pathway and cell cycle were also
affected. KEGG analysis showed 31 deregulated genes involved in the cell cycle pathway. The expression of genes such as E2F, RB1, p21, CDK4/6, CYCLIN D1 and CDK2 was perturbed which could be the ultimate cause behind the decrease in the G1 to S phase transition (as shown in section 4.11), leading to an adverse effect on the proliferative potential of PIWIL4 depleted AML cells. Contrary to results obtained in ChIP-seq and western blot analysis, epigenetic enzymes or factors that establish H3K4me3 marks did not show any deregulated expression while SETDB1, an enzyme which establishes H3K9me3 marks, showed a marginal increase in expression. In addition, Thymidine DNA glycosylase (TDG) an enzyme important for TET dependent DNA demethylation, showed decreased expression in the PIWIL4 depleted samples.

Figure 25. KEGG Actin cytoskeleton pathway enriched in genes expressed in both low PIWIL4 expressing MLL-AF9 patients and PIWIL4 depleted THP-1 cells.

When differentially regulated genes in PIWIL4 depleted THP-1 cells were manually enriched for commonalities against genes deregulated in low PIWIL4
expressing *MLL-AF9* patients (described in section 4.7), the PI3K-Akt and actin cytoskeletal pathways were again found to be deregulated (figure 25). Phosphoinositide kinase subunit and adapter protein (*PIK3R5* and *PIK3AP1*), JAK1, the tyrosine kinase upstream to PI3K-Akt3 pathway, secondary messenger proteins such as guanine nucleotide binding proteins, CDC42 related proteins – *CDC42BPA, RHOU, ARHGAP31, LIMK2, ENAH*, integrin *ITGA7* and myosin light chain kinases were all significantly downregulated at the mRNA level in both PIWIL4 depleted THP-1 cells and low PIWIL4 expressing *MLL-AF9* patients.

An important aspect of piwi proteins is its association with repetitive elements and TEs. Since the read depth of the RNA-seq was not sufficient, repetitive elements could not be checked for differential expression.

### 3.17 Changes in histone marks in promoter/regulatory regions show low correlation with gene expression changes

To assess if global H3K9me3 changes and localized H3K4me3 changes had an impact on gene expression, ChIP-seq data for histone modifications was compared to gene expression data. For this, the differentially expressed genes list from the RNA-seq data was compared to the list of genes that showed H3K9me3 hypomethylation or H3K4me3 hypermethylation in their promoter/regulatory region i.e. 1kb to 8kb upstream of TSS. Since H3K9me3 mark is a repressive mark, an increase in expression is expected upon its loss from regulatory regions. Conversely, H3K4me3 is a euchromatic mark and its hypermethylation is associated with increased gene expression. Although 183 genes (7% of total deregulated genes) showed H3K9me3 hypomethylation in the promoter/regulatory regions, only 99 (8% of upregulated genes) showed statistically significant increase in expression. Similarly, gain in H3K4me3 marks found in regulatory regions of 81 genes (3% of total deregulated genes), coincided with increase in expression in only 38 (3% of upregulated genes).
Figure 26. Venn diagram showing relation between differential gene expression and changes in (A) H3K9me3 (B) H3K4me3 histone methylation. Venn diagram showing the overlap between differentially regulated genes which were upregulated and genes which lost H3K9me3 or gained H3K4me3 marks at 1 to 8 kb upstream of transcription start site (TSS).

When genes showing H3K9me3 hypomethylation and H3K4me3 hypermethylation in their regulatory regions were analyzed independent of their expression status through KEGG pathway analysis, again, PI3K-Akt pathway was one of the highest deregulated pathways. Additionally, pro-senescence pathways such NF-KB and the Rap1 signaling pathway were upregulated upon loss of H3K9me3 in their regulatory regions. When common genes from RNA seq of PIWIL4 depleted THP-1 cells and low PIWIL4 expressing MLL-AF9 patients were compared to genes that lost H3K9me3 in their regulatory regions, growth suppressing genes such as insulin receptor binding protein GRB10, tumor suppressor NEO1, NF-KB associated transcription factor and enzyme HIVEP3 and PRKCQ respectively, were found to be upregulated. Cytoskeleton regulatory protein CDC42BP was also found to be upregulated while also exhibiting a loss of H3K9me3 marks on its regulatory region.

Although loss of H3K9me3 marks was global in nature, only 7% of deregulated genes showed loss in their regulatory regions. This is indicative of a weak correlation between gene expression changes and H3K9me3 loss in protein coding regions of the genome. Despite this over 2711 genes were found deregulated in the shRNA arm compared to scrambled control. This effect could be attributed to the impact of loss of H3K9me3 in regions that fall outside the purview of the ‘reference sequence’ i.e. repetitive elements. In an ideal
scenario, the analysis of repetitive elements from RNA-seq data could be superimposed on regions that lose H3K9me3 to identify the role of PIWIL4 and its associated machinery in TE silencing. As mentioned before in the previous section, the lack of depth in the sequencing did not permit such an analysis so far.

3.18 Overexpression of PIWIL4 in THP-1 and RAJI cell lines leads to increased proliferation in vitro

Due to the relation of PIWIL4 with the proliferation associated PI3K-Akt pathway, an overexpression of PIWIL4 was performed in high endogenous PIWIL4 expressing THP-1 cells and RAJI cells that have no detectable PIWIL4 expression. As shown in figure 27A, an increase of 10 and 3000 fold overexpression of PIWIL4 were achieved in THP-1 and RAJI cells, respectively.

RAJI showed a considerably higher increase in PIWIL4 overexpression when normalized to its empty vector control. This could be attributed to the otherwise lack of detectable levels of PIWIL4 in RAJI cells.

In proliferation assays on day 6 both THP-1 and RAJI showed 1.4 and 1.6 times increase in cell number, respectively, further bolstering the concept of a pro-survival and proliferation related role of PIWIL4 in AML cells (figure 27B).
Results

Figure 27B. Overexpression of PIWIL4 in AML cell lines THP-1 and RAJI leads to increased cell growth. PIWIL4 overexpressing cells were analyzed using deep sequencing. (A) A heatmap shows 2559 genes that showed changes in gene expression.

3.19 piRNA like RNA is expressed in AML cells - THP-1 and MV4-11

piRNA are associated with the epigenetic function of PIWIL4 in lower organisms, but the presence of piRNA has not been established in human leukemic cells. Therefore to establish if piRNA or piRNA-like RNA are expressed in AML cells, initially a piRNA microarray was performed using wild type THP-1 cells and cells transduced with scrambled or shPIWIL4-A (figure 28A). Out of 23,000 probe-sets for known and predicted piRNA, 9,957 unique piRNA could be detected in wild type THP-1 cells.

To assess if knockdown of PIWIL4 perturbs the expression profile of known and predicted piRNA, microarray profile of THP-1 cells transduced with shRNA was compared to that of scrambled transduced cells. 1105 piRNA were found deregulated in the shRNA arm (p>0.05, fold change<2) of which 510 were upregulated and 595 downregulated compared to scrambled control.

Models that describe mechanism of PIWI protein function state that the PIWI associated epigenetic machinery is guided to loci through sequence complimentarity of piRNA (see section 1.3.6).
Results

Figure 28. Array based expression of piRNA in PIWIL4 depleted THP-1 cells. (Upper panel) Heatmap showing differentially regulated known and predicted piRNA in THP-1 cells transduced with scrambled vs shPIWIL4-A. (Lower panel) piRNAs deregulated upon PIWIL4 depletion compared to scrambled control. PIWIL4 depleted THP-1 cells were analyzed using a piRNA microarray. Differentially regulated piRNA were mapped to coding regions of the human genome (using UCSC genome browser - http://genome.ucsc.edu/). The genes that piRNA mapped to were compared to the genes which were deregulated upon PIWIL4 depletion in THP-1 cells (RNA-seq data, Section 3.16).

In accordance with this model, using the reference sequence from UCSC genome browser, deregulated piRNAs were mapped to the human genome. Most deregulated piRNA sequences mapped to more than one locus in the genome with 100% complementarity for the entire stretch of the sequence. Some piRNA mapped to intronic/non-geneic regions while others mapped to coding sequences of single or several genes. In all, deregulated piRNAs were found to map to 811 unique regions consisting of pseudogenes, protein coding genes and loci which were yet unannotated. Out of the 811 deregulated

<table>
<thead>
<tr>
<th>ProbeName</th>
<th>Up/DN regulated</th>
<th>Sequence</th>
<th>Loci mapped to</th>
<th>Gene mapped to</th>
<th>RNA-seq status of gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQ579883P</td>
<td>DN</td>
<td>ACAGCTGATATCTCCTGGTGCTATCCCCA</td>
<td>Chr17</td>
<td>CDC42BPA</td>
<td>Upregulated</td>
</tr>
<tr>
<td>DQ593902P</td>
<td>DN</td>
<td>TGGTCGCGCCGGTTTCTTAAGCGGTCTGTA</td>
<td>Chr17</td>
<td>MYH10</td>
<td>Upregulated</td>
</tr>
<tr>
<td>DQ601792P</td>
<td>DN</td>
<td>TCTAGGAGGTAGGAGCGGCCTTCGCCGCC</td>
<td>Chr 1</td>
<td>ENAH</td>
<td>Downregulated</td>
</tr>
<tr>
<td>DQ601642P</td>
<td>DN</td>
<td>AAAGTGAGGAGGCTCCGCCGGCCCGGCC</td>
<td>Chr6</td>
<td>PHACTR1</td>
<td>Downregulated</td>
</tr>
<tr>
<td>DQ600093P</td>
<td>DN</td>
<td>TACCACCAGTTCCCTGAACGCCCTTC</td>
<td>Chr3</td>
<td>HYAL3</td>
<td>Upregulated</td>
</tr>
<tr>
<td>DQ579111P</td>
<td>DN</td>
<td>TGAAGGTGTTTCTGGAACCTTGATC</td>
<td>Chr6</td>
<td>HIST1H4E</td>
<td>Upregulated</td>
</tr>
<tr>
<td>DQ597391P</td>
<td>DN</td>
<td>TGCTAGCTCAGGGTTGTGGTGACCTGCACC</td>
<td>Chr12</td>
<td>TMTC2</td>
<td>Upregulated</td>
</tr>
</tbody>
</table>
piRNAs, 83 mapped to genes that were found to be deregulated in RNA seq of PIWIL4 depleted THP-1 cells.

This list of genes analyzed through KEGG pathway analysis yielded the growth associated RAS signaling, actin/cytoskeletal regulation and Rap1 signaling pathway. Notably, deregulated piRNA mapped to coding regions of cytoskeletal regulation associated genes such as \textit{CDC42BPA}, \textit{MYH10}, \textit{PHACTR1} which also suffered a loss of H3K9me3 upon PIWIL4 depletion. These genes were also found to be deregulated in \textit{MLL-AF9} patients and AML cell lines. The figure above summarizes a few deregulated piRNAs that were mapped to coding sequences of genes that were also identified via RNA-seq as differentially expressed genes in PIWIL4 depleted THP-1 cells.

To further bolster evidence in favor of existence of piRNA-like RNA in human AML cells, in addition to an array based method, deep sequencing of small RNA (25-32nts) was performed using RNA from \textit{MLL} translocation harboring MV4-11 cells transduced with scrambled control and shRNA. Using Bowtie, the sequenced library was depleted of known microRNA and ribosomal RNA sequences and sorted for 26 to 32 nucleotides long sequences with a sequence bias of uridine at position 1 or adenine at position 10, a characteristic feature of piRNA. Firstly, the percentage of reads sized 26-32 nucleotides that mapped to the genome were depleted in the shRNA arm compared to scrambled control (upper half of figure 29). In the scrambled arm, over 1 million reads were present in the fraction of 26-32 nucleotides long RNA while in the shRNA arm this number was reduced to 600,000. Furthermore, reads that harbored the characteristic sequence bias of uridine at position 1 or adenine at position 10 were depleted in the shRNA treated arm compared to control (lower half figure 29).
Figure 29. Small RNA deep sequencing for identifying piRNA like RNA in AML cells. (Top panel) Distribution of reads according to length of RNA in base pairs in scrambled vs shPIWIL4-A arm. (Bottom panel) Distribution of reads containing uridine base at position 1 and Adenine at position 10 in scrambled vs shPIWIL4-A arm. Small RNA deep sequencing in MLL rearrangement harboring MV4-11 cells yielded 26-32 nucleotides long RNA with characteristic features such as Uridine residue at position 1 and Adenine at position 10. The percentage of matching reads of total as well as sequence specific piRNA like RNA were reduced in PIWIL4 depleted AML cells.

Although more reads were present in the scrambled arm, the shRNA arm was more enriched for known piRNA. Out of the 600,000 reads, the shRNA arm contained 147,000 reads (24% of total reads) of known piRNAs when matched to piRNA databases RNAdb and piRNA bank.
Figure 30. Known piRNA found in AML cell line MV4-11 through small RNA deep sequencing. (Top panel) piRNA found in scrambled transduced MV4-11 cells upon small RNA deep sequencing and matching reads to piRNA databases. (Bottom panel) piRNA found in shPIWIL4-A transduced MV4-11 cells upon small RNA deep sequencing and matching reads to piRNA databases. Small RNA deep sequencing in MLL rearrangement harboring MV4-11.
cells yielded 26-32 nucleotides that could be mapped to databases containing sequences for identified piRNA. 24% of 26-32 nucleotide long sequences found in the shRNA arm mapped to known piRNA while this number was reduced to only 11% in the scrambled arm.

On the other hand the scrambled arm was relatively less enriched for known piRNA with 113,000 reads for known piRNA (11% of total reads). In spite of this, both scrambled and shRNA arms contained known piRNA that were more enriched in one arm than the other (as shown in figure 30). PIWIL4 is classically known to bind longer - 28 to 32 nt long piRNA while other PIWI proteins, PIWIL1 and PIWIL2 preferentially bind relatively smaller piRNA (26 to 28 nt). In light of this fact, it was notable that in PIWIL4 depleted MV4-11 cells, piRNA-like RNA associated with repetitive sequences (Repbase, Genetic Information Research Institute) showed both an accumulation and a decrease at different size distributions (figure 31).

Figure 31. piRNA like RNA in MV4-11 cells mapped to repetitive sequences. 26-32 nucleotide long sequences that could be mapped to repetitive sequences using Repbase showed an increase at the 26 to 27 nucleotide length but a decrease in 28 to 30 nucleotide sized piRNA like RNA.
3.20 PIWIL4 associates with epigenetically active proteins and proteins involved in piRNA biogenesis in AML cells

Mass spectrometry results for PIWIL4 binding proteins in THP-1 cells over expressing PIWIL4 showed a host of binding partners (Table 5). As noted in previous sections, PIWIL4 was found to associate with cytokeletal proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Description</th>
<th>Unique peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cytoskeletal proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYH9</td>
<td>Cellular myosin heavy chain A</td>
<td>57</td>
</tr>
<tr>
<td>ACTN4</td>
<td>Alpha actinin 4</td>
<td>41</td>
</tr>
<tr>
<td>IQGAP1</td>
<td>IQ Motif Containing GTPase Activating Protein 1</td>
<td>25</td>
</tr>
<tr>
<td><strong>Epigenetic proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLL1</td>
<td>Mixed lineage leukemia-1</td>
<td>2</td>
</tr>
<tr>
<td>NSD1</td>
<td>Nuclear Receptor Binding SET Domain Protein 1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Miscellaneous</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHX9</td>
<td>DEAH box helicase</td>
<td>9</td>
</tr>
<tr>
<td>EF2</td>
<td>Elongation factor 2</td>
<td>8</td>
</tr>
<tr>
<td>HS90B</td>
<td>Heat shock protein HSP 90-beta OS</td>
<td>8</td>
</tr>
<tr>
<td>HS90A</td>
<td>Heat shock protein HSP 90-alpha</td>
<td>6</td>
</tr>
<tr>
<td>HSPA9</td>
<td>Heat Shock 70kDa Protein 9</td>
<td>2</td>
</tr>
<tr>
<td>DDX5</td>
<td>DEAD box helicase</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5. Binding partners of PIWIL4 in THP-1 cells. Binding partners of PIWIL4 identified using a combination of immunoprecipitation and mass spectrometry in THP-1 cells transduced with PIWIL4 overexpression vector.

Epigenetic proteins such as MLL1, the enzyme which establishes H3K4me3 marks and NSD1 that establishes the heterochromatic marks H4K20me3 were also identified as potential binding partners. Heat-shock proteins HSP90 and HSP70 that have been shown to interact with piwi proteins in lower organism
also binds with it in AML cells. Moreover, DDX5 and DHX9, putative RNA helicases that also collaborate with piwi during piRNA biogenesis in lower organisms were also identified as potential binding partners.
4. DISCUSSION

AML is a stem cell disease characterized by aberrant expression of stem cell regulatory genes and deregulation of epigenetic marks. Epigenetic modifiers are often aberrantly expressed in AML while recent advancements in high throughput sequencing has led to the identification of several novel recurrent mutations in genes coding for epigenetic enzymes. However, the knowledge pertaining to the mechanism behind the influence of these mutations and deregulated expression of epigenetic modifiers on AML pathogenesis is limited. PIWIL proteins are stem cell self-renewal associated proteins that together with piwi interacting RNA (piRNA) mediate epigenetic silencing of transposable elements in lower organisms (Cox, Chao et al. 1998; Cox, Chao et al. 2000; Aravin, Gaidatzis et al. 2006; Aravin, Hannon et al. 2007; Brennecke, Aravin et al. 2007; Aravin, Sachidanandam et al. 2008; Brennecke, Malone et al. 2008). Recent studies have identified PIWIL genes as putative proto-oncogenes in solid cancers. To this date, not many studies have contributed to the understanding of the functional role of PIWIL proteins in hematopoietic malignancies.

In our study, compared to other PIWIL genes, we observe a high expression of PIWIL4 in human HSPCs (figure 7A). This comes as no surprise since similar results have been observed in murine HSCs (Nolde, Cheng et al. 2013). Similarly, prior studies in solid cancers have described PIWIL4 as a proto-oncogene due to its high expression in tumor cells compared to corresponding normal tissue (Li, Yu et al. 2010; Greither, Koser et al. 2012; Su, Ren et al. 2012; Chen, Liu et al. 2013; Keam, Young et al. 2014). In agreement with this data, we observe PIWIL4 to be aberrantly expressed in AML patients (figure 9). It is of interest though to understand why PIWIL4 has such a broad range of expression in AML subtypes. Particularly, why is PIWIL4 highly expressed in MLL-AF9 and lower expressed in PML-RARA and INV16 AML positive patients and cell lines (figure 10 and 11)?

MLL translocation harboring leukemia in adults is mostly associated with AML and is considered to have an intermediate/poor prognosis while PML-RARA and INV16 promyelocytic leukemia have a favorable outcome. It is possible
that proto-oncogenic expression of PIWIL4 is related to increased proliferation in cancer cells thereby leading to a more rapidly dividing and therefore more aggressive disease. Interestingly, bioinformatic analysis of the promoter region of PIWIL4 using UCSC genome browser (http://genome.ucsc.edu/) in AML cell line K562 yields an enrichment of euchromatic/activating histone marks H3K79me2 and H3K4me3, both of which are associated with MLL leukemia (data not shown). Moreover, high PIWIL4 expressing MLL-AF9 patients showed increased expression of AML and proliferation associated genes such as FLT3, CBL and NRAS compared to low PIWIL4 expressing patients (figure 12). These data bolster the assumption that high PIWIL4 expression could be associated with MLL leukemia and possibly with increased growth rate of leukemic cells in AML. In the future, it would be interesting to further test our hypothesis by checking if the broad range of expression of PIWIL4 within the various MLL translocations harboring patients has any bearing on prognosis or blast percentage or other important prognostic factors. In the absence of such an analysis, the effect of PIWIL4 knockdown in AML cell lines and primary patient cells served as a good alternative to understand the functional role of PIWIL4 and its potential effect on growth rate. We could show that PIWIL4 depleted AML cell lines, particularly MLL-AF9 translocation harboring cells, showed a marked decrease in proliferation rate in vitro and in vivo (figure 14-16). Similarly, an adverse effect on growth was also observed in vitro in PIWIL4 depleted primary MLL-AF9 positive patient BMNCs (figure19). This decrease in proliferation was not a result of apoptosis but could be due to a cell cycle defect as the PIWIL4 depleted AML cell lines showed a slower transition from G0/G1 phase to S phase of the cell cycle (figure 17). The abrogation in growth potential was rescued upon re-expression of shRNA resistant PIWIL4 (figure22). Moreover, overexpression of PIWIL4 in wild type THP-1 and RAJI cell lines resulted in further increase in growth potential in vitro which ties growth of leukemic cells, particularly MLL-AF9 harboring AML cells, firmly with PIWIL4 expression (figure 27). Additionally, it could be that the higher the PIWIL4 levels, the more addicted the AML cells are to its expression, which could be the reason behind why MLL translocation harboring AML cell lines
Discussion

(THP-1, MONO-MAC-6 and MV4-11) and primary patient cells displayed a substantial effect on growth upon its depletion.

Notably, depletion of PIWIL4 in AML cells, particularly in MLL-AF9 harboring cells, was associated with changes in epigenetic marks in the form of global H3K9me3 hypomethylation and localized H3K4me3 hypermethylation (figure 20 and 21). Interestingly, H3K9me3 marks are associated with Piwil function in lower organisms and were directly related to PIWIL4 in human AML since re-expression of PIWIL4 in knockdown experiments resulted not only in the rescue of the retarded growth phenotype but also the re-establishment of H3K9me3 marks. Meanwhile, H3K4me3 gain occurred upon depletion of PIWIL4, therefore, it could be assumed to be independent of PIWIL4 function. While the gain of the H3K4me3 mark was confined to gene regulatory regions, the loss of H3K9me3 marks occurred mostly in non-coding regions as opposed to gene-rich or gene regulatory regions. Therefore, expectedly, differential gene expression in AML cells caused due to PIWIL4 depletion did not strongly correlate with H3K9me3 loss in their gene regulatory regions. These results raise questions, such as-

Is growth retardation related to loss of H3K9me3 in PIWIL4 depleted cells? Was epigenetic change a result of growth retardation or did epigenetic change precede the change in growth potential?

Firstly, genes that suffered H3K9me3 hypomethylation and H3K4me3 hypermethylation in their regulatory regions mostly belonged to cell growth related pathways such as - PI3K-AKT, actin cytoskeletal pathway, and the anti-mitogenic NF-KB and Rap1 signaling pathways (figure 23-25). Additionally, negative regulators of growth such as tumor suppressor NEO1, insulin receptor binding protein GRB10, regulator of NF-KB signaling PRCKQ and CDC42BPA, regulator of cytoskeleton structure lost repressive H3K9me3 marks in their regulatory regions and showed an increase in expression in RNA seq. Moreover, these genes were also found to be deregulated in low PIWIL4 expressing MLL-AF9 harboring AML patients compared to high expressing patients (figure 12). Secondly, differential gene expression analysis from PIWIL4 high vs PIWIL4 low expressing MLL-AF9 patients and PIWIL4
depletion in *MLL-AF9* AML cell line also yielded the PI3K-AKT and actin cytoskeleton pathways. These facts are suggestive of a possible causal relationship between H3K9me3 hypomethylation and growth retardation in PIWIL4 depleted cells and points to epigenetic changes preceding changes in gene expression.

Though this insight offers some explanation, it still does not account for the impact of the majority of H3K9me3 hypomethylation, which exists in intragenic regions, not in gene regulatory regions or on gene bodies. As mentioned before (in the Results section – 3.14, 3.16 and 3.17), due to lack of sequencing depth repetitive regions could not be assayed for H3K9me3 hypomethylation through ChIP-seq or for expression through RNA-seq. It is well known that repetitive elements such as TEs make up approximately 50% of the human genome. Since TEs possess the ability to cause genomic instability by transcribing and inserting themselves in coding/regulatory regions they must be silenced by way of heterochromatin formation through epigenetic modifications such as H3K9me3, H4K20me3 and H3K27me3 which cause condensation of chromatin and prevention of transcription. It would be interesting to know if TEs suffer loss of H3K9me3 marks in PIWIL4 depleted AML cells. If so, does it impact their expression and does their expression affect cell growth? Recent publications have highlighted that global loss of H3K9me3 in human cells could result in TE transcription, causing genomic instability leading to cell cycle defect and cellular senescence (Sidler, Woycicki et al. 2014). We had a similar assumption for AML cells, i.e. the primary role of PIWIL4 is to silence TEs through H3K9me3, and therefore, depletion of PIWIL4 would lead to activation of TEs leading to increase in genomic instability thereby affecting cell growth and viability in AML. Hence, in the future, an analysis of TE sequences and their impact on cell growth must be performed to completely understand the function of PIWIL4 in AML cells.

In addition to establishing the function of PIWIL4, it was also necessary to understand the mechanism behind its effect on H3K9me3 marks. In lower organism Piwil proteins have been proposed to establish H3K9me3 by binding histone methyltransferases through partner proteins and localize to specific loci through their interaction with piRNA (figure 32).
In our experiments on AML cell lines, we observed that a PAZ domain lacking PIWIL4 mutant could not rescue the effects of PIWIL4 depletion i.e. the mutant could not rescue the growth phenotype or the H3K9me3 hypomethylation (figure 22). Since the PAZ domain anchors the 3’end of piRNA to the PIWI protein it is possible that the mutant form does not interact with piRNA and therefore does not localize to the necessary loci in the genome. It could also be that the mutant protein loses all functionality and the protein is quickly degraded. This remains to be tested using western blot and intracellular localization assays.

Meanwhile, in support of the proposed mechanistic model for PIWIL4-piRNA function, using microarray and small RNA deep sequencing we could show that piRNA-like RNAs are indeed expressed in AML cell lines and their expression is perturbed upon PIWIL4 depletion (figure 28-31). This supports the theory that piRNA biogenesis could be linked to PIWIL4 in human AML cells, the same as it is in germ cells of lower organisms.

In *MLL* translocation harboring AML cell line MV4-11 through small RNA deep sequencing we could establish that piRNA like RNA with classical features of
piRNA i.e. uridine at first position or adenine at 10th position could be identified, while reads obtained using this criteria were lower in PIWIL4 depleted cells (figure 29). We could also show that reads of a larger length i.e. 28-32 nucleotides which is the classical length of PIWIL4 bound piRNA were decreased in PIWIL4 depleted AML cells (figure 31). Since piRNA biogenesis is a cyclic process, an accumulation of piRNA is expected if PIWIL4 is lost from the piRNA biogenesis process. In this regard, we also saw an increase in read lengths of 26-28 nucleotides in length which are classically PIWIL2 associated piRNA.

piRNA microarray in THP-1 cells also yielded similar results where upon PIWIL4 depletion piRNA were found to be both upregulated and downregulated (figure 28). Interestingly, deregulated piRNA from our microarray data mapped to various locations in the human genome including coding sequences of 83 genes, which were deregulated upon PIWIL4 depletion. As the model for PIWIL4 function suggests, in the absence of PIWIL4 and its associated piRNA, specific loci should lose H3K9me3 marks and hence show deregulated expression. We could identify 8 genes (CDC42BPA, MYH10, PHACTR1, ACER3, ABHD12, PLXDC2, PLEKHB2 and TMTC2) which upon PIWIL4 depletion showed deregulated expression, loss of H3K9me3 marks and sequence complementarity to deregulated piRNA. Notably, most these genes belonged to the actin cytoskeleton regulation pathway. It was also interesting to note that a few proteins binding to PIWIL4 were also cytoskeletal in nature such as, MYH9 and ACTN4 (table 5). Similar observations have been made in breast cancer cells as well (Keam, Young et al. 2014). In addition, Heat shock proteins that have been intimately linked with PIWI function in lower organisms were also identified as partner proteins in our mass spectrometry analysis. Epigenetic proteins such as NSD1 and MLL1 itself were found bound to PIWIL4. Of note, NSD1 and MLL1 are associated with heterochromatin associated H4K20me3 and euchromatin associated H3K4me3 modifications, respectively. Interestingly, no epigenetic proteins pertaining to H3K9me3 marks were identified. This could be due to indirect or weak interactions between H3K9me3 establishing HMTs and PIWIL4 protein. There is also the possibility that PIWIL4-piRNA machinery take part in
maintenance rather than establishment of H3K9me3 marks. Therefore, PIWIL4 possibly does not associate with HMTs and only localizes to loci where H3K9me3 marks have been established. Proving either theory requires further experiments.

PIWIL4 also binds with numerous RNA helicases, which in lower organisms have been shown to be essential for piRNA biogenesis. Interestingly, a RNA helicase DDX5, which in our data associates with PIWIL4 in AML cell line THP-1, has been shown to effect the growth of AML cells in vitro and in vivo but its depletion has no effect on normal HSPCs (Mazurek, Park et al. 2014). This similarity in effect on growth of AML cells could be down to the deregulation of the PIWI-piRNA pathway, which exclusively effects the growth of AML cells but not normal cells.

In summary, firstly, our data suggests that the PIWI-piRNA pathway targets RhoGTPase related (specifically CDC42 related) genes which play a role in governing cellular cytoskeleton and in turn affect growth rate of cells. It is well known that the RhoGTPase regulated cellular cytoskeleton has a role to play in mitosis, cytokinesis and cell attachment which have bearing on cellular proliferation (Heng and Koh 2010; Provenzano and Keely 2011). CDC42 associated Rho GTPase effectors, activators and the actin cytoskeleton pathway was recurrently deregulated in microarray of MLL-AF9 patients and RNA seq of PIWIL4 depleted MLL-AF9 harboring AML cell lines. Moreover, H3K9me3 ChIP and piRNA microarray showed that genes that suffered H3K9me3 loss in their regulatory regions and had deregulated piRNA map to them also belonged to the actin cytoskeletal pathway. Furthermore, numerous RhoGTPase and cytoskeleton related genes such as integrins and cadherins and downstream pathways like the NF-KB and adhesion pathways identified by KEGG analysis were deregulated in MLL-AF9 patients and AML cell lines alike. PIWIL4 itself was found to be associated with cytoskeletal proteins in AML cells. Our data suggests that PIWIL4, independently through protein-protein interactions, and by way of the piRNA pathway and epigenetic marks such as H3K9me3, affects cellular cytoskeleton and hence proliferation. It is possible that the genes responsible for regulating the cytoskeletal pathway might not be sole targets of the PIWIL4-piRNA epigenetic machinery. Silencing
of TEs via H3K9me3 could yet be of importance and activation of TEs could have a role to play in the massive deregulation of gene expression observed in our dataset, in the form of 2700 genes in PIWIL4 depleted AML cells.

Secondly, our data has established that PIWIL4 depletion does not impact normal HSPCs in vitro (figure 18). Though the effect on long term engraftment has to be checked using NSG mice, prior studies in murine hematopoiesis suggest that a similar result might be attained in vivo for human HSPCs. It is unclear why AML cells are more addicted to PIWIL4 expression. Future experiments in the direction of understanding protein binding partners, the profile of piRNA that bind to PIWIL4 in normal HSPCs and potential differences in the impact on global H3K9me3 and TEs between normal and leukemic cells could help to identify the reason behind our observations.

To conclude, we could show for the first time that PIWIL4, an aberrantly expressed gene in AML, is important for AML growth and is associated with establishment/maintenance of global H3K9me3 marks. Furthermore, we documented that loss of epigenetic marks associated with depletion of PIWIL4 had an impact on gene expression and on AML cell growth in vitro and in vivo.

Furthermore, we could show for the first time that piRNA like RNA are expressed in human AML cells and map to various regions in the human genome. We identified genes that could potentially bind piRNA through sequence complimentarity, which lost H3K9me3 and showed deregulated expression upon PIWIL4 depletion. Most of the genes that fell into this category belonged to the cellular proliferation related RhoGTPase-actin cytoskeleton regulation pathways. PIWIL4 itself was found bound to cytoskeletal proteins as well as epigenetic modifiers such as MLL-1 and NSD1.

These findings suggest that the RNA binding protein, PIWIL4, which is known to direct epigenetic modifications in germ stem cells of lower organisms, also plays a similar role in AML cells by establishing/maintaining H3K9me3 and possibly other epigenetic modifications which have an overall effect on growth of AML cells.
5. **Summary**

PIWIL proteins are stem cell self-renewal and maintenance associated proteins that bind Piwi interacting RNA (piRNA) to mediate epigenetic modifications in lower vertebrates. In solid cancers, PIWIL proteins have been described as putative proto-oncogenes. We could show that a human PIWI like protein - *PIWIL4* is overexpressed in majority of AML patients, particularly in patients harboring the *MLL-AF9* translocation. The knockdown of *PIWIL4* in *MLL-AF9* harboring cell lines and primary patient samples leads to a marked depreciation in growth *in vitro* and *in vivo*, and a global loss of H3K9me3 marks in cell lines. The loss of growth potential and decrease in global levels of H3K9me3 in cell lines could be rescued via overexpression of wild type *PIWIL4*, but not by a *PIWIL4* mutant lacking the piRNA binding - PAZ domain. No growth inhibition or effect on colony formation was observed upon PIWIL4 depletion in normal human hematopoietic stem progenitor, *in vitro*.

*PIWIL4* depleted AML cell lines showed a deregulation of the actin cytoskeleton pathway. Cytoskeletal genes were also deregulated in *PIWIL4* high vs *PIWIL4* low expressing *MLL-AF9* harboring AML patients. Using microarray we could determine that PIWIL4 knockdown induced differential expression of 981 known and predicted piRNAs while through small RNA deep sequencing we could identify piRNA-like-RNA in *MLL-AF9* harboring cells. Immunoprecipitation of PIWIL4 identified cytoskeletal proteins, epigenetic enzymes such as MLL1 and NSD1 and heatshock proteins as binding partners. Thus, collectively, we could show for the first time that *PIWIL4* expression is deregulated in human AML, decreases leukemic growth (while not impacting normal hematopoietic cell growth), shapes epigenetic marks and impacts piRNA expression in this disease.

Future experiments will enable us to examine if *PIWIL4* depletion affects repetitive elements and if an impact on these elements shares a causal relationship with the effect on cell growth. Since we could show that PIWIL4 and heat shock proteins interact in leukemic cells and given the current use of heat shock protein inhibitors for treating MLL patients, it is
6. REFERENCES


